2009-12-08

Studies on the Performance and Impact of Channel Estimation in MIMO and OFDM Systems

Michael David Larsen
Brigham Young University - Provo

Studies on the Performance and Impact of Channel Estimation in MIMO and OFDM Systems

Michael David Larsen

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

A. Lee Swindlehurst, Chair
Brian D. Jeffs
Michael A. Jensen
Michael D. Rice
Karl F. Warnick

Department of Electrical and Computer Engineering

Brigham Young University

April 2010

ABSTRACT

Studies on the Performance and Impact of Channel Estimation

in MIMO and OFDM Systems

Michael David Larsen

Department of Electrical and Computer Engineering

Doctor of Philosophy

The need for reliable, high-throughput, mobile wireless communication technologies has never been greater as increases in the demand for on-the-go access to information, entertainment, and other electronic services continues. Two such technologies, which are at the forefront of current research efforts, are orthogonal frequency division multiplexing (OFDM) and multiple-input multiple-output (MIMO) systems, their union being known simply as MIMO-OFDM. The successful performance of these technologies depends upon the availability of accurate information concerning the wireless communication channel. In this dissertation, several issues related to quality of this channel state information (CSI) are studied. Specifically, the first part of this dissertation considers the design of optimal pilot signals for OFDM systems. The optimization is addressed via lower bounds on the estimation error variance, which bounds are given by formulations of the Cramér-Rao bound (CRB).

The second part of this dissertation uses the CRB once again, this time as a tool for evaluating the potential performance of MIMO-OFDM channel estimation and prediction. Bounds are found for several parametric time-varying wideband MIMO-OFDM channel models, and numerical evaluations of these bounds are used to illuminate several interesting features regarding the estimation and prediction of MIMO-OFDM channels.

The final part of this dissertation considers the problem of MIMO multiplexing using SVD-based methods when only imperfect CSI is available. For this purpose, general per-MIMO-subchannel signal and interference-plus-noise power expressions are derived to quantify the effects of CSI imperfections, and these expressions are then used to find robust MIMO-SVD power and bit allocations which maintain good overall performance in spite of imperfect CSI.

Keywords: OFDM, MIMO, MIMO-OFDM, channel estimation, channel prediction, channel state information, performance bounds, perturbation theory

ACKNOWLEDGMENTS

My deepest appreciation goes to my advisor, Dr. A. Lee Swindlehurst. While the course of my study at BYU went quite differently then probably either of us expected, Dr. Swindlehurst remained a source of continuing guidance, support, and confidence though major changes in both of our lives. I also owe a debt of gratitude to Dr. Michael Jensen for his continued interest in my work and success while Dr. Swindlehurst was away from BYU. It was nice feel like part of a group even when I was no longer really part of one. My thanks also to Dr. Thomas Svantesson for his collaboration and feedback on the material presented in Chapter 4 of this dissertation. Thank you, as well, to my committee for their support and guidance on my research. I would also like to express my love and appreciation to my wife Cathy, our children Evelyn and Paul, and our families, all of whose continual love and support kept me moving forward. Finally, thanks to the Lord for all of my opportunities and successes during the course of this work. I have felt His love and care continually.

Table of Contents

List of Tables **xiii**

List of Figures **xvi**

1 Introduction **1**

1.1 Summary of Contributions 3

1.1.1 OFDM Pilot Design 3

1.1.2 MIMO-OFDM Channel Estimation/Prediction Performance 4

1.1.3 Robust MIMO Multiplexing 6

1.2 Outline of Dissertation 8

2 Background and Modeling **11**

2.1 Notation 11

2.2 Background and Prior Work 12

2.2.1 OFDM 12

2.2.2 The Cramér-Rao Bound 13

2.2.3 OFDM Pilot Design 14

2.2.4 SISO and MIMO Channel Prediction 16

2.2.5 MIMO and Waterfilling 18

2.2.6 Adaptive Modulation in MIMO 20

2.3 OFDM Signal Model 22

2.4 Parametric MIMO Wideband Models 24
2.4.1 Structured DOD/DOA Model 25
2.4.2 Unstructured Vector Model 26
2.4.3 Unstructured Matrix Model 26
2.5 MIMO Perturbation Model 27

3 Optimal Pilot Design for Joint Time Delay and Channel Estimation in OFDM 29

3.1 Introduction 29
3.2 Signal Model 30
3.3 The CRB 31
3.4 Optimal Pilot Selection 34
3.4.1 Uncoupled CRB Cases 35
3.4.2 Coupled CRB Case 36
3.5 Pilot Selection using the CRB-FOP 39
3.5.1 Derivation of the CRB-FOP 39
3.5.2 The CRB-FOP and Pilot Selection 43
3.6 Numerical Simulations 44
3.7 Conclusions 49

4 Performance Bounds for MIMO-OFDM Channel Estimation 53

4.1 Introduction 53
4.2 Channel Models 54
4.2.1 DOD/DOA Model 55
4.2.2 DOD/DOA Model with Calibration Errors 56
4.2.3 Vector Spatial Signature Model 57

4.2.4 Matrix Spatial Signature Model 58
4.3 Lower Bound on Estimation/Prediction Error 59
4.3.1 Direct vs. Indirect Use of Pilots 62
4.3.2 DOD/DOA Model 64
4.3.3 DOD/DOA Model with Calibration Errors 66
4.3.4 VSS CRB 67
4.3.5 MSS CRB 67
4.4 Numerical Simulations 68
4.4.1 The Channel Models 70
4.4.2 Calibration Error 72
4.4.3 Array Sizes 75
4.4.4 Number of Multipath Components 77
4.5 Conclusions 78

5 Power Allocation and Bit Loading for Spatial Multiplexing with Imperfect CSI 83

5.1 Introduction 83
5.2 Derivation of the Subchannel SINRs 84
5.2.1 General Case: MCSI 86
5.2.2 Special Case: CCSI 90
5.2.3 Effect of Channel Estimation Errors 91
5.2.4 Effect of Channel Mobility 91
5.2.5 SINR for Specified Outage Rates 92
5.2.6 Numerical Examination of the SINR 97
5.3 Subchannel Power Allocation 101
5.3.1 MCSI Subchannel Power Allocation 102

5.3.2 CCSI Subchannel Power Allocation . . . 103
5.4 Beamforming vs. Multiplexing Thresholds . . . 105
5.4.1 CCSI Thresholds . . . 105
5.4.2 CCSI Thresholds for Specified Outage Rate . . . 106
5.4.3 Separate Estimation Error and the Gauss-Markov Model . . . 107
5.4.4 MCSI Thresholds . . . 107
5.5 Numerical Simulations . . . 110
5.6 Conclusions . . . 115

6 Conclusions and Open Problems **119**
6.1 Conclusions . . . 119
6.2 Open Problems . . . 121

Bibliography **124**

A Derivation of a CRB for Biased Estimators **133**

B Derivation of CRB Matrices **135**
B.1 Derivatives for the Jacobian and Fisher Matrices . . . 135
B.1.1 Derivatives with Respect to σ . . . 135
B.1.2 Derivatives with Respect to Re[$\boldsymbol{\alpha}$] (DOD/DOA) . . . 135
B.1.3 Derivatives with Respect to Im[$\boldsymbol{\alpha}$] (DOD/DOA) . . . 135
B.1.4 Derivatives with Respect to $\boldsymbol{\omega}_d$ . . . 135
B.1.5 Derivatives with Respect to $\boldsymbol{\tau}$ . . . 136
B.1.6 Derivatives with Respect to $\boldsymbol{\Omega}_t$ (DOD/DOA) . . . 136
B.1.7 Derivatives with Respect to $\boldsymbol{\Omega}_\mathbf{r}$ (DOD/DOA) . . . 136
B.1.8 Derivatives with Respect to $\mathbf{A}_t$ (VSS) . . . 136

B.1.9 Derivatives with Respect to $\mathbf{A}_r$ (VSS) 136
B.1.10 Derivatives with Respect to $\mathbf{A}$ (MSS) 137
B.2 Derivation of CRB Submatrix 137
B.3 Expressions for FIM Subblocks 138
B.3.1 DOD/DOA CRB 138
B.3.2 DOD/DOA CRB with Calibration Errors 139
B.3.3 VSS CRB 140
B.3.4 MSS CRB 141

C Additional Beamforming/Multiplexing Thresholds 143

C.1 Thresholds Accounting for Estimation Error 143
C.2 Thresholds for the Gauss-Markov Channel Model 143

List of Tables

2.1 Channel State Information Relationships 28

3.1 Channel Impulse Response Coefficients 44

4.1 Number of Real-Valued Model Parameters 59

List of Figures

3.1 Example of DPS and CPD pilot design using $\mathbf{G}(\alpha, \mathbf{b})$ with $\alpha = 0$. 45

3.2 Example of DPS and CPD pilot design using $\mathbf{G}(\alpha, \mathbf{b})$ with $\alpha = 1$. 46

3.3 Example of DPS and CPD pilot design using $\mathbf{G}(\alpha, \mathbf{b})$ with a $\gamma_{\tau_d}^{-1}$ constraint. 48

3.4 Demonstration of the influence of α on the terms of $\mathbf{G}(\alpha, \mathbf{b})$. 49

3.5 Example of DPS and CPD pilot design using the CRB-FOP method with a constraint on $\lambda_{\tau_d}^{-1}$. 50

4.1 Example of 2D normalized error bound results for the DOD/DOA and VSS models. 71

4.2 Comparison of bounds for all discussed models in frequency. 72

4.3 Comparison of bounds for all discussed models as a function of wavelength. . 73

4.4 Examination of the normalized error bound as a function of calibration error. 74

4.5 Plot of the normalized error bound in frequency for various array sizes. . . . 75

4.6 Plot of the normalized error bound versus wavelength for various array sizes. 76

4.7 The effect of frequency measurement spacing on performance. 78

4.8 The effect array sizes and frequency measurement spacing on performance. . 79

4.9 The normalized error bound for various path numbers L in frequency. 80

4.10 The normalized error bound for various path numbers L versus wavelengths. 81

5.1 Comparison of the MCSI SINR expression results with Monte Carlo results for the first and fourth subchannels. 98

5.2 A comparison of the MCSI and CCSI cases via an examination of the per-subchannel SINRs as a function of ρ. 99

5.3 A comparison of the per-subchannel SINR expressions with Monte Carlo simulation. 100

5.4 Example of multiplexing/beamforming thresholds on $\sigma^2_{H,1}$ and $\sigma^2_{H,2}$. 111

5.5 A comparison of designed bit-loading levels and average bits per transmission achieved for uniformed and informed bit-loading schemes. 116

5.6 A comparison of designed bit-loading levels and average bits per transmission achieved for uniformed and informed bit-loading schemes with confidence level $p = 0.1$. 117

5.7 A comparison of designed bit-loading levels and average bits per transmission achieved for the MCSI scenario. 118

Chapter 1

Introduction

The last several years have seen a remarkable growth in the demand for and deployment of mobile wireless communication systems with large data-throughput capabilities. In the past, the demand for such systems came more from the realm of the military. Today, however, the use of such systems has become mainstream, driven by consumer devices such as wireless netbooks and laptops, smartphones, and other readily available mobile communications devices. As the number of people using such systems increases, so does the need for reliable and efficient techniques for both high-throughput and mobile communications.

One way to increase the throughput of a communication system is to increase the system bandwidth or the efficiency of the bandwidth utilization. Wireless wideband systems frequently must operate in both indoor and outdoor multipath channels, i.e., in cluttered environments that result in multiple copies of transmitted signals arriving at the receiving devices at slightly different times. These multipath signals result in wireless communications channels which are frequency selective, altering the transmitted signals differently at different frequencies. When not properly handled, frequency selective channels may severely corrupt signals passing through them.

An early and competitive method for combating frequency selective channels is frequency division multiplexing (FDM), in which a single information signal is transmitted over several lower-rate frequency subchannels whose individual narrowband channel responses appear constant in frequency [1,2]. In order to prevent interference between signal components modulated over the different subchannels in an FDM system, the subchannels are generally required to be orthogonal in frequency. While this may be accomplished by isolating the subchannel bandwidths one from another in frequency, the resulting bandwidth utilization will be inefficient. A more efficient bandwidth utilization is obtained by allowing the fre-

quency subchannels to overlap while maintaining the requirement that they be orthogonal. This method, aptly named orthogonal frequency division multiplexing (OFDM), was first proposed during the early research into FDM methods [3–6]. While early technical hurdles prevented its widespread adaption immediately following its conception, OFDM is now widely used in wired and wireless communication systems [3].

For many years, efforts to increase communication throughput focused almost exclusively on the increase and efficient utilization of frequency bandwidth. It was found more recently, however, that in multipath-rich environments, the spatial environment itself could be exploited in order to increase throughput through the use of multiple antennas [7–9]. Analytical results for systems using multiple antennas at both sides of the communication link, which are also known as multiple-input multiple-output (MIMO) systems, indicate a direct relationship between the number of antennas and the system capacity, a theoretical measure of throughput. Thus, MIMO provided system designers a previously-untapped way to increase system throughput.

The recent popularization of both OFDM and MIMO led naturally to the merging of these to individually advantageous communication techniques. In the combination, known as MIMO-OFDM, multiple antennas are used to transmit and receive the OFDM signals, resulting in a signal which is potentially multiplexed in both space and frequency. While the challenges posed by such systems are greater than those posed by the individual systems alone, the benefits are also great. Currently, many of the next generation communication systems proposed, such as IEEE 802.16 (WiMAX) [10–12] and 3G's Long-Term Evolution (LTE) [13–15], incorporate MIMO-OFDM [16, 17].

It is said that there is no such thing as a free lunch, and that statement is borne out in both OFDM and MIMO system design. The increases in throughput gained when using OFDM and/or MIMO methods come at the cost of the additional complexity inherent in these systems. Moreover, the performance of coherent transmission schemes depends upon the communication system's ability to obtain and maintain accurate estimates of parameters related to the transmission environment, which we refer to simply as the channel parameters. As a consequence, studies related to channel parameter estimation for OFDM, MIMO, and MIMO-OFDM systems are well warranted. Additionally, the investigation of transmission

techniques demonstrating robustness to errors in the channel parameter estimates is also worthwhile. Such topics are the focus of this dissertation.

1.1 Summary of Contributions

1.1.1 OFDM Pilot Design

The performance of coherent ODFM systems is dependent on the accurate estimation, either implicitly or explicitly, of the parameters of the communications channel. Accurate channel estimation is necessary for the successful reception and decoding of the transmitted data, while good estimates of timing and frequency offsets are critical in order to prevent OFDM inter-subcarrier and inter-symbol interference. Since such estimation is often accomplished via the transmission of known pilot symbols or tones, the study and design of optimal pilot signals may result to large dividends in terms of system performance.

A good overview of many of the issues related to training design for channel parameter estimation in OFDM systems is presented in [3]. This includes a discussion of many of the channel parameters of interest in OFDM systems, such as the channel impulse response, the timing offset or time delay, and the frequency offset parameters. In [3], it is noted that OFDM channel estimation is typically carried out through the use of pilot tones interspersed among the data-bearing symbols in frequency. Optimal pilot designs of this type, utilizing both single and multiple OFDM symbols for channel estimation, are investigated in [18–23]. In particular, it is shown in [19, 20, 23] that the optimal pilot configuration in terms of the MMSE is equipowered pilot tones equally spaced over the OFDM subchannels. It is further shown in [19] that the capacity bound is maximized when the smallest required number of pilots are used.

Several papers also address the issue of OFDM training design for the estimation of time and frequency offsets. In [24–27], the correlation properties of the cyclic prefix are used for coarse timing and frequency offset estimation without the use of additional training. In [28–35], preamble training sequences dedicated solely to the synchronization problem are presented. Such sequences are used to estimate the timing and/or frequency offsets prior to the transmission of data. In [36, 37], the design and placement of null and pilot tones in the OFDM subchannels is discussed with relation to frequency offset estimation. Additionally,

several methods jointly consider channel estimation as a means to refine timing and frequency offset estimates [32, 38, 39]. However, training signals used in these papers are typically not designed to be optimal for the joint channel and offset estimation problem.

In the first part of this dissertation, the problem of optimal subchannel pilot design and placement for joint channel and timing offset estimation in OFDM systems is studied, with particular attention given to the problem of pilot design for the minimization of a lower bound on the channel estimation error variance while meeting a performance constraint on the variance of the timing offset estimate. To the author's knowledge, this problem has not been previously addressed in the literature. This pilot design problem is addressed through the use of lower bounds on the channel parameter estimation error variances, which are found via two separate Cramér-Rao bound (CRB) formulations. The first bound is found by direct application of a well-known formula for the CRB [40], and the above-mentioned optimization problem is then cast as the minimization of a weighted trace of this CRB, where the weighting is introduced as a means to control the relative importance given to the timing offset and channel estimation components of the problem. This cost function, whose minimization in terms of the pilot structure may not be solved in closed form, is then used to determine many of the general properties of the optimal solution. The second bound is found directly using a CRB-as-a-function-of-parameters formulation [41]. This derivation, which reduces the coupling between timing-offset- and channel-parameter-related regions of the CRB, allows the above-mentioned optimization problem to be separated into two simpler steps, the satisfaction of the timing error constraint followed by the channel impulse response error minimization.

1.1.2 MIMO-OFDM Channel Estimation/Prediction Performance

As with OFDM, the performance of MIMO-OFDM critically depends on its ability to deal with the frequency- and time-selective nature of the wideband propagation channel. The general problem of channel estimation in MIMO-OFDM systems is typically addressed through the use of pilot symbols strategically placed at different times and on different subcarriers [3, 12, 42–44]. Once the channel is estimated at the time-frequency locations of the pilot symbols, estimates at other times and frequencies are typically found using interpola-

tion. In addition, it is often advantageous to predict or extrapolate channel estimates into the future as well. Channel prediction is particularly useful for bridging the gap between the channel estimates and the current channel state in schemes that employ adaptive modulation or power control [45–48]. The problem of channel prediction has been explored for single-input single-output (SISO) channels in [47, 49–52], for single-input multiple-output (SIMO) channels in [53], and for MIMO channels in [54–58].

As in any estimation application, it is useful to quantify the best performance that may be achieved from channel interpolation and prediction. Performance bounds can serve as a standard against which various estimation and prediction techniques are evaluated. Such bounds may also indicate characteristics of the problem that require extra attention for optimal estimation and prediction performance. In [59], the CRB was used to evaluate prediction performance for narrowband SISO channels, and it was found that channel knowledge over several wavelengths of motion was required to achieve acceptable prediction performance even for short intervals in the future. A similar performance bound was found for narrowband, time-fading MIMO channels [60], which demonstrated that the prediction and interpolation performance for MIMO channels exceeds those possible for the SISO case. Apparently, the increase in the number of parameters required for the MIMO case is more than offset by the extra information about the physical channel provided by the multiple antennas. In a related work, a constructive bound is found in [61] for the error performance of a frequency-selective, block-fading channel. This bound is based on an FIR channel model and does not consider training to be distributed in frequency as in an OFDM system.

In the second part of this dissertation, the theoretical performance of pilot-based channel interpolation and prediction found in [60] is expanded from narrowband MIMO channels to frequency-selective, time-fading, wireless MIMO-OFDM channels, where the analysis is carried out via bounds for the interpolation and prediction error of the channel. This analysis is based on several extended ray-based parametric channel models for MIMO channels, and the lower bounds on the interpolation and prediction error for these models are derived using vector formulations of the CRB as a function of parameters for unbiased estimators. A new CRB formulation is also derived (see Appendix A) to enable a similar approach for a specially-structured biased estimator based on one of the channels models.

The analysis of these bounds demonstrates that better estimation and prediction performance of OFDM channels can be obtained using MIMO systems. It is also shown that while models including angle-based antenna array representations enjoy a performance advantage for perfectly calibrated arrays, the advantage is quickly lost when modeling errors are taken into account. In particular, the derived bounds allow one to compute the size of the calibration error required before less-structured models yield better results. The results of the analysis also have implications for pilot design, with numerical results demonstrating that MIMO-OFDM systems may function with a lower time and frequency pilot density than a corresponding SISO implementations, even after taking into account the fact that more pilot symbols are required to estimate a MIMO channel.

1.1.3 Robust MIMO Multiplexing

The first two parts of this dissertation consider issues related to channel parameter estimation and performance for OFDM and MIMO-OFDM systems. In contrast, the final part of this dissertation focuses on quantifying and mitigating the effects of errors in the channel estimates for narrowband MIMO systems. Given perfect channel state information (CSI) at the transmitter and receiver of a MIMO communications link, along with no restrictions on codebook design, the well-known waterfilling solution maximizes information throughput given a fixed transmission power. The waterfilling solution achieves this goal through use of the singular value decomposition (SVD) to separate the MIMO channel into independent single-input single-output subchannels enabling interference-free data multiplexing [7,9]. Based on the signal-to-noise ratio (SNR) of the resulting subchannels, bit-loading and coding schemes are typically devised to approach the available channel capacity.

In practice, two difficulties arise that must be considered when designing an SVD-based communication scheme. First, it is not possible to obtain perfect CSI since realizable channel estimates are formed from noisy measurements. Mobility in the communication system adds additional complications for CSI estimation, since the MIMO channel changes rapidly when the transmit or receive arrays are in motion. When noisy or outdated CSI is used in conjunction with an SVD-based multiplexing method, the MIMO subchannels become coupled, resulting in potentially severe subchannel power loss and mutual interference.

As a consequence, subchannel bit-loading levels selected based on the assumption of perfect CSI may no longer meet given error performance constraints. For these constraints to be met, back-off strategies are often employed to reduce the subchannel bit rates and create an error margin that accounts for the CSI mismatch. However, without knowledge of the levels of the signal loss and interference induced by the subchannel coupling, it is not clear how to systematically carry out such reductions.

A second difficulty arises from the finite nature of the symbol constellations used for bit allocation. In order for the optimality of the waterfilling power levels to hold, subchannel bit loading requires an infinite-length codebook with continuous modulation order. The use of discrete codebooks and finite constellations such as quadrature amplitude modulation (QAM) results in additional performance loss when used with the power levels chosen by the waterfilling method which were selected assuming idealized symbol constellations [62–64]. When practical symbol constellations are to be used, subchannel power level selection incorporating information about the symbol constellations may help reduce such losses due to mismatches between theory and practice.

Capacity-optimal transmission strategies when only imperfect CSI is available have been studied in [65–68] and the references therein. Minimum symbol-error-rate (SER) optimization and adaptive modulation have also been studied in [69, 70]. These papers generally advocate multiplexing or beamforming using other information, such as channel covariance information, to guide their design, and, unlike SVD-based schemes, often require complex decoding schemes at the receiver.

Several other studies seeking to retain the elegant structure of MIMO SVD-based solutions have focused on the impact of imperfect CSI and finite modulation. The SER of MIMO SVD-based multiplexing methods in the presence of imperfect CSI has been examined in [71] for BPSK transmission. The effect of CSI errors on bit error rate (BER) using more general constellations including M-ary QAM (M-QAM) has been analyzed in [64, 72–75]. With an understanding of the effect of imperfect CSI and finite modulation schemes on SVD-based multiplexing, the primary design task becomes the selection of subchannel power and bit-loading levels. For instance, having derived an expression for the BER, [73] proposes an ad-hoc method to select power and bit-loading levels in each subchannel to meet

a chosen BER constraint. For the special case of uniform power allocation over all of the MIMO subchannels, [64] develops an expression for an adjusted per-subchannel signal-to-interference-plus-noise ratio (SINR) that may be used to select subchannel bit-loading levels [76]. The work of [63] derives the optimal bit loading assuming finite symbol constellations and perfect CSI, and proposes an ad-hoc method to approximate that optimal solution.

The final part of this dissertation develops closed-form expressions for the subchannel SINRs in order to quantify the impact of imperfect CSI on SVD-based signaling schemes. Unlike similar expressions appearing in literature which typically assume the CSI is identical at both ends of the communications link or that perfect CSI is available at the receiver, the expressions of this work allow for different levels of CSI error at the transmitter and receiver. These SINR expressions, developed using the first-order SVD perturbation analysis of [77], are then used to facilitate robust bit-loading design through the systematic selection of the subchannel power levels. Bit-loading designs for the general imperfect CSI case are carried out using ad-hoc iterative techniques. However, for the special case of identical CSI at both ends of the link, the SINR expression is used to derive approximately optimal subchannel power levels for M-QAM signaling under a subchannel SER constraint. SINR expressions are also used to find thresholds for the amount of channel uncertainty and measurement noise above which beamforming should be used instead of spatial multiplexing over the SVD-based subchannels.

1.2 Outline of Dissertation

The remainder of this dissertation will proceed as follows. In Chapter 2, some basics of OFDM and MIMO systems will be discussed. This will include the presentation of channel models for various OFDM, MIMO, and MIMO-OFDM systems that will be studied in this work. Prior work related to the specific topics of this dissertation will also be presented, and important terms and notation will be defined.

Chapter 3 of this dissertation focuses on single antenna OFDM systems. Based upon the assumption that channel parameter estimation will be accomplished using pilot tones in the OFDM subchannels, this chapter explores the problem of optimal pilot signaling relative

to parameter estimation in OFDM systems by using performance bounds on the variance of channel parameter estimates of interest.

In Chapter 4, the focus is expanded to MIMO-OFDM systems, of which single antenna systems are a special case. In this chapter, though, it is assumed that the training signals are known. The problem then becomes determining the best channel parameter estimation performance possible given a training signal distributed in both time and frequency. The resulting bounds have implications concerning both the benefits of using MIMO in wideband systems and the design of training signals for such systems.

Chapter 5 considers the case when inaccurate channel parameter estimates are used in a narrowband MIMO system. In this chapter, expressions are derived which quantify the impact of the channel estimate inaccuracies on the system performance when using a popular MIMO transmission scheme. These expressions are then used to modify this scheme, increasing its robustness to estimation inaccuracies. Thresholds are also derived which address the question of whether or not spatial multiplexing should be employed in the MIMO system.

A summary of the results of these efforts and a discussion of the potential for future research are presented in Chapter 6.

Chapter 2

Background and Modeling

The purpose of this chapter is to prepare the reader for the discussions which follow in the remainder of this dissertation. Specifically, in what follows we will list commonly used notation; present general background information relative to orthogonal frequency division multiplexing (OFDM), multiple-input multiple-output (MIMO) systems, and the Cramér-Rao bound (CRB); and introduce specific models that will be used in the remaining chapters.

2.1 Notation

The following mathematical notation will be used frequently. Non-bold uppercase and lowercase letters such as A and a represent scalar variables and functions, lowercase bold letters such as $\mathbf{a}$ denote column vectors, and uppercase bold letters such as $\mathbf{A}$ represent matrices. The identity matrix is denoted by $\mathbf{I}$ with or without a subscript. If it has a subscript, such as $\mathbf{I}_N$, the subscript denotes the number of rows and/or columns of the matrix. The transpose of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}^T$, the conjugate (or Hermitian) transpose is denoted by $\mathbf{A}^H$, and the Moore-Penrose pseudoinverse is denoted by $\mathbf{A}^\dagger$. The trace of a matrix $\mathbf{A}$ is written as $\mathrm{tr}[\mathbf{A}]$. The matrix vectorization operator applied to the matrix $\mathbf{A}$, $\mathrm{vec}[\mathbf{A}]$, denotes the vector created by stacking the columns of $\mathbf{A}$. The binary operator $\otimes$ indicates the matrix Kronecker product, e.g., $\mathbf{A} \otimes \mathbf{B}$, while the binary operator $\odot$ indicates the element-wise or Hadamard matrix product. The i,j-th element of a matrix and the i-th element of a vector are represented as $[\mathbf{A}]_{ij}$ and $[\mathbf{a}]_i$, respectively. For the notation $(\cdot)^+$, $(a)^+ = a$ if $a > 0$; otherwise, $(a)^+ = 0$. The notation $(\cdot)^*$ is used for two purposes (the selection of which should be clear in context): to indicate the conjugate of a complex value, e.g., $\mathrm{conjugate}(\mathbf{A}) = \mathbf{A}^*$, and to indicate the optimality of a specific quantity. If the matrix $\mathbf{A}$ is complex-valued, then $\mathrm{Re}[\mathbf{A}]$ and $\mathrm{Im}[\mathbf{A}]$ are the real and imaginary parts

of $\mathbf{A}$. The set notation $\{a_i\}$ is used to represent the set of items $a_1, a_2, \ldots, a_N$. The notation $x \sim \mathcal{N}(y, z)$ indicates that value of x is drawn from a normal or Gaussian distribution with mean y and variance z. For complex proper random variables, a similar notation is given by $x \sim \mathbb{C}\mathcal{N}(y, z)$, where x and y are complex-valued and z is real-valued. When the previous notation is applied to vectors, a multivariate Gaussian distribution is assumed. Finally, the expectation of a random variable X is denoted by $E[X]$.

2.2 Background and Prior Work

2.2.1 OFDM

Orthogonal frequency division multiplexing is a signal modulation technique which is used to transmit high-rate signals by transmitting several lower-rate signals in parallel over orthogonal frequency subchannels. A good introduction to OFDM is given in [3]. An OFDM symbol of duration T consists of the sum of N modulated subcarriers and may be written as

$$s(t) = \begin{cases} \mathrm{Re}\left[\sum_{k=-\frac{N}{2}+1}^{\frac{N}{2}} b_{k+N/2-1} e^{j2\pi\left(f_c - \frac{k+1/2}{T}(t-t_0)\right)}\right], & t_0 \le t \le t_0 + T \\ 0, & \text{otherwise} \end{cases}, \tag{2.1}$$

where t_0 is the symbol start time, b_k, $k = 0, \ldots, N-1$, are the data symbols, f_c is the symbol carrier frequency, and N is assumed to be a multiple of two. The OFDM symbol is often expressed in complex baseband notation as

$$s(t) = \begin{cases} \sum_{k=-\frac{N}{2}+1}^{\frac{N}{2}} b_{k+N/2-1} e^{j2\pi\frac{k}{T}(t-t_0)}, & t_0 \le t \le t_0 + T \\ 0, & \text{otherwise} \end{cases}, \tag{2.2}$$

which is essentially the inverse Fourier transform of the N input symbols b_k, $k = 0, \ldots, N-1$. The individual subcarriers given by the complex exponentials in (2.2) are orthogonal over the interval $t_0 \le t \le t_0 + T$. When using a discrete-time equivalent of this model, the inverse Fourier transform may be replaced by the inverse discrete Fourier transform (IDFT). In practice, the subchannel modulation is carried out very efficiently using the inverse fast Fourier transform (IFFT). At the receiver, the OFDM symbol is sampled at least N times

and the fast Fourier transform (FFT) is then used on these samples to demodulate the data symbols.

Because the individual subcarrier transmissions occur at a much lower rate, or, equivalently, because the subcarrier frequency bands are relatively narrow, individual OFDM symbols are remarkably robust against the detrimental effects of multipath channels. However, it is still possible for channel multipath to cause inter-symbol interference between neighboring OFDM symbols. To counter this effect, a guard time which is longer than the multipath delay spread is added to each OFDM symbol so that multipath components from one symbol cannot interfere with the subsequent symbol. The OFDM symbol is cyclicly extended over this guard time so that all of the multipath components of the signal will maintain the orthogonality properties of the original signal, thus preventing inter-carrier interference between the subchannels.

2.2.2 The Cramér-Rao Bound

In the study of parameter estimation techniques, it is useful to quantify the best performance that may be achieved by a particular estimator. Performance bounds can serve as a standard against which various estimation and prediction techniques are evaluated. Such bounds may also be used as a design tool. In this context, particular features of a system are selected with the objective of minimizing the performance bound, and those selections which achieve this goal are considered optimal. One of the most popular of these bounds is the Cramér-Rao bound, which provides a lower bound for the variance of unbiased estimators [41, 78, 79]. If an estimator's performance achieves the CRB, the estimator is said to be an efficient estimator.

The CRB has several key features that make it an attractive choice as a lower bound. First, the CRB is relatively easy to determine when compared to other performance bounds. Next, CRB theory often allows us to determine whether or not estimators exist that achieve the CRB. If it is determined that no CRB-achieving estimator exists, it is still often possible to find estimators that approximately achieve the bound. Finally, CRB theory is strongly connected with the theory of maximum likelihood (ML) estimation. ML estimators are

known to be asymptotically efficient and estimators which achieve the CRB are ML estimators.

We will use the CRB in Chapters 3 and 4 of this dissertation. In Chapter 3, the CRB is used as a design tool for selecting optimal pilot sequences for OFDM systems. In Chapter 4, the CRB is used to evaluate the relative performance of several channel parameter models in the context of MIMO-OFDM channel estimation. Note that several CRB formulations exist, of which we will use only a few, including one derived for this work, in the subsequent chapters. The specifics of those formulations are discussed within the chapters as needed.

2.2.3 OFDM Pilot Design

As may be noted from the earlier discussion in the Chapter 1.1.1, extensive work has been carried out relative to the estimation of OFDM channel parameters. It is not our intention here to provide an exhaustive review of those previous studies. Instead, we wish to highlight a few works that are directly related to or provide understanding for the methods to be used in this dissertation.

Pilot design for OFDM channel impulse response estimation has been extensively studied in the literature, and many different pilot structures have been proposed [3]. One of the most popular general training structures involves the interspersal of pilot tones among the data bearing symbols in the MIMO subchannels. Relative to the OFDM model of (2.2), this amounts to using a subset of the N symbols $\{b_k\}$ as pilot tones and the remainder as data bearers. In [19, 20, 23], it was found that the optimal pilot allocations in terms of mean square error (MSE) minimization for this style of training are given by evenly spacing the selected number of pilot tones over the OFDM subchannels and by allocating equal power to each pilot. Using the notation of [20], the equispacing of the pilot tones has the following mathematical description. Let $\{n_p\}$ be the set of subchannel indices into $\{b_k\}$ corresponding the the pilot tones. Then the optimal pilot tone locations are given by

$$n_p = p_0 + pV, \tag{2.3}$$

where $p = 1, \ldots, P$ is the index into the pilot tones set, $V \in \mathbb{Z}$ is the pilot tone spacing, and $p_0 \in \{0, \ldots, V-1\}$ is an initial location offset. For the optimal pilot tones, it is required that the spacing V satisfy $PV\phi/N \in \mathbb{Z}$ and $V\phi/N \notin \mathbb{Z}$ for all $\phi \in \{-L+1, \ldots, L+1\} \backslash \{0\}$, where L is the channel impulse response length. For a given number of pilot tones P, multiple spacings V may be found which satisfy these conditions. Also, there are several choices for P which satisfy the above optimality conditions. Clearly, then, the optimal pilot configuration is not unique. Note that such pilot configurations are also optimal relative to the maximum likelihood criterion assuming additive white Gaussian measurement noise at the receiver. Thus, it is not surprising that this solution appears as a limiting case in the CRB-based optimal pilot discussion of Chapter 3.

In OFDM systems, the channel timing offset or time delay is generally estimated in two ways. The first makes use of the cyclicly-extended guard time, called the cyclic prefix, that is added to OFDM symbols to prevent multipath-induced inter-symbol interference [24]. Because the cyclic prefix is simply a periodic extension of the OFDM symbol, it introduces correlations into the OFDM symbol that may be used to obtain coarse estimates of the time delay, as well as estimates of the frequency offset. The second method makes use of special preamble training sequences. These sequences are specifically designed to have good time correlation properties, allowing the start time of the OFDM symbol to be estimated via a correlator at the receiver. One of the principal studies dealing with this type of preamble design is [28], though many various flavors of this technique exist in the literature. Note that both of these techniques give only coarse timing estimates. If better estimates are needed, other techniques must be used to refine the estimates. In the work of Chapter 3, it is assumed that the OFDM system has coarse timing estimates available. Thus, the developments of that chapter with regard to time delay estimation fit under the umbrella of timing estimate refinement.

It is instructive to note that the time delay and frequency offset parameters are modeled in a very similar manner, with the expressions containing the frequency offset in the OFDM model prior to the FFT at the receiver appearing almost identical to the time-delay-containing expressions following the FFT. Thus, time-domain techniques for estimating the frequency offset may be very similar to frequency-domain techniques for estimating the

time delay. In [37], optimal pilot sequences were found relative to frequency offset estimation using the CRB. As a basic result, this study determines that the optimal pilot design for frequency offset estimation consists of a single pilot located at the end of the frequency band (assuming a flat channel frequency response). This result is very similar to a limiting case of the results found in Chapter 3 of this dissertation.

In [29], the problem of optimal pilot design via the CRB for the joint estimation of the OFDM channel impulse response and carrier frequency offset is considered. Interesting, at the outset of this paper, the development of the CRB parallels that found in the first part of Chapter 3. However, the training sequences considered in [29] are preamble training sequences and, instead of considering pilot design relative to the full joint CRB, this work utilizes the modified CRB (MCRB) [80], which uncouples the time delay and channel impulse response portions of the CRB. Once the MCRB is obtained, asymptotic approximations are made to the MCRB in order to find the minimax optimal preamble training sequence, which they determine is a white time sequence.

2.2.4 SISO and MIMO Channel Prediction

In Chapter 4 of this dissertation, bounds are developed for the interpolation and prediction performance of wideband MIMO (e.g. MIMO-OFDM) channels given previously-obtained estimates of the channel at particular times and frequencies. While the interpolation of channels is discussed frequently in literature, the prediction of channels less frequently addressed. As a reference for the discussion of Chapter 4, we now wish to briefly discuss a few prediction methods appearing in the literature. In general, channel prediction methods rely on models of the communications channel. Prediction is carried out by first estimating the model parameters, and then using the resulting estimates within the model to extrapolate the channel. As a result, the different methods are often most distinguishable by the models upon which they rely. With respect to prediction, one of the most common ways to model the channel is as a sum of complex sinusoids, each with different magnitudes and phase.

In [52], the single-input single-output (SISO) channel is modeled as

$$h(t) = \sum_{i=1}^{L} A_i e^{j(2\pi f_i + \theta_i)}, \tag{2.4}$$

where L is the number discrete signal paths and A_i, f_i, and θ_i are the path attenuation, Doppler shift, and phase shift for the i-th path, respectively. Having this model, [52] then uses the ROOT-MUSIC method of [81] to estimate the unknown parameters of the sinusoid.

As another example, in [49], a narrowband SISO channel is modeled as

$$h(x) = \sum_{i=1}^{N} a_i e^{jkx\cos(\theta_i)}, \tag{2.5}$$

a sum of spatially-dependant sinusoids, with each sinusoid representing a signal path arriving from one of a set of N physical scatterers. In this model, x is the receiver position, k is the wavenumber, a_i is the complex scattering amplitude of path i, and θ_i is the angle between the direction of receiver motion and the direction of the i scatterer. Assuming a constant velocity, this model is simplified to a corresponding spatial sampling model as

$$h(m) = \sum_{i=1}^{N} a_i e^{jmu_i}, \tag{2.6}$$

where $u_i = k\Delta x\cos(\theta_i)$ are normalized Doppler shifts and Δx is the distance between two samples. Given this simplified model, the model parameters are estimating using a variant of the ESPRIT method [82]. This same method was used in [50] to estimate the parameters of a related wideband channel model. Note that the model of (2.6) may also be considered to be a finite impulse response model, which class of models are sometimes considered separately from the class of sum-of-sinusoid models. The ESPRIT algorithm was also used to estimate the parameters of a MIMO channel model for the purposed of prediction in [56].

Another popular class of models used for channel prediction is the class of autoregressive (AR) models. In these models, the channel evolves in time according to an evolution model such as

$$h(t) = \sum_{i=1}^{L} A_i h(t-\tau_i) + w(t), \tag{2.7}$$

where L is the order of the AR process, $\{A_i\}$ are the AR coefficients, $\{\tau_i\}$ are delays, which are typically assumed to be known and equally spaced in time, and $w(t)$ is a random input process. This same type of model may be used for MIMO channels, where the scalar channel $h(t)$ is replaced by a vector and the scalar coefficients $\{A_i\}$ are replaced by matrix coefficients. As examples, AR models are used for SISO prediction in [53] and for MIMO prediction in [60, 61].

The results of SISO prediction using the above-mentioned models and methods all confirm the conclusion of [59], that channel knowledge over several wavelengths of motion is required to achieve acceptable prediction performance even for short intervals in the future. The MIMO prediction examples confirm the results of [60] and Chapter 4, that the use of multiple antennas leads to improved prediction abilities relative to SISO systems. However, the implementations of such predictors have, to this point, all fallen short of the performance possibilities suggested in [60] and this dissertation.

2.2.5 MIMO and Waterfilling

In a narrowband multiple-input multiple-output systems assuming additive white Gaussian noise (AWGN) with N_T antennas at the transmitter and N_R antennas at the receiver, the signal at the i-th antenna is given by the sum of the received signals from each transmit antenna plus noise, and it may be written as

$$y_i = \sum_{j=1}^{N_T} h_{ij} x_j + n_i, \tag{2.8}$$

where h_{ij} is the complex channel coefficient between the j-th transmit antenna and the i-th receive antenna, x_j is the signal transmitted by the j-th transmit antenna, and $n_i \sim \mathbb{C}\mathcal{N}(0, \sigma_n^2)$ is zero-mean AWGN with variance σ_n^2. Stacking the received signals at the N_R receive antennas into a vector of length N_R, the overall received signal model may be written in matrix/vector form as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2.9}$$

where the i,j-th element of $\mathbf{H}$ is given by h_{ij}, the i-th element of the column vector $\mathbf{y}$ is given by y_i, and the other vectors are defined in a like manner.

In information theory and communications, a popular measure indicating the maximum amount of information that may be transmitted over a communications channel is given by the channel capacity, which is defined as the maximum mutual information between the transmitted and received signals, i.e., the capacity C is given by

$$C = \max_{p(x)} I(\mathbf{X};\mathbf{Y}), \tag{2.10}$$

where the maximum is taken over all possible input distributions $p(x)$ [83]. For narrowband AWGN models such as (2.8), it is known that the maximizing input distribution is Gaussian. For the model of (2.9), it was further shown in [7,9] that given a fixed total transmit power P, the capacity may be realized when $\mathbf{x} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{V}\mathbf{Q}\mathbf{V}^H)$, where $\mathbf{V}$ is the matrix of eigenvectors from the eigen-decomposition $\mathbf{H}^H\mathbf{H} = V\mathbf{\Lambda}\mathbf{V}^H$ and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues such that $[\mathbf{\Lambda}]_{ii} = \lambda_i$. The matrix $\mathbf{Q}$ is a diagonal matrix whose entries, given by

$$[\mathbf{Q}]_{ii} = \left(\mu - \frac{\sigma_n^2}{\lambda_i}\right)^+ \tag{2.11}$$

with

$$\mu = \frac{P + \sum_{i=1}^{N_T} \frac{\sigma_n^2}{\lambda_i}}{N_T}, \tag{2.12}$$

are the powers allocated to the corresponding elements of $\mathbf{x}$. The MIMO-subchannel power allocation given by (2.11) is commonly referred to as waterfilling, the name coming from a physical analogy between the subchannel power allocation problem and the problem of determining water depth when filling a container having a multi-level base using a fixed amount of water. Note that the optimality of waterfilling power levels presupposes that the codebook of possible transmission vectors $\mathbf{x}$ is infinite and continuous, an assumption that may only be approximated in practice.

Considering the signal model of (2.9), the waterfilling solution may be realized through the used of the singular value decomposition (SVD). Let $\mathbf{H}$ have the SVD given by $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where the $\mathbf{V}$ defined here is the same as that defined in Section 2.2.5. The matrices

$\mathbf{U}$ and $\mathbf{V}$ are orthonormal and the elements of the diagonal diagonal matrix $\mathbf{\Sigma}$ are the square-roots of the values of λ_i which are used in the waterfilling power level determination of (2.11). Now note that if $\mathbf{x} = \mathbf{V}\mathbf{Q}^{\frac{1}{2}}\mathbf{s}$ with $\mathbf{s} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{I})$, then $\mathbf{x}$ meets the conditions given for the input signal by waterfilling, and an estimate of the information signal $\mathbf{s}$ may be obtained as

$$\hat{\mathbf{s}} = \mathbf{U}^H\mathbf{y} \tag{2.13}$$

$$= \mathbf{\Sigma}\mathbf{Q}^{\frac{1}{2}}\mathbf{s} + \mathbf{U}\mathbf{n}. \tag{2.14}$$

Note that this transmission scheme decomposes the MIMO channel into independent SISO subchannels, each of which may now be dealt with using SISO techniques.

2.2.6 Adaptive Modulation in MIMO

In the previous discussion, it was noted that, assuming a perfectly known channel (i.e., perfect channel state information (CSI)), the SVD-based waterfilling solution decomposes the MIMO channel into independent SISO subchannels. Based on the signal-to-noise ratio (SNR) of the resulting subchannels, bit-loading and coding schemes may be devised using SISO techniques to approach the available channel capacity. However, as was discussed in Chapter 1.1.3, CSI is never perfect in practice and practical symbol constellation designs do not meet the capacity-achieving input conditions assumed by the waterfilling solution. Thus, the power levels proscribed by the waterfilling solution are not guaranteed to be optimal with respect to the actual MIMO channel, and bit-loading levels will be chosen based on incorrect SNR information.

Given knowledge about the nature of the CSI imperfections and/or the symbol constellations to be used, it is often possible to find modified MIMO SVD-subchannel power and bit-loading levels that outperform those selected assuming perfect conditions. Such techniques, which attempt to adjust the power and bit-loading levels based upon symbol constellation and/or CSI imperfection information, are referred to as adaptive modulation techniques. Note that by this definition, the SVD-based signaling schemes discussed in Chapter 5 of this dissertation are considered adaptive modulation techniques.

A great many papers in literature consider the general problem of adaptive modulation, out of which we will only discuss a few most relevant to this dissertation. In [63,64,73], adaptive modulation techniques are are presented for SVD-based MIMO transmission assuming perfect CSI. In other words, these techniques focus solely on the problem of choosing the best bit-loading levels given a selected symbol constellation. The operation of these methods depends upon expressions which relate a performance criterion, such as bit error rate (BER) or symbol error rate (SER), to the MIMO-subchannel SNRs. In [63], the selection of this relation is left to the reader, though a particular bit-loading method is assumed for the simulations. In [64, 73], this relation, which is between BER for M-QAM with square constellations and SNR, is given by [84]

$$\text{BER} \approx 0.2 \exp\left[\frac{-1.5 \text{ SNR}}{M-1}\right]. \tag{2.15}$$

Given a constraint on BER, $\text{BER}_{\text{Target}}$, which is chosen to ensure successful transmission, bit-loading levels may be determined for the SVD subchannels by finding the maximum number of bits possible such that

$$b \leq \log_2\left[\frac{-1.5 \text{ SNR}}{\ln(5 \text{ BER}_{\text{Target}})} + 1\right], \tag{2.16}$$

where b, the number of bits, is related to M by $M = 2^b$ for square constellations.

The papers of [64, 73] also present adaptive modulation schemes assuming imperfect CSI. These methods are similar to those presented for the perfect CSI case except that the subchannel SNR expressions, which depended upon waterfilling subchannel power levels, must be replaced with with signal-to-interference-plus-noise ratio (SINR) expressions that quantify the effect of CSI imperfections on the SVD subchannels. The SINR expression used in [73] is equivalent to one of the SINR expressions derived in Chapter 5 for the special case of identical CSI at both the transmitter and receiver, referred to in the chapter as the common CSI (CCSI) case. The SINR expression used in [64] is also similar to this CCSI expression, though it is not as accurate.

Note that an expression similar to that of (2.16) is presented in Chapter 5 (see (5.57)) to relate the SER for M-QAM constellations and SNR. Its distinguishing feature (other than

using SER instead of BER) is that it permits the use of non-square M-QAM constellations, which allows a greater range of bit-loading levels to be considered. It also allows the derivations in Chapter 5 that are based upon it to be carried out more easily than would be possible using (2.16).

Also note that the adaptive modulation algorithm of [73], which is the best performing method from the papers discussed here, is used for the purpose of comparison in Chapter 5. The general details of this algorithm are presented in this chapter, so we do not include them here.

2.3 OFDM Signal Model

For the channel parameter estimation and pilot selection problem to be discussed in Chapter 3, consider the frequency-selective channel impulse response model given by

$$h(t) = \sum_{l=0}^{L-1} h_l \delta(t - lT_s - \tau_d), \tag{2.17}$$

where L is the number of discrete multipath components, which is assumed to be known, T_s is the system sampling period, h_l is the complex channel coefficient for the l-th path, and τ_d is the timing offset or first-path time delay. The discrete-time OFDM symbol $s(n)$, $n = 0, \ldots, N-1$, is constructed from N data and pilot symbols $\{b_p\}$ by means of the length-N IDFT as

$$s(n) = \frac{1}{\sqrt{N}} \sum_{p=-N/2+1}^{N/2} b_p \, e^{j\frac{2\pi}{N}np} \tag{2.18}$$

$$= \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} b_{m-N/2+1} \, e^{j\frac{2\pi}{N}(m-N/2+1)n}. \tag{2.19}$$

Note that we have indexed the OFDM subchannels with the zeroth frequency at the center of the frequency band, that is, $p \in [-N/2+1, -N/2+2, \ldots, 0, 1, \ldots, N/2]$, instead of the more conventional 0 to $N-1$ IDFT indexing to allow for a more traditional physical interpretation of the results. Transforming (2.19) into the continuous-time domain and adding a cyclic

prefix of duration T_G, we may represent the baseband transmitted signal as

$$s_{CT}(t) = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} b_{m-N/2+1}\ e^{j\frac{2\pi}{T}(m-N/2+1)t}, \quad -T_G \leq t \leq T, \tag{2.20}$$

where T is the OFDM symbol length, which is related to the sampling time T_s and the IDFT size N by $T = NT_s$.

Assuming that $\nu(t)$ is AWGN, the baseband received OFDM symbol after convolution with the channel is given by

$$y_{CT}(t) = \frac{1}{\sqrt{N}} \int s(\tau)h(t-\tau)\ d\tau + \nu(t) \tag{2.21}$$

$$= \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \sum_{L=0}^{L-1} b_{m-N/2+1} h_l e^{j\frac{2\pi}{T}(m-N/2+1)(t-\tau_d-lT_s)} + \nu(t). \tag{2.22}$$

For the purposes of Chapter 3, we restrict our focus to the zero inter-carrier and inter-symbol interference case. Therefore, we assume that the carrier frequency is perfectly synchronized at the transmitter and receiver and that the guard interval is larger than the delay spread, i.e., $(L-1)T_s+\tau_d < T_G$. With these assumptions, the sampled received signal after discarding the cyclic prefix is given by

$$\begin{aligned} y(k) &= y_{CT}(kT_s) \\ &= \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \sum_{L=0}^{L-1} h_l b_{m-N/2+1}\ e^{-j\frac{2\pi}{T}(m-N/2+1)\tau_d}\ e^{j\frac{2\pi}{N}(m-N/2+1)(k-l)} + \nu(kT_s). \end{aligned} \tag{2.23}$$

By taking the DFT of $x(n)$, we obtain the frequency-domain received signal

$$x(p) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n)\ e^{-j\frac{2\pi}{N}np} + n(p) \tag{2.24}$$

$$= \left(\sum_{l=0}^{L-1} h_l\ e^{-j\frac{2\pi}{N}pl} \right) b_p\ e^{-j\frac{2\pi}{T}p\tau_d} + n(p), \tag{2.25}$$

where $p \in [-N/2+1, -N/2+2, \ldots, 0, 1, \ldots, N/2]$ and $n(p)$ is the noise frequency response which, under the present assumptions, is zero-mean Gaussian distributed with variance σ_n^2.

In matrix/vector notation, this may be written as

$$\mathbf{x} = \mathbf{\Gamma B F}_L \mathbf{h} + \mathbf{n}, \tag{2.26}$$

where

$$\mathbf{x} = [\ x(-N/2+1) \quad \cdots \quad x(N/2)\ ]^T, \tag{2.27}$$

$$\mathbf{\Gamma} = \mathrm{diag}(w_T^{-(-N/2+1)\tau_d}, \ldots, w_T^{-(N/2)\tau_d}), \tag{2.28}$$

$$\mathbf{B} = \mathrm{diag}(b_{-N/2+1}, \ldots, b_{N/2}), \tag{2.29}$$

$$\mathbf{h} = [\ h_0 \quad \cdots \quad h_{L-1}\ ]^T, \tag{2.30}$$

and

$$\mathbf{n} = [\ n(-N/2+1) \quad \cdots \quad n(N/2)\ ]^T, \tag{2.31}$$

with $w_x^y = e^{-j\frac{2\pi}{x}y}$, and $\mathbf{F}_L$ composed of the first L columns of the zero-frequency-centered $N \times N$ DFT matrix. The model of (2.26) forms the basis of the development and discussions presented in Chapter 3.

2.4 Parametric MIMO Wideband Models

The MIMO wideband models considered in Chapter 4 are discrete-multipath or ray-based channel models, i.e., the models assume that that the signal at the receiver is a sum of a finite number of copies of the signal sent by the transmitter, each copy experiencing its own attenuation, delay, and Doppler. These models may be derived by extending the common narrowband time-varying SISO channel model [85]

$$h(t) = \sum_{l=1}^{L} \alpha_l e^{j\omega_{d,l}t}, \tag{2.32}$$

where α_l is the complex scattering coefficient of path l, $\omega_{d,l}$ is the Doppler frequency in rad/s of path l at time t, and L denotes the total number of paths traveled by the signal between the transmitter and the receiver. This model may be easily extended from a SISO to a MIMO system by including array responses for the tranmsitter and receiver. Let $\mathbf{a}_{t,l}$ and

$\mathbf{a}_{r,l}$ be array responses for the transmitter and receiver, respectively. Then the narrowband MIMO channel model may be written as

$$\mathbf{H}(t) = \sum_{l=1}^{L} \alpha_l \mathbf{a}_{r,l} \mathbf{a}_{t,l}^T e^{j\omega_{d,l} t}, \tag{2.33}$$

where $\mathbf{a}_{t,l}$ is a vector of length M_t, the number of tranmsit antennas, and $\mathbf{a}_{r,l}$ is a vector of length M_r, the number of receive antennas. This model has been used in several channel measurement campaigns [86–88] and forms the basis for the derivations and analysis of [60]. The narrowband model of (2.33) may be extended further to accommodate the description of wideband channels as

$$\mathbf{H}(\omega, t) = \sum_{l=1}^{L} \alpha_l \mathbf{a}_{r,l} \mathbf{a}_{t,l}^T e^{j((\omega_c - \omega)\tau_l - \omega_{d,l} t)}, \tag{2.34}$$

where ω is the frequency variable in rad/s, ω_c is the center or reference frequency of the band of interest, and τ_l is the delay in seconds associated with path l. While we have presented this model as an extension of (2.33), it may also be derived directly as a time-varying multipath channel kernel [89]. One of the advantages of this model is that the channel is explicitly defined for every time and frequency. Thus, it is directly applicable to the MIMO-OFDM problem where information symbols are transmitted at particular times and frequencies.

The developments in Chapter 4 use several specific channel models based upon the general model of (2.34). The principal differences of these individual models arise from assumptions made about the array response term $\mathbf{a}_{r,l} \mathbf{a}_{t,l}^T$.

2.4.1 Structured DOD/DOA Model

For the first models used in Chapter 4, the array response vectors are assumed to be structured functions of the direction of departure (DOD) and direction of arrival (DOA) of the signal paths relative to the transmit and receive antenna arrays, respectively. The first of these models assumes that the antenna configurations are known and are precisely calibrated. It may also be modified to form a new model for situations in which there is a small level of uncertainty concerning the exact antenna configurations by adding a perturbation term to

each array response vector so that the overall response term for the l-th signal path becomes $(\mathbf{a}_{r,l} + \mathbf{v}_{r,l})(\mathbf{a}_{t,l} + \mathbf{v}_{t,l})^T$, where $\mathbf{v}_{t,l}$ and $\mathbf{v}_{r,l}$ are the array perturbation terms for the transmit and receive arrays, respectively. The specifics of these structured models will be discussed in more detail in the context of Chapter 4.

2.4.2 Unstructured Vector Model

The structured DOD/DOA model assumes specific array configurations that depend on the signal path DODs and DOAs. The estimation of DOD and DOA parameters can be difficult and the results may be very sensitive to calibration errors. To avoid such problems, a more general model may be applied to the channel parameter estimation problem. Note that in the context of Chapter 4, this model is not assumed to be a ground truth representation of the actual channel, but is instead used as an approximation of the structured DOD/DOA model for the purposes of simplifying channel estimation. In this unstructured vector model, the path gains and the angle- and position-dependent array responses of the DOD/DOA model are replaced by unstructured vectors, termed *spatial signatures*. That is, the vectors $\mathbf{a}_{t,l}$ and $\mathbf{a}_{r,l}$ for this model are not assumed to be explicit functions of DOD or DOA, but instead abstractly represent the transmit and receive array responses for path l with delay τ_l and Doppler $\omega_{d,l}$. This model will be referred to Chapter 4 as the vector spatial signature (VSS) model, and will be presented in more detail prior to its use there.

2.4.3 Unstructured Matrix Model

The model of (2.34) may be abstracted one step further by replacing the vector response vectors of $\mathbf{a}_{t,l}$ and $\mathbf{a}_{r,l}$ of path l by a matrix spatial signature $\mathbf{A}_l$. As in the VSS model, the matrices $\mathbf{A}_l$ are not assumed to be explicit functions of DOD or DOA, but instead abstractly represent the channel characteristics for a particular path with associated delay τ_l and Doppler $\omega_{d,l}$. Also like the VSS, this model is not assumed to be a ground truth representation of the actual channel, but is instead used as an approximation of the DOD/DOA model. Since it is parameterized in a different manner, we expect the performance of channel estimation based upon this model to differ from that achieved when using the DOD/DOA model. One significant difference is that since no restriction is placed on the

elements of the matrices $\mathbf{A}_l$, they may be of arbitrary rank, in contrast to the rank-one matrices formed by the outer products $\mathbf{a}_{r,l}\mathbf{a}_{t,l}^T$. The consequences of this fact will be reflected in the system performance. Note that this matrix spatial signature (MSS) model is essentially a two-dimensional filter in ω and t with matrix filter taps. At a fixed time t, it is similar to the multi-channel finite impulse response (FIR) channel models frequently used in the literature, although the taps are not evenly spaced in time.

2.5 MIMO Perturbation Model

In MIMO communication systems, accurate knowledge of the communication channel, referred to as channel state information (CSI), is a vital factor enabling good system performance. In many systems requiring CSI at both the transmitter and receiver of a communications link, CSI is obtained through channel estimation at the receiver with the aid of training data sent from the transmitter, and that information is then periodically communicated back to the transmitter. These estimates are typically imperfect, with both the estimation and feedback processes contributing to CSI imperfections. In Chapter 5 of this dissertation, we investigate the impact of such imperfections on MIMO system performance. For the purposes of Chapter 5, we mathematically describe this imperfect CSI by letting $\mathbf{H}$ be the channel as known at the transmitter, $\mathbf{H} + \Delta\mathbf{H}_1$ the channel as known at the receiver, and $\mathbf{H} + \Delta\mathbf{H}_1 + \Delta\mathbf{H}_2$ the actual MIMO channel, where the preceding matrices all have dimension $N_r \times N_t$, the number of receive and transmit antennas, respectively. This particular formulation allows the expressions derived in Chapter 5 to depend only upon the known CSI $\mathbf{H}$ and the assumed statistics for the CSI imperfections, instead of requiring more structured knowledge of the CSI imperfections. In the context of the above CSI notation, the matrix $\Delta\mathbf{H}_1$ represents the discrepancies between the actual channel and the transmit CSI that are unique to the transmitter alone, the source of which may include quantization effects and errors in the feedback channel, as well as variations in the MIMO channel when assuming that the receiver updates its CSI more frequently than the transmitter. Channel variations considered in this context may occur, for example, if the transmitter or receiver moves during the interval between channel estimation and channel feedback. On the other hand, the matrix $\Delta\mathbf{H}_2$ represents CSI errors that are common to both the transmitter and

Table 2.1: Channel State Information Relationships

Transmitter	$\mathbf{H}$
Receiver	$\mathbf{H} + \Delta\mathbf{H}_1$
Actual Channel	$\mathbf{H} + \Delta\mathbf{H}_1 + \Delta\mathbf{H}_2$

receiver. Such errors are typically introduced at the receiver and are then fed back to the transmitter. Factors contributing to this term include estimation errors and outdated CSI present at the receiver. The CSI setup is summarized in Table 2.1. Note that when $\Delta\mathbf{H}_1 = 0$, the transmitter and receiver possess the same CSI.

Chapter 3

Optimal Pilot Design for Joint Time Delay and Channel Estimation in OFDM

3.1 Introduction

In this chapter, we examine the problem of optimal subchannel pilot design and placement for joint channel and timing offset estimation in OFDM systems. In particular, we consider the problem of pilot design in order to minimize a lower bound on the channel estimation error variance while meeting a performance constraint on the variance of the timing offset estimate. In order to address this problem, we consider two Cramér-Rao bound (CRB) formulations. The first of these is found directly by applying the well-known Bangs formula [40]. Our optimization problem is then cast as the minimization of a weighted trace of this CRB, which allows us to control the relative importance of the timing offset and channel estimation problems when designing pilots. While this minimization may not be solved for the pilots in closed form, the cost function, which may be solved numerically, provides valuable insights into the structure of optimal solutions. Following this presentation, a second CRB is found indirectly using a CRB as a function of parameters (CRB-FOP) formulation [41]. This derivation, which reduces the coupling between timing-offset- and channel-parameter-related regions of the CRB, is particularly suited to our needs, allowing the problem to be separated into two simpler components.

The remainder of this chapter will proceed as follows. In Section 3.2, the OFDM signal model is presented. In Section 3.3, the joint CRB derivation is then shown. Using these results, the problem of optimal pilot selection is presented in Section 3.4. The ensuing discussion provides motivation for CRB-FOP derivation, which is carried out in Section 3.5. Finally, simulation examples are presented in Section 3.6, followed by the concluding remarks in Section 3.7.

3.2 Signal Model

Consider the OFDM signaling model of (2.25) derived in Chapter 2.3, which is reproduced here:

$$x(p) = \left(\sum_{l=0}^{L-1} h_l \, e^{-j\frac{2\pi}{N}pl} \right) b_p \, e^{-j\frac{2\pi}{T}p\tau_d} + n(p), \tag{3.1}$$

where L is the number of discrete multipath components, which is assumed to be known, T_s is the system sampling period, h_l is the complex channel coefficient for the l-th path, and τ_d is the timing offset. Since this timing offset is equivalent to the primary channel time delay, i.e., the time delay of the first path of the impulse response, we will refer to it simply as the time delay or delay in the remainder of this chapter. We assume that the channel coefficients $\{h_l\}$ and the time delay τ_d are unknown and must be estimated through the use of pilot tones transmitted as part of an OFDM symbol $s(n)$, $n = 0, \ldots, N-1$, which is constructed from N data and pilot symbols $\{b_p\}$ by means of the length-N inverse discrete Fourier transform (IDFT). The additive noise $\{n(p)\}$ is assumed to be independent, identically-distributed zero-mean complex Gaussian noise so that $n(p) \sim \mathbb{C}\mathcal{N}(0, \sigma_n^2)$. Recall that, for the purposes of this work, we have restricted our focus to the zero inter-carrier and inter-symbol interference case. Therefore, we assume the carrier frequency is perfectly synchronized at the transmitter and receiver and that the guard interval is larger than the delay spread, i.e., $(L-1)T_s + \tau_d < T_G$. Equivalently, we may assume that a good estimate of the carrier frequency offset and a course estimate of the delay spread have been obtained previously by means of correlations in the cyclic prefix, a preamble training sequence, or another related method.

As previously shown in Chapter 2.3, the model shown in (3.1) may be written in matrix/vector notation as

$$\mathbf{x} = \mathbf{\Gamma B F}_L \mathbf{h} + \mathbf{n}, \tag{3.2}$$

where

$$\mathbf{x} = [\ x(-N/2+1) \quad \cdots \quad x(N/2)\ ]^T, \tag{3.3}$$

$$\mathbf{\Gamma} = \mathrm{diag}(w_T^{-(-N/2+1)\tau_d}, \ldots, w_T^{-(N/2)\tau_d}), \tag{3.4}$$

$$\mathbf{B} = \mathrm{diag}(b_{-N/2+1}, \ldots, b_{N/2}), \tag{3.5}$$

$$\mathbf{h} = [\ h_0 \quad \cdots \quad h_{L-1}\ ]^T, \tag{3.6}$$

and

$$\mathbf{n} = [\ n(-N/2+1) \quad \cdots \quad n(N/2)\ ]^T, \tag{3.7}$$

with $w_x^y = e^{-j\frac{2\pi}{x}y}$ and $\mathbf{F}_L$ composed of the first L columns of the zero-frequency-centered $N \times N$ DFT matrix.

3.3 The CRB

Since the overall goal of this chapter is to find optimal pilot configurations for the joint estimation of the time delay τ_d and the channel impulse response $\mathbf{h}$, we require a way to quantify the strengths of particular pilot designs in terms of system performance. To this end, we now seek to find the CRB with respect to τ_d and $\mathbf{h}$ given the OFDM signal model of (3.2). Let the parameter vector $\boldsymbol{\Theta}$ be defined as

$$\boldsymbol{\Theta} \triangleq [\ \tau_d \quad \mathrm{Re}[\mathbf{h}^T] \quad \mathrm{Im}[\mathbf{h}^T]\ ]^T, \tag{3.8}$$

where $\mathrm{Re}[a]$ and $\mathrm{Im}[a]$ are the real and imaginary parts, respectively, of the complex variable a. Then the CRB with respect to $\boldsymbol{\Theta}$, can be calculated using Bangs formula [40]

$$\begin{aligned}\left[(\mathrm{CRB})^{-1}\right]_{ij} &= \mathrm{Tr}\left[\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \theta_i}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \theta_j}\right] + 2\mathrm{Re}\left[\frac{\partial \mathbf{x}^H}{\partial \theta_i}\mathbf{C}^{-1}\frac{\partial \mathbf{x}}{\partial \theta_j}\right] \\ &= \mathbf{J}(\boldsymbol{\Theta})_{ij},\end{aligned} \tag{3.9}$$

where $\mathbf{J}(\boldsymbol{\Theta})$ is the Fisher information matrix (FIM) and we assume $\mathbf{C} = E[\mathbf{n}\mathbf{n}^H] = \sigma_n^2\mathbf{I}$ for this problem. Applying Bangs formula to our signaling model (3.2), the FIM is found to be

$$J(\boldsymbol{\Theta}) = \frac{2}{\sigma_n^2}\begin{bmatrix} J_{\tau_d} & \mathbf{J}_{21}^T \\ \mathbf{J}_{21} & \mathbf{J}_{\mathbf{h}} \end{bmatrix} \tag{3.10}$$

with

$$J_{\tau_d} = \mathbf{h}^H\mathbf{F}_L^H\mathbf{B}^H\mathbf{D}^2\mathbf{B}\mathbf{F}_L\mathbf{h}, \tag{3.11}$$

$$\mathbf{J}_{21} = \begin{bmatrix} \text{Im}[\mathbf{F}_L^H \mathbf{B}^H \mathbf{D} \mathbf{B} \mathbf{F}_L \mathbf{h}] \\ -\text{Re}[\mathbf{F}_L^H \mathbf{B}^H \mathbf{D} \mathbf{B} \mathbf{F}_L \mathbf{h}] \end{bmatrix}, \tag{3.12}$$

$$\mathbf{J}_{\mathbf{h}} = \begin{bmatrix} \text{Re}[\mathbf{F}_L^H \mathbf{B}^H \mathbf{B} \mathbf{F}_H] & -\text{Im}[\mathbf{F}_L^H \mathbf{B}^H \mathbf{B} \mathbf{F}_H] \\ \text{Im}[\mathbf{F}_L^H \mathbf{B}^H \mathbf{B} \mathbf{F}_H] & \text{Re}[\mathbf{F}_L^H \mathbf{B}^H \mathbf{B} \mathbf{F}_H] \end{bmatrix}, \tag{3.13}$$

and

$$\mathbf{D} = \frac{2\pi}{T_s} \begin{bmatrix} -\frac{N}{2}+1 & & \\ & \ddots & \\ & & \frac{N}{2} \end{bmatrix}. \tag{3.14}$$

In order to invert the FIM, we need to use two inversion properties. First, given a nonsingular complex-valued matrix $\mathbf{A}$, we have the following easily-verifiable inversion relationship [90]:

$$\begin{bmatrix} \text{Re}[\mathbf{A}] & -\text{Im}[\mathbf{A}] \\ \text{Im}[\mathbf{A}] & \text{Re}[\mathbf{A}] \end{bmatrix}^{-1} = \begin{bmatrix} \text{Re}[\mathbf{A}^{-1}] & -\text{Im}[\mathbf{A}^{-1}] \\ \text{Im}[\mathbf{A}^{-1}] & \text{Re}[\mathbf{A}^{-1}] \end{bmatrix}. \tag{3.15}$$

Second, given a nonsingular matrix, the inverse may be found as a function of partitions of the matrix as

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{X}^{-1} & -\mathbf{X}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{X}^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{X}^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}, \tag{3.16}$$

where $\mathbf{X} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ is the Schur complement of $\mathbf{D}$ [79]. Applying (3.16) to (3.13), we find

$$\mathbf{J}_{\mathbf{h}}^{-1} = \begin{bmatrix} \text{Re}[\mathbf{A}^{-1}] & -\text{Im}[\mathbf{A}^{-1}] \\ \text{Im}[\mathbf{A}^{-1}] & \text{Re}[\mathbf{A}^{-1}] \end{bmatrix}. \tag{3.17}$$

Note $\mathbf{J}_{\mathbf{h}}^{-1}$ is the CRB with respect to $\mathbf{h}$ assuming τ_d is known. We denote this matrix as $\mathbf{CRB}_{\mathbf{h}}$ so that $\mathbf{CRB}_{\mathbf{h}} = \mathbf{J}_{\mathbf{h}}^{-1}$. Likewise, $J_{\tau_d}^{-1}$ is the CRB with respect to τ_d assuming $\mathbf{h}$ is known, i.e, $\text{CRB}_{\tau_d} = J_{\tau_d}^{-1}$. Having obtained $\mathbf{J}_{\mathbf{h}}^{-1}$, the inverse of (3.10) may now be found by

applying (3.16) with the result

$$\mathbf{CRB}_{\tau_d,\mathbf{h}} = \frac{\sigma_n^2}{2}\begin{bmatrix} \gamma_{\tau_d}^{-1} & \mathbf{CRB}_{21}^T \\ \mathbf{CRB}_{21} & \mathbf{CRB}_{22} \end{bmatrix}, \tag{3.18}$$

where

$$\gamma_{\tau_d} = \mathbf{h}^H\mathbf{F}_L^H\mathbf{B}^H\mathbf{D}\mathbf{\Pi}^{\perp}_{\mathbf{BF}_L}\mathbf{DBF}_L\mathbf{h}, \tag{3.19}$$

$$\mathbf{\Pi}^{\perp}_{\mathbf{BF}_L} = \mathbf{I} - \mathbf{BF}_L(\mathbf{F}_L^H\mathbf{B}^H\mathbf{BF}_L)^{-1}\mathbf{F}_L^H\mathbf{B}^H, \tag{3.20}$$

$$\mathbf{CRB}_{21} = \begin{bmatrix} -\gamma_{\tau_d}^{-1}\mathrm{Im}[\mathbf{q}] \\ \gamma_{\tau_d}^{-1}\mathrm{Re}[\mathbf{q}] \end{bmatrix}, \tag{3.21}$$

and

$$\mathbf{CRB}_{22} = \begin{bmatrix} \mathrm{Re}[\mathbf{Q}^{-1}] + \gamma_{\tau_d}^{-1}\mathrm{Im}[\mathbf{q}]\mathrm{Im}[\mathbf{q}^T] & -\mathrm{Im}[\mathbf{Q}^{-1}] - \gamma_{\tau_d}^{-1}\mathrm{Im}[\mathbf{q}]\mathrm{Re}[\mathbf{q}^T] \\ \mathrm{Im}[\mathbf{Q}^{-1}] - \gamma_{\tau_d}^{-1}\mathrm{Re}[\mathbf{q}]\mathrm{Im}[\mathbf{q}^T] & \mathrm{Re}[\mathbf{Q}^{-1}] + \gamma_{\tau_d}^{-1}\mathrm{Re}[\mathbf{q}]\mathrm{Re}[\mathbf{q}^T] \end{bmatrix} \tag{3.22}$$

with

$$\mathbf{Q} = \mathbf{F}_L^H\mathbf{B}^H\mathbf{BF}_L \tag{3.23}$$

and

$$\mathbf{q} = \mathbf{A}^{-1}\mathbf{F}_L^H\mathbf{B}^H\mathbf{DBF}_L\mathbf{h}. \tag{3.24}$$

Comparing the individual CRB expressions for CRB_{τ_d} and (3.17) with their related subblock expressions (3.19) and (3.22) of (3.18), we note that the expressions comprising (3.18) contain the uncoupled CRB expressions of (3.11) and (3.17), but each subblock also includes a penalty term as a consequence of the coupling between the estimation of τ_d and $\mathbf{h}$.

The CRB of (3.18) depends on particular instances of the channel response $\mathbf{h}$. While this is useful in understanding system performance given particular instantaneous channel realizations, it may also be useful to have a CRB formulation that does not depend on the particular channel realization. Such a CRB formulation may be obtained though the use of the expected value operation, i.e., $E[\mathbf{CRB}_{\tau_d,\mathbf{h}}]$, where the expectation is over the channel realizations $\mathbf{h}$. Unfortunately, $E[\mathbf{CRB}_{\tau_d,\mathbf{h}}]$ (specifically $E[\mathbf{CRB}_{22}]$) cannot, in general, be calculated in closed form. One popular way to address this type of problem has been to

use the modified CRB (MCRB) [29, 80], which, like the CRB, lower bounds the minimum unbiased estimation error variance. The MCRB accomplishes this feat by performing the expectation relative the nuisance parameters, in our case, $\mathbf{h}$, on the FIM prior to matrix inversion. However, for our problem, the MCRB leads to the uncoupling of the CRB expressions relative to the estimation error variance of τ_d and $\mathbf{h}$, which is not acceptable for our purposes. Note that the study of the $\mathbf{h}$-specific expressions is still warranted since the features of the optimal pilot designs developed from these expressions are widely applicable.

3.4 Optimal Pilot Selection

We now address the issue of optimal pilot design. Specifically, given a fixed total transmit power P_T, we wish to find diagonal entries for the matrix $\mathbf{B}$, denoted by the vector $\mathbf{b}$, which optimize a weighted sum of the diagonal entries of the CRB, i.e.,

$$\mathbf{b}^* = \arg\min_{\mathbf{b}\in\mathcal{B}} \mathbf{G}(\alpha, \mathbf{b}), \tag{3.25}$$

where the cost function $\mathbf{G}(\alpha, \mathbf{b})$ is given by

$$\mathbf{G}(\alpha, \mathbf{b}) = \left[\alpha \left(\frac{2\pi N}{T_s}\right)^2 \gamma_{\tau_d}^{-1} + (1-\alpha)\left(\frac{1}{L}\right) \mathrm{tr}(\mathbf{CRB}_{22})\right]. \tag{3.26}$$

The coefficients $2\pi N/T_s$ and $1/L$ in $\mathbf{G}(\alpha, \mathbf{h})$ serve to normalize the τ_d and $\mathbf{h}$ portions of the CRB matrix while the weighting parameter $\alpha \in [0, 1]$ may be used to control the relative importance attached to the estimation performance for τ_d and $\mathbf{h}$ when selecting $\mathbf{b}^*$.

Note that in the optimization statement of (3.25), we have designated that the vector $\mathbf{b}^*$ be chosen from some set of possible pilots $\mathcal{B}$. While the possible sets that may reasonably be selected for $\mathcal{B}$ is quite large, we will restrict our attention in this work to two possible sets from which the optimal pilot vector may be selected: $\mathcal{B}_{DP}$, the set of all combinations of P discrete and equipowered pilots (the remaining $N - P$ elements of $\mathbf{b}$ are set to zero), and $\mathcal{B}_{CP} = \mathbb{R}^N$, the set of all real valued vectors of length N. By examining the expressions for the entries of (3.10), it may be noted that no generality is lost by considering only pilots in $\mathbb{R}^N$ as opposed to $\mathbb{C}_N$, the set all complex vectors of length N, since only the magnitude

of $\mathbf{b}$ influences the FIM expression. We will refer to the optimization over $\mathcal{B}_{DP}$ as the discrete pilot set (DPS) scenario and the optimization over $\mathcal{B}_{CP} = \mathbb{R}^N$ as the continuous pilot distribution (CPD) scenario. These two scenarios are probably the most commonly considered in OFDM pilot design when considering OFDM symbols individually.

3.4.1 Uncoupled CRB Cases

With these two scenarios in mind, we may now discuss the optimization problem of (3.25) in the context of these strategies starting with the limiting uncoupled CRBs $J_{\tau_d}^{-1}$ and $\mathbf{CRB_h}$, for which it is assumed that $\mathbf{h}$ or τ_d are known, respectively. First, reconsider the expression for J_{τ_d}, the FIM with respect to τ_d assuming $\mathbf{h}$ is known:

$$J_{\tau_d} = \mathbf{h}^H \mathbf{F}_L^H \mathbf{B}^H \mathbf{D}^2 \mathbf{B} \mathbf{F}_L \mathbf{h} \tag{3.27}$$

$$= \|\mathbf{DBg}\|^2, \tag{3.28}$$

where $\mathbf{g} = \mathbf{F}_L \mathbf{h}$ is the discrete frequency response of the channel impulse response $\mathbf{h}$. For the DPS case, the optimal P equipowered pilot locations are chosen as the set corresponding to the locations of the largest P products $[\mathbf{D}]_{ii}^2 |g_i|^2$, locations which are generally near the band edges where the channel frequency response magnitude is large. For the CPD case, all of the pilot power is given to a single pilot tone whose location is given by

$$i^* = \arg \max_{i \in [-N/2+1, N/2]} [\mathbf{D}]_{ii}^2 |g_i|^2. \tag{3.29}$$

Note that if $P = 1$ for the DPS case, then the DPS and CPD solutions are identical.

Now consider the case of choosing the pilots to maximize the trace of the CRB relative to $\mathbf{h}$ assuming τ_d is known. This trace may be written as

$$\mathrm{tr}[\mathbf{CRB_h}] = \mathrm{tr}\left[(\mathbf{F}_L^H \mathbf{B}^H \mathbf{B} \mathbf{F}_L)^{-1} \right]. \tag{3.30}$$

The solution in this case was found for the CPD in [19,20,23], though not under the guise of a CPD optimization. In these works, it was found that (3.30) was minimized when the pilots are equipowered and are spaced evenly over the the frequency bins, where the meaning

of evenly-spaced in terms of the pilot locations is defined in [20] with the pilot locations specifically given by $n_p = p_0 + p_V$, $\forall \phi \in \{-L+1, \ldots, L+1\} \backslash \{0\}$ where $V \in \mathbb{Z}$ such that $PV\phi/N \in \mathbb{Z}$ and $V\phi/N \notin \mathbb{Z}$ and $p_0 + N/2 - 1 \in \{0, \ldots, V-1\}$. Obviously, there are several pilot configurations that satisfy the equispace requirement and thus minimize (3.30), including spreading the pilot power evenly over all of the frequency bins, though [19] does show that capacity-wise, it is advantageous to use the minimum number of required pilots. As long as $P \geq L$ is chosen to satisfy the above condition, the DPS case will correspond to a possible solution of the CPD case. Otherwise, the optimal DPS solution will be close to equispaced and will not correspond to a CPD solution.

3.4.2 Coupled CRB Case

Now let us consider the optimization problem of (3.25) involving the CRB subblocks $\gamma_{\tau_d}^{-1}$ and $\mathbf{CRB}_{22}$ from the joint CRB of (3.18). Unfortunately, this optimization cannot be solved for the optimal pilots in closed form, even for the cases of $\alpha = 0$ and $\alpha = 1$ which focus the optimization fully on the estimation performance of $\mathbf{h}$ or τ_d, respectively. Fortunately, the optimization may be solved numerically. Since this is a signal design problem, it is typically done offline and we are justified in considering techniques that may require a moderate level of computational complexity.

The optimization for the DPS involves a simple but computationally-intensive search over the possible pilot combinations. Without any prior knowledge of the likely solution form, finding $\mathbf{b}^*$ requires that all *N-choose-P* possible pilots be investigated, a number which increases rapidly with increases in both N and $P < N/2$. However, with prior knowledge of the general structure of the solution, good solutions may be obtained over a significantly reduced search space (e.g., see Section 3.6).

For the CPD, solutions may be found using numerical optimization algorithms, e.g., Matlab's unconstrained optimization routine *fminunc*, which uses a linesearch algorithm when applied to this type of problem, may be used. The computational complexity for even moderate values of N is much lower than that for the comparable DPS solution, making CPD using numerical optimization an appealing route to a solution. It should be noted that the resulting solutions to (3.25) may not necessarily be unique, and they are not guaranteed

to be globally optimal since the cost function $\mathbf{G}(\alpha, \mathbf{b})$ is not convex in $\mathbf{b}$. However, we have found this type of optimization generally yields good solutions with the CPD solution always equaling or outperforming the DPS solution, i.e, $\mathbf{G}(\alpha, \mathbf{b}^*_{CPD}) \leq \mathbf{G}(\alpha, \mathbf{b}^*_{DPS})$ in all of the simulations we have carried out. This suggests at least local convexity around local minima, which correspond to good solutions.

While we cannot, in general, find closed form solutions to the optimization of (3.25), we can gain some insights into the general structure of the solution by examining the constituent components of the cost function $\mathbf{G}(\alpha, \mathbf{b})$, which we will do for the CPD, though similar inferences may be made for the DPS. First, consider the minimization of $\gamma_{\tau_d}^{-1}$, which is equivalent to minimization of $\mathbf{G}(1, \mathbf{b})$ and to the maximization γ_{τ_d}. We may rewrite the expression for γ_{τ_d} in (3.19) as

$$\gamma_{\tau_d} = \mathbf{h}^H \mathbf{F}_L^H \mathbf{B}^H \mathbf{D} \mathbf{\Pi}^{\perp}_{\mathbf{BF}_L} \mathbf{D} \mathbf{B} \mathbf{F}_L \mathbf{h} \tag{3.31}$$

$$= \mathbf{h}^H \mathbf{F}_L^H \mathbf{B}^H \mathbf{D}^2 \mathbf{B} \mathbf{F}_L \mathbf{h} - \mathbf{h}^H \mathbf{F}_L^H \mathbf{B}^H \mathbf{D} \mathbf{\Pi}_{\mathbf{BF}_L} \mathbf{D} \mathbf{B} \mathbf{F}_L \mathbf{h} \tag{3.32}$$

$$= \left\| \mathbf{D} \mathbf{B} \mathbf{F}_L \mathbf{h} \right\|^2 - \left\| \mathbf{\Pi}_{\mathbf{BF}_L} \mathbf{D} \mathbf{B} \mathbf{F}_L \mathbf{h} \right\|^2, \tag{3.33}$$

where

$$\mathbf{\Pi}_{\mathbf{BF}_L} = \mathbf{B} \mathbf{F}_L (\mathbf{F}_L^H \mathbf{B}^H \mathbf{B} \mathbf{F}_L)^{-1} \mathbf{F}_L^H \mathbf{B}^H. \tag{3.34}$$

Note that $\mathbf{\Pi}_{\mathbf{BF}_L}$ (and $\mathbf{\Pi}^{\perp}_{\mathbf{BF}_L}$) is projection matrix so that $\mathbf{\Pi}_{\mathbf{BF}_L} \mathbf{\Pi}_{\mathbf{BF}_L} = \mathbf{\Pi}_{\mathbf{BF}_L}$, a property used to obtain the equality of (3.33). Observe also that $\gamma_{\tau_d} \leq \left\| \mathbf{D} \mathbf{B} \mathbf{F}_L \mathbf{h} \right\|^2$, the FIM matrix for the uncoupled case. As was mentioned previously, that FIM was maximized by allocating all of the pilot power to a single element of $\mathbf{b}$ corresponding to a point near the band edge where the channel frequency response is large. This solution is modified in the coupled expression by the penalty term $\left\| \mathbf{\Pi}_{\mathbf{BF}_L} \mathbf{D} \mathbf{B} \mathbf{F}_L \mathbf{h} \right\|^2$. While the uncoupled term favors allocating all of the available power at a single or possibly a few points near the band edges, the inverse in (3.34) requires that at least L elements of $\mathbf{b}$ are nonzero and that they have a fairly even spacing. The result is a combination of the two. Typically, the solution involves $L + 1$ discrete pilot clusters, the number of unknowns to be estimated, with the outermost pilots shifted closer to the band edges and the remaining plots spread across the band. The term pilot cluster

refers to two consecutive elements of $\mathbf{b}$ given allocated power, isolated from other powered elements, which essentially act as a single pilot. It is likely that if power allocation were possible over a continuous range of frequencies, these pilot clusters would then be realized as single pilot tones at frequencies between the discrete frequency possibilities allowed here. The pilots away from the band edges are allocated a relatively small amount of power when compared to those at the band edges, and that power progressively reduces towards the band center. An example of this pilot allocation will be shown in the simulations.

Now consider the term involved in the minimization of $\mathbf{G}(0, \mathbf{b})$, which is equivalent to the minimization of $\mathrm{tr}[\mathbf{CRB}_{22}]$. This term may be written as

$$\mathrm{tr}[\mathbf{CRB}_{22}] = \mathrm{tr}[\mathbf{Q}]^{-1} + \gamma_{\tau_d}^{-1} \mathrm{tr}[\mathbf{q}\mathbf{q}^H] \tag{3.35}$$

$$= \mathrm{tr}\left[(\mathbf{F}_L^H \mathbf{B}^H \mathbf{B}\mathbf{F}_L)^{-1}\right] + \frac{\left\|(\mathbf{B}\mathbf{F}_L)^\dagger \mathbf{D}\mathbf{B}\mathbf{F}_L\mathbf{h}\right\|^2}{\left\|\mathbf{\Pi}^\perp_{\mathbf{B}\mathbf{F}_L}\mathbf{D}\mathbf{B}\mathbf{F}_L\mathbf{h}\right\|^2}. \tag{3.36}$$

The first part of this expression corresponds to the uncoupled case and, as discussed previously, is minimized by equispaced and equipowered pilots. The second term in (3.36), then, is a penalty term resulting from the coupling between the estimation of τ_d and $\mathbf{h}$. Considered alone, the numerator of this penalty term, $\left\|(\mathbf{B}\mathbf{F}_L)^\dagger \mathbf{D}\mathbf{B}\mathbf{F}_L\mathbf{h}\right\|^2$, is minimized by placing all of the available pilot power in the center of the frequency band at the zeroth frequency or at a combination of this and locations at which the channel frequency response has a zero. When considered in combination with the denominator, which we examined earlier for the $\alpha = 1$ case, the result requires L or more pilots ($L+1$ pilot clusters is typical), spread across the frequency band. However, in this case the non-band-edge pilots receive the largest allocation of power, with the power decreasing towards the band edges, and the pilots tend to cluster more towards the center of the band, though still relatively well spaced. In the full expression of (3.36), the uncoupled term tends to be the dominant term. Thus, the resulting pilot allocation is relatively equipowered and equispaced. However, the penalty term alters this solution, dictating a low number of discrete pilot clusters (usually $L+1$) with outermost pilots tending towards the band edges and inner pilots shifting slightly towards locations of small channel frequency response. Also, a greater fraction of the power is allocated towards the center of the frequency band.

For values of $\alpha \in (0,1)$, resulting pilot allocations combine the features of the $\alpha = 0$ and $\alpha = 1$ limiting cases. Thus, as α moves from 1 to 0, the optimal pilots move from roughly equipowered and equispaced to a configuration of pilots with most of the power at the band edges. The selection of α will depend on desired or required constraints. We also note that $\mathbf{G}(\alpha, \mathbf{b})$ varies linearly with $\frac{\sigma_n^2}{P_T}$. This implies that for a particular α, the structure of $\mathbf{b}^*$ is not a function of the total pilot power or measurement noise level, only its norm is.

In this chapter, we are interested in finding the pilot configuration which minimizes the variance on the estimation of the channel impulse response $\mathbf{h}$ while constraining the error variance of the time delay estimate τ_d to be at or below some threshold, assuming a fixed total pilot power P_T. Because of the coupling of the τ_d and $\mathbf{h}$ terms in the CRB, we cannot design the pilots independently. But assuming P_T is sufficient so that the constraint on the error variance on τ_d may be achieved, this problem may be solved by searching for the α that results in a value for $\frac{\sigma_n^2}{2}\gamma_{\tau_d}^{-1}$ which meets the required constraint. The resulting pilot configuration automatically minimizes the error variance on the estimate of $\mathbf{h}$ by design.

3.5 Pilot Selection using the CRB-FOP

Once again, consider the problem of finding the pilot configuration which minimizes the channel impulse response estimation error variance while meeting a constraint on the error variance of the time delay estimate. When addressed in terms of the expression of (3.25), the search for the solution, while feasible, may be intimidating since we must search over α, solving for a different pilot configuration at each α, until we find the solution that satisfies our conditions. It would be desirable, therefore, to have a method for solving this problem that did not necessitate such a search. In this section, we derive another CRB which allows us to address the above design problem directly without the need for a trade-off parameter such as α.

3.5.1 Derivation of the CRB-FOP

We begin by reconsidering the signal model of (3.2) in terms of the estimation problem. Assume that pilot tones are transmitted over $P \leq N$ frequency subchannels. For the purposes of intuition, also assume that the estimation of the channel impulse response will

be performed in two stages. The channel will first be estimated in frequency at $K < P$ of the frequency subchannels, and then these estimates will be used to obtain the channel impulse response. Note that while this estimation procedure aids intuition, the CRB to be derived is general in that it does not depend on this particular estimation procedure. Let

$$\mathbf{y}_P = \mathbf{E}_P^T \Gamma \mathbf{B} \mathbf{g} + \mathbf{n}_P. \tag{3.37}$$

The matrix $\mathbf{E}_P$ is defined as

$$\mathbf{E}_P = \left[\begin{array}{cccc} \mathbf{e}_{p_1} & \mathbf{e}_{p_2} & \cdots & \mathbf{e}_{p_P} \end{array} \right], \tag{3.38}$$

where $\mathbf{e}_i$ is the i-th column of the $N \times N$ identity matrix and $\{p_i\}$, $i = 1, \ldots, P$ is the set of frequency subchannel indexes for the P pilot locations. Thus, $\mathbf{y}_P$ and $\mathbf{n}_P = \mathbf{E}_P^T \mathbf{n}$ are the received signal and additive noise, respectively, at the OFDM frequency subchannels corresponding to pilot locations. The vector $\mathbf{g} = \mathbf{F}_L \mathbf{h}$ is the frequency response of the channel. Now define

$$\mathbf{g}_K = \mathbf{E}_K^T \mathbf{g}, \tag{3.39}$$

where $\mathbf{E}_K$ is defined in the same manner as $\mathbf{E}_P$ but over the index set $\{k_i\} \subset \{p_j\}$, i.e., $\mathbf{g}_K$ is the channel frequency response corresponding to $K < P$ of the pilot locations. We now wish to find a CRB with respect to our parameters

$$\mathbf{\Theta}_\mathbf{F} = \left[\begin{array}{ccc} \tau_d & \mathrm{Re}[\mathbf{g}_K^T] & \mathrm{Im}[\mathbf{g}_K^T] \end{array} \right]^T. \tag{3.40}$$

This initial CRB corresponds with the first part of our intuitional problem in which the channel is first estimated in frequency at K OFDM subchannels. As was previously done, we may use Bangs formula (see (3.9)) to find the FIM, the inverse of which is the CRB. The process of finding the derivatives required for (3.9) is straight forward, so we do not include it here. The resulting FIM is given by

$$J_F(\mathbf{\Theta}_F) = \frac{2}{\sigma_n^2} \left[\begin{array}{cc} J_{F,\tau_d} & \mathbf{J}_{F,21}^T \\ \mathbf{J}_{F,21} & \mathbf{J}_{F,\mathbf{h}} \end{array} \right] \tag{3.41}$$

with

$$J_{F,\tau_d} = \mathbf{g}^H \mathbf{B}^H \mathbf{D} \mathbf{Z}_P \mathbf{D} \mathbf{B} \mathbf{g}, \tag{3.42}$$

$$\mathbf{J}_{F,21} = \begin{bmatrix} \mathrm{Im}[\mathbf{E}_K^T \mathbf{B}^H \mathbf{D} \mathbf{B} \mathbf{g}] \\ -\mathrm{Re}[\mathbf{E}_K^T \mathbf{B}^H \mathbf{D} \mathbf{B} \mathbf{g}] \end{bmatrix}, \tag{3.43}$$

and

$$\mathbf{J}_{F,\mathbf{h}} = \begin{bmatrix} \mathbf{E}_K^T \mathbf{B}^H \mathbf{B} \mathbf{E}_K & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_K^T \mathbf{B}^H \mathbf{B} \mathbf{E}_K \end{bmatrix}, \tag{3.44}$$

where $\mathbf{D}$ is given in (3.14) and $\mathbf{Z}_P = \mathbf{E}_P \mathbf{E}_P^T$ is an $N \times N$ diagonal matrix with ones on the diagonal corresponding to the locations given by $\{p_i\}$ and zeros elsewhere. Using the block inversion formula of (3.16), the CRB may now be directly found, and it is given by

$$\mathbf{CRB}_F = \frac{\sigma_n^2}{2} \begin{bmatrix} \lambda_{\tau_d}^{-1} & \mathbf{CRB}_{F,21}^T \\ \mathbf{CRB}_{F,21} & \mathbf{CRB}_{F,22} \end{bmatrix}, \tag{3.45}$$

where

$$\begin{aligned} \lambda_{\tau_d} &= \mathbf{g}^H \mathbf{B}^H \mathbf{D} (\mathbf{Z}_P - \mathbf{Z}_K) \mathbf{D} \mathbf{B} \mathbf{g} \\ &= \mathbf{g}^H \mathbf{B}^H \mathbf{D} \mathbf{Z}_{P \backslash K} \mathbf{D} \mathbf{B} \mathbf{g}, \end{aligned} \tag{3.46}$$

$$\mathbf{CRB}_{F,21} = \begin{bmatrix} -\lambda_{\tau_d}^{-1} \mathrm{Im}[\mathbf{q}_F] \\ \lambda_{\tau_d}^{-1} \mathrm{Re}[\mathbf{q}_F] \end{bmatrix}, \tag{3.47}$$

$$\mathbf{CRB}_{F,22} = \begin{bmatrix} \mathbf{Q}_F^{-1} + \lambda_{\tau_d}^{-1} \mathrm{Im}[\mathbf{q}_F] \mathrm{Im}[\mathbf{q}_F^T] & -\lambda_{\tau_d}^{-1} \mathrm{Im}[\mathbf{q}_F] \mathrm{Re}[\mathbf{q}_F^T] \\ -\lambda_{\tau_d}^{-1} \mathrm{Re}[\mathbf{q}_F] \mathrm{Im}[\mathbf{q}_F^T] & \mathbf{Q}_F^{-1} + \lambda_{\tau_d}^{-1} \mathrm{Re}[\mathbf{q}_F] \mathrm{Re}[\mathbf{q}_F^T] \end{bmatrix}, \tag{3.48}$$

$$\mathbf{Q}_F = \mathbf{E}_K^T \mathbf{B}^H \mathbf{B} \mathbf{E}_K, \tag{3.49}$$

and

$$\mathbf{q}_F = (\mathbf{B} \mathbf{E}_K)^\dagger \mathbf{D} \mathbf{B} \mathbf{g}. \tag{3.50}$$

The matrix $\mathbf{Z}_{P \backslash K}$ is the $N \times N$ diagonal matrix with ones on the diagonal corresponding to the locations given by $\{p_i\} \backslash \{k_i\}$ and zeros elsewhere.

We now have a CRB with respect to the parameters $\boldsymbol{\Theta}_F$. However, what we would really like is a CRB with respect to the parameters $\boldsymbol{\Theta}$, that is, with respect to τ_d and the channel impulse response $\mathbf{h}$. Note that $\boldsymbol{\Theta}$ is a function of $\boldsymbol{\Theta}_F$ when $K \geq L$, i.e.,

$$\begin{aligned}\boldsymbol{\Theta} &= \begin{bmatrix} 1 & \mathbf{0}^T & \mathbf{0}^T \\ \mathbf{0} & \mathrm{Re}\left[(\mathbf{E}^T\mathbf{F}_L)^\dagger\right] & -\mathrm{Im}\left[(\mathbf{E}^T\mathbf{F}_L)^\dagger\right] \\ \mathbf{0} & \mathrm{Im}\left[(\mathbf{E}^T\mathbf{F}_L)^\dagger\right] & \mathrm{Re}\left[(\mathbf{E}^T\mathbf{F}_L)^\dagger\right] \end{bmatrix} \boldsymbol{\Theta}_F \\ &= \mathbf{T}\boldsymbol{\Theta}_F \end{aligned} \tag{3.51}$$

for $K \geq L$. Because of this relationship, we may find the desired CRB by using the CRB as a function of parameters (CRB-FOP) formulation [41]. That is, given a CRB as a function of $\boldsymbol{\Theta}_0$ and the functional relationship $\boldsymbol{\Theta}_1 = \mathbf{f}(\boldsymbol{\Theta}_0)$, the CRB with respect to $\boldsymbol{\Theta}_1$ may be found from the CRB with respect to $\boldsymbol{\Theta}_0$ as

$$\mathbf{CRB}(\boldsymbol{\Theta}_1) = \mathbf{J}'\mathbf{CRB}(\boldsymbol{\Theta}_0)\mathbf{J}'^H \tag{3.52}$$

where $\mathbf{J}'$ is the Jacobian of the function $\mathbf{f}(\boldsymbol{\Theta}_0)$ with respect to $\boldsymbol{\Theta}_0$, i.e.,

$$\mathbf{J}' = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial [\boldsymbol{\Theta}_0]_1} & \frac{\partial \mathbf{f}}{\partial [\boldsymbol{\Theta}_0]_2} & \cdots & \frac{\partial \mathbf{f}}{\partial [\boldsymbol{\Theta}_0]_K} \end{bmatrix}. \tag{3.53}$$

For our problem, the relationship between $\boldsymbol{\Theta}_F$ and $\boldsymbol{\Theta}$ is linear so the Jacobian is simply $\mathbf{J}' = \mathbf{T}$. Using (3.52), we find the CRB with respect to $\boldsymbol{\Theta}$ from the CRB of (3.45), and it is given by

$$\mathbf{CRB}_{\boldsymbol{\Theta}} = \frac{\sigma_n^2}{2} \begin{bmatrix} \lambda_{\tau_d}^{-1} & \mathbf{CRB}_{\boldsymbol{\Theta},21}^T \\ \mathbf{CRB}_{\boldsymbol{\Theta},21} & \mathbf{CRB}_{\boldsymbol{\Theta},22} \end{bmatrix}, \tag{3.54}$$

where

$$\mathbf{CRB}_{\boldsymbol{\Theta},21} = \begin{bmatrix} -\lambda_{\tau_d}^{-1}\mathrm{Im}[\mathbf{q}_{\boldsymbol{\Theta}}] \\ \lambda_{\tau_d}^{-1}\mathrm{Re}[\mathbf{q}_{\boldsymbol{\Theta}}] \end{bmatrix}, \tag{3.55}$$

$$\mathbf{CRB}_{\boldsymbol{\Theta},22} = \begin{bmatrix} \mathrm{Re}[\mathbf{Q}_{\boldsymbol{\Theta}}] + \lambda_{\tau_d}^{-1}\mathrm{Im}[\mathbf{q}_{\boldsymbol{\Theta}}]\mathrm{Im}[\mathbf{q}_{\boldsymbol{\Theta}}^T] & -\mathrm{Im}[\mathbf{Q}_{\boldsymbol{\Theta}}] - \lambda_{\tau_d}^{-1}\mathrm{Im}[\mathbf{q}_{\boldsymbol{\Theta}}]\mathrm{Re}[\mathbf{q}_{\boldsymbol{\Theta}}^T] \\ \mathrm{Im}[\mathbf{Q}_{\boldsymbol{\Theta}}] - \lambda_{\tau_d}^{-1}\mathrm{Re}[\mathbf{q}_{\boldsymbol{\Theta}}]\mathrm{Im}[\mathbf{q}_{\boldsymbol{\Theta}}^T] & \mathrm{Re}[\mathbf{Q}_{\boldsymbol{\Theta}}] + \lambda_{\tau_d}^{-1}\mathrm{Re}[\mathbf{q}_{\boldsymbol{\Theta}}]\mathrm{Re}[\mathbf{q}_{\boldsymbol{\Theta}}^T] \end{bmatrix}, \tag{3.56}$$

$$\mathbf{Q}_{\boldsymbol{\Theta}} = (\mathbf{E}_K^T\mathbf{F}_L)^\dagger \mathbf{Q}_F^{-1} (\mathbf{E}_K^T\mathbf{F}_L)^{\dagger H}, \tag{3.57}$$

and

$$\mathbf{q}_{\boldsymbol{\Theta}} = (\mathbf{E}_K^T \mathbf{F}_L)^{\dagger} \mathbf{q}_F. \tag{3.58}$$

3.5.2 The CRB-FOP and Pilot Selection

Given the CRB-FOP formulation of (3.54), let us make a few observations relative to the CRB and optimal pilot selection. First, we should emphasis that the quality of the CRB is not degraded as a result of this two-step procedure for its obtaining. In particular, it is shown in [41] that if an efficient estimator exists for the first set of parameters $\boldsymbol{\Theta}_0$, then an efficient estimator also exists for the second set of parameters $\boldsymbol{\Theta}_1$ assuming that the function relating the two parameter sets is linear, which it is in our case. Therefore, the CRB is not loosened by taking this derivation approach.

Next, we observe that the CRB subblock given by (3.56) is very similar in structure to that given in (3.22). Therefore, much of the intuition gained in Section 3.3 for pilot design relative to the trace of (3.22) also applies for this subblock. In fact, if the pilots are constrained to be equipowered, such as in the DPS case, it is easy to verify that the two expressions are identical except for the differences in $\lambda_{\tau_d}^{-1}$ and $\gamma_{\tau_d}^{-1}$.

Finally, we observe that λ_{τ_d} is almost identical in form to the uncoupled J_{τ_d} of (3.11). This does not mean that the expression for λ_{τ_d} is uncoupled in terms of the estimation of τ_d and $\mathbf{h}$. It is coupled. However, the form of this expression does suggest a very simple two-step procedure for addressing our design problem, finding a pilot configuration which minimizes the channel impulse response estimation error variance while meeting a constraint on the error variance of the time delay estimate. For the first step, pilot locations and powers may be selected to ensure that the constraint on τ_d estimation performance is met. This may be done for both the DPS and CPD scenarios in the same manner as in the uncoupled case. For the second step, with the remaining power, pilots are selected to minimize $\mathrm{tr}[\mathbf{CRB}_{\boldsymbol{\Theta},22}]$, and this may be done in a manner identical to the previously discussed coupled case with $\alpha = 0$, except that the pilot locations selected for the first part of the problem are excluded from consideration here. The important point is that this secondary pilot selection does not alter the value of $\lambda_{\tau_d}^{-1}$ realized by the initial design.

Table 3.1: Channel Impulse Response Coefficients

Element	Coefficient Value
h_0	$0.3802 + j0.2254$
h_1	$1.2968 - j0.9247$
h_2	$-1.5972 - j0.3066$
h_3	$0.6096 + j0.2423$

Note that this design procedure does not guarantee the optimal solution to our desired problem. This is because, as was mentioned previously, the CRB subbolocks relating to τ_d and $\mathbf{h}$ are not truly uncoupled. It is possible, for example, that selecting slightly less optimal locations when designing pilots for λ_{τ_d}, may make possible an improved pilot design for the secondary pilot selection. That being said, such a scenario is not likely when $P - K$ is small, so we expect this design procedure to yield optimal results in most instances.

3.6 Numerical Simulations

We now proceed to present simulation results demonstrating the design and properties of the optimal pilot signals chosen based upon the CRB formulations presented previously. For the results that follow, assume a channel impulse response is of length $L = 4$. We use an OFDM signal using $N = 32$ subchannels, a total pilot power of $P_T = L + 1$, and a measurement noise variance of $\sigma_n^2 = -20$ dB. The channel impulse response is selected randomly with coefficients drawn from a zero-mean unit-variance complex Gaussian distribution. These impulse response coefficients, ordered by increasing delay, are given in Table 3.1. When designing pilots for the DPS case, $P = 5$ pilots are used. For the CPD case, the Matlab function *fminunc* with a random initialization is used for minimization.

To begin, we first look at optimal pilot design results obtained when minimizing over the cost function $\mathbf{G}(\alpha, \mathbf{b})$. In Fig. 3.1, the optimal pilot configurations are shown for both the DPS and CPD cases when $\alpha = 0$. The magnitude of the channel frequency response, $|\mathbf{g}(k)|$, is superimposed on the plot as a reference. Recall that when $\alpha = 0$, full performance emphasis is placed on $\mathbf{h}$ estimation. For the CPD case, we observe $L + 1$ pilot clusters with

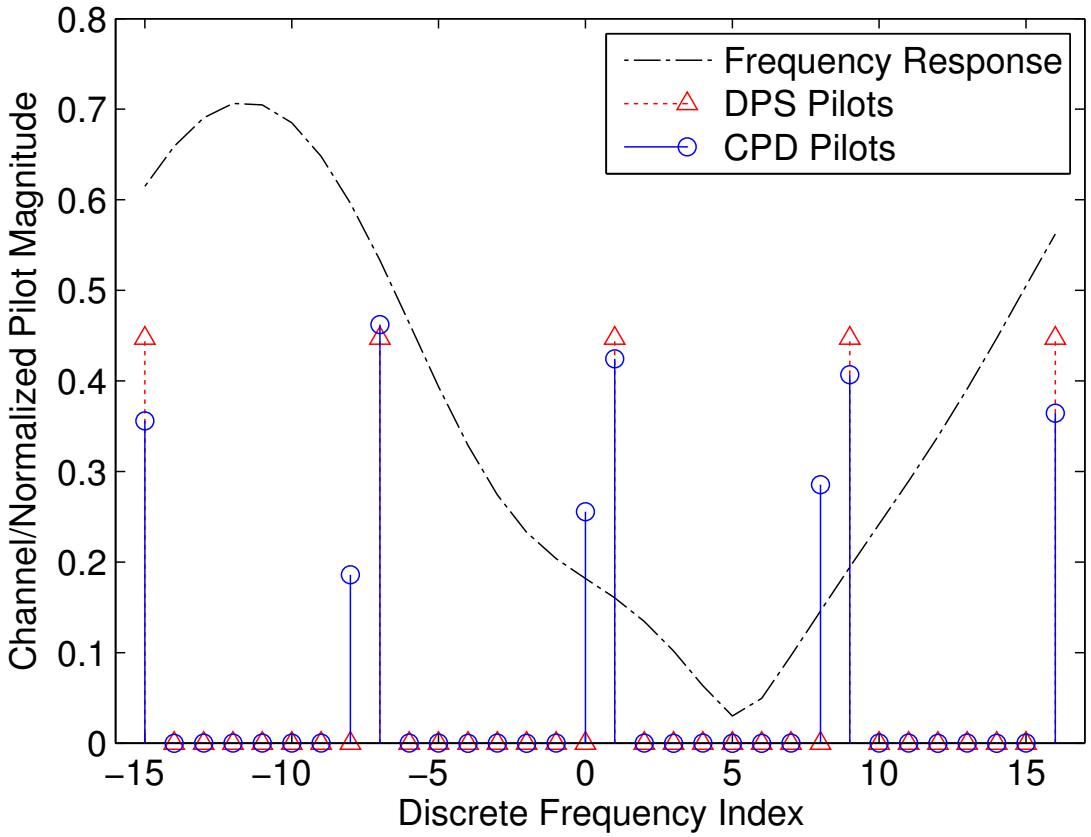


Figure 3.1: Example of DPS and CPD pilot design using $\mathbf{G}(\alpha, \mathbf{b})$ with $\alpha = 0$.

a relatively even spacing over the frequency subchannels, and the positions of these CPD pilots correspond well with the optimal pilot locations selected for the DPS scenario.

Note that it may be arguable whether or not there are actually L or $L+1$ CPD pilot clusters since pilot spacings may wrap around the edges of the frequency band. While pilot power is generally allocated near both band edges for $\alpha = 0$, we have noted in other simulations that these allocations do not tend to behave as a cluster, e.g., one or the other may move away from the band edge from time to time. Also, while the CPD pilot cluster locations correspond well with those found for the DPS scenario with $L+1$ equipowered pilots, that is not the case when using only L DPS pilots.

If we conclude that there are $L+1$ pilot clusters, the equispacing seen in this figure does not correspond to the notion of equispacing as defined in [20]. Relative to that case, the outermost pilots have been shifted towards the band edges and positions of the non-edge

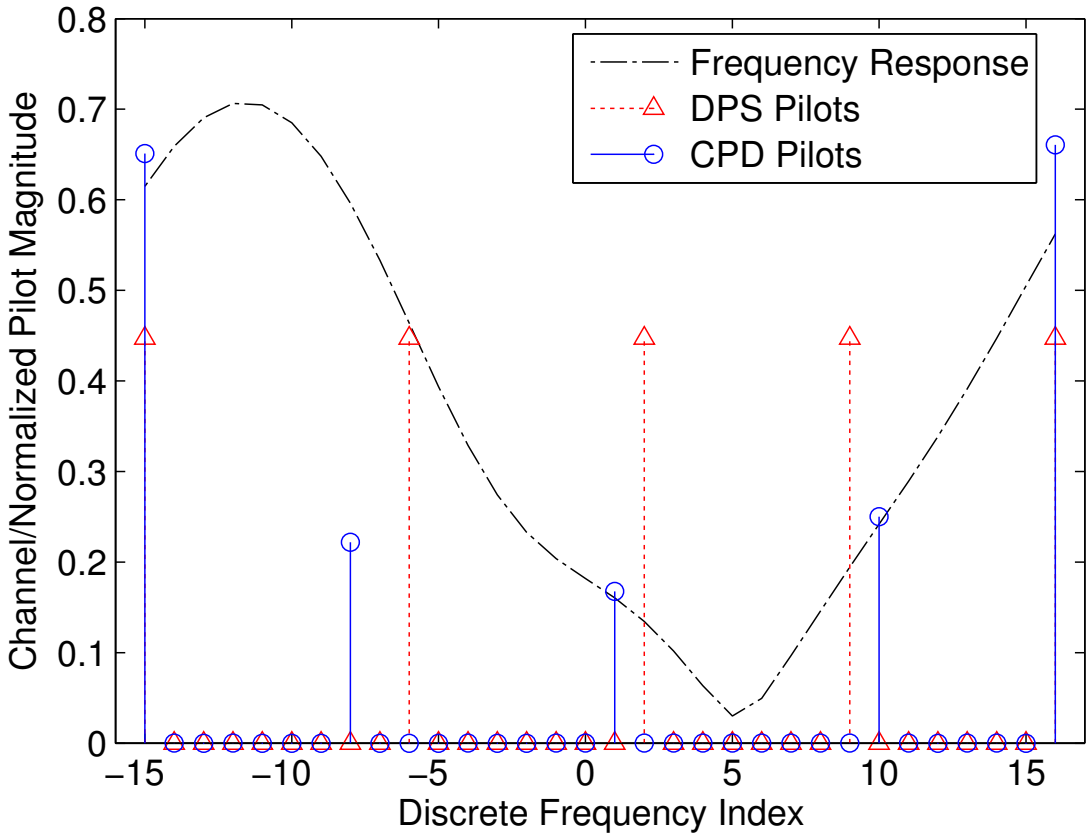


Figure 3.2: Example of DPS and CPD pilot design using $\mathbf{G}(\alpha, \mathbf{b})$ with $\alpha = 1$.

pilots have shifted to maintain approximately equal spacing within the interior of the band. In terms of power allocation, the interior pilot clusters are allocated a larger percentage of the available power than those on the band edges. All of these results agree with the discussion of Section 3.4.

For the complementary limiting case of $\alpha = 1$, simulation results are presented in Fig. 3.2. Recall that when $\alpha = 1$ the pilot locations are designed with sole emphasis placed on the time delay estimation performance. With this shift in emphasis comes a corresponding shift in relative power allocation among the pilots. For this case, the pilots near the band edges receive a larger relative power allocation than those towards the band center. In addition, the pilot locations are not as equally spaced as they were in the $\alpha = 0$ case, with some pilots shifting closer to the band edges for the CPD case and close to the locations

of lower channel response magnitude in the DPS case. Once again, the simulation example supports the intuition developed earlier.

It is useful to note in Fig. 3.2 that, like in Fig. 3.1, the locations of the DPS pilots have a relatively good correspondence with the locations of the CPD pilots, though the correspondence is not as strong in this case. This relationship between the pilot locations selected by the DPS when using $L+1$ pilots and those selected for the CPD is typical in the simulations we have carried out. It is greatest when α is small (though not, interesting, when $\alpha = 0$) and weakens as α is increased. Since for typical values of N, L, and P, the minimization used for the CPD optimization requires much less computation than the DPS exhaustive search, this relationship could be used to significantly reduce the computational burden of the DPS optimization. For example, the CPD optimization could be used to restrict the DPS pilot search space considerably, which would result in dramatic computational reductions.

We now present an example in which the minimization of (3.25) is used to find pilot designs which minimize the variance of the channel coefficient estimates while satisfying a constraint on the variance of the time delay estimate. The constraint on the delay variance, or, correspondingly, on $\gamma_{\tau_d}^{-1}$, is chosen to be $\gamma_{\text{TARGET}}^{-1} = 6 \times 10^{-5} \times (T_s/(2\pi\sigma_n))^2$. Given this constraint, the simulation is carried out by varying α over thirty-two evenly-spaced points on $[0, 1]$, solving the minimization of (3.25) at each point, and selecting the solution which satisfies and is closest to the desired constraint. For the CPD case, the best pilot design is found at a value of $\alpha = 0.61$ with $\gamma_{\tau_d}^{-1} = 5.9 \times 10^{-5} \times (T_s/(2\pi\sigma_n))^2$ and $\text{tr}[\mathbf{CRB}_{22}]/L = 0.077$. The DPS, however, was unable to meet the constraint with the available power, achieving only $\gamma_{\tau_d}^{-1} = 6.6 \times 10^{-5} \times (T_s/(2\pi\sigma_n))^2$ and $\text{tr}[\mathbf{CRB}_{22}]/L = 0.019$ with $\alpha = 1$. The resulting pilot configurations are shown in Fig. 3.3.

Note that for the CPD case, the values of the two terms of $\mathbf{G}(\alpha, \mathbf{b}^*)$ at the optimal $\mathbf{b}*$ change relatively smoothly with α. This observation may be used to simplify the search for the value of α which satisfies the design problem above. In Fig. 3.4, the values of these terms are plotted relative to α for the optimal pilot locations. Except for the occasional outlier, most channel realizations investigated in our simulations have demonstrated trade-off curve features comparable to those seen in the figure.

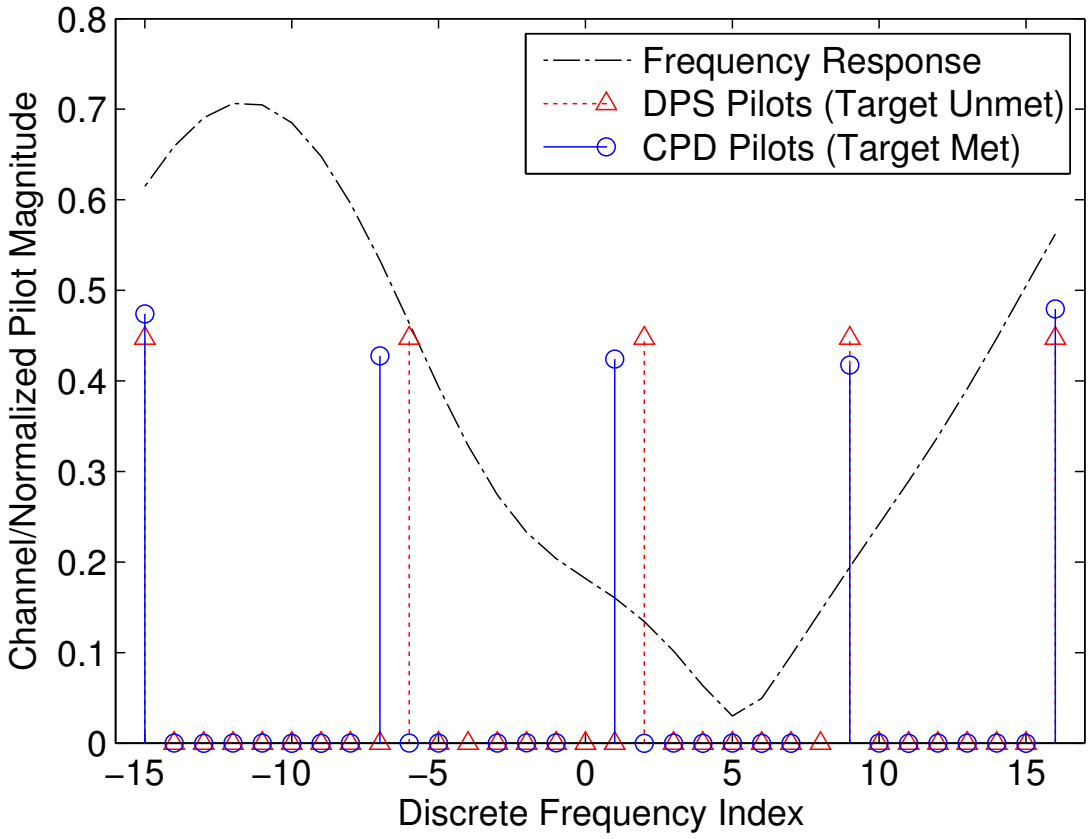


Figure 3.3: Example of DPS and CPD pilot design using $\mathbf{G}(\alpha, \mathbf{b})$ with constraint $\frac{\sigma_n^2}{2}\gamma_{\tau_d}^{-1} \leq 3 \times 10^{-5} \times (T_s/(2\pi))^2$; the target is not achieved in the DPS case resulting in $\alpha = 1$, $\frac{\sigma_n^2}{2}\gamma_{\tau_d}^{-1} = 3.3 \times 10^{-5} \times (T_s/(2\pi))^2$, and $\mathrm{tr}[\mathbf{CRB}_{22}]/L = 0.019$, but is achieved in the CPD case with $\alpha = 0.61$, $\frac{\sigma_n^2}{2}\gamma_{\tau_d}^{-1} = 2.9 \times 10^{-5} \times (T_s/(2\pi))^2$, and $\mathrm{tr}[\mathbf{CRB}_{22}]/L = 0.077$.

Finally, we consider an example in which the CRB-FOP formulation is used for optimal pilot design. Once again, we are considering the problem of finding pilot designs which minimize the variance of the channel coefficient estimates while satisfying a constraint on the variance of the time delay estimate. Figure 3.5 presents the optimal pilot designs for our problem using the CRB-FOP formulation and targeting the same constraint as in the previous design example. Recall that, unlike the previous case, it is not necessary with this formulation to search for a constraint-satisfying solution over a trade-off parameter such as α. Instead, the pilots are designed in two steps, first by directly satisfying the constraint, and then by minimizing the channel estimation variance with the remaining available power. Both the DPS and CPD where able to solve this problem, though the DPS required three pilots to

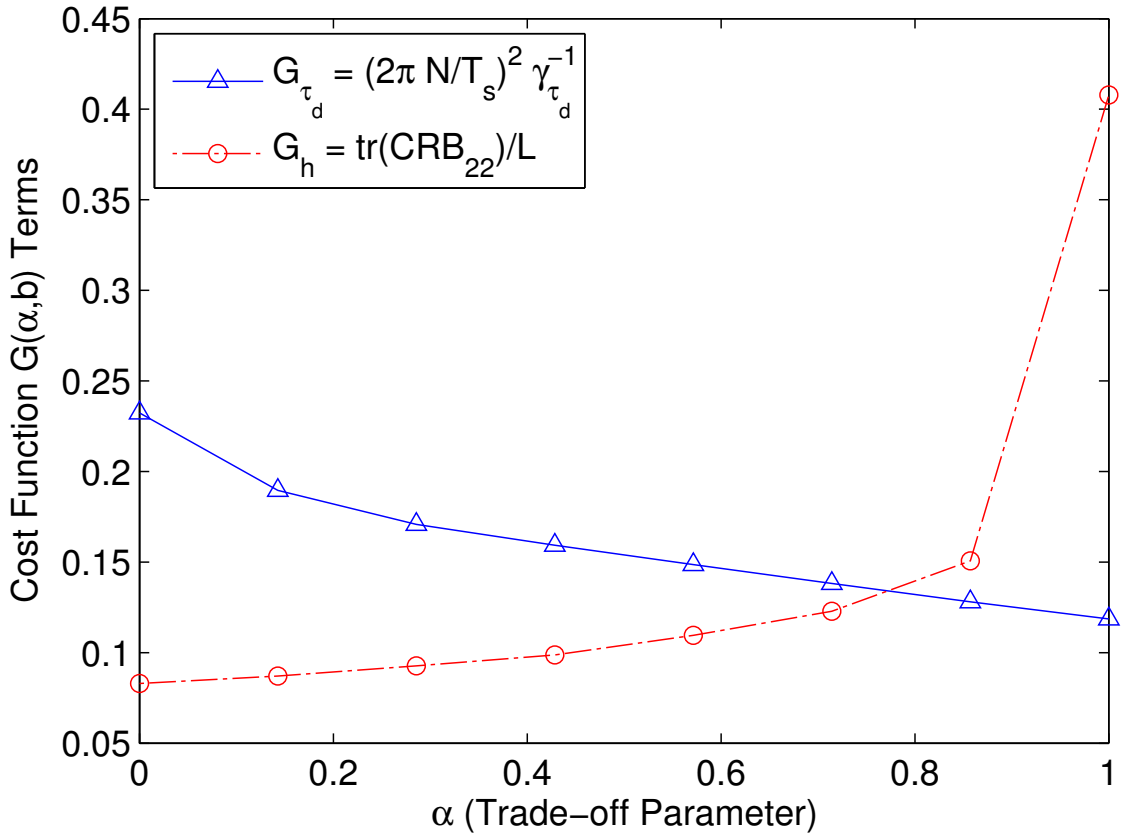


Figure 3.4: Demonstration of the influence of α on the terms of $\mathbf{G}(\alpha, \mathbf{b})$.

satisfy the delay constraint. For the pilot solutions shown, $\lambda_{\tau_d}^{-1} = 5.6 \times 10^{-5} \times (T_s/(2\pi\sigma_n))^2$ and $\text{tr}[\mathbf{CRB}_{\boldsymbol{\Theta},22}]/L = 0.117$ for the DPS case and $\lambda_{\tau_d}^{-1} = 6 \times 10^{-5} \times (T_s/(2\pi\sigma_n))^2$ and $\text{tr}[\mathbf{CRB}_{\boldsymbol{\Theta},22}]/L = 0.099$ for the CPD scenario. Note that in both cases, the channel estimation performance is lower than that found using the standard CRB formulation. This is a result of the model used to derived the CRB as a function of parameters. The model prohibits the channel estimation from using information garnered by pilots allocated exclusively for the delay estimation. However, this loss may be a small price to pay for the design simplifications available when using this CRB formulation.

3.7 Conclusions

In this chapter, we have considered the design of optimal pilot signals for the joint estimation of the time delay and channel impulse response parameters in an OFDM com-

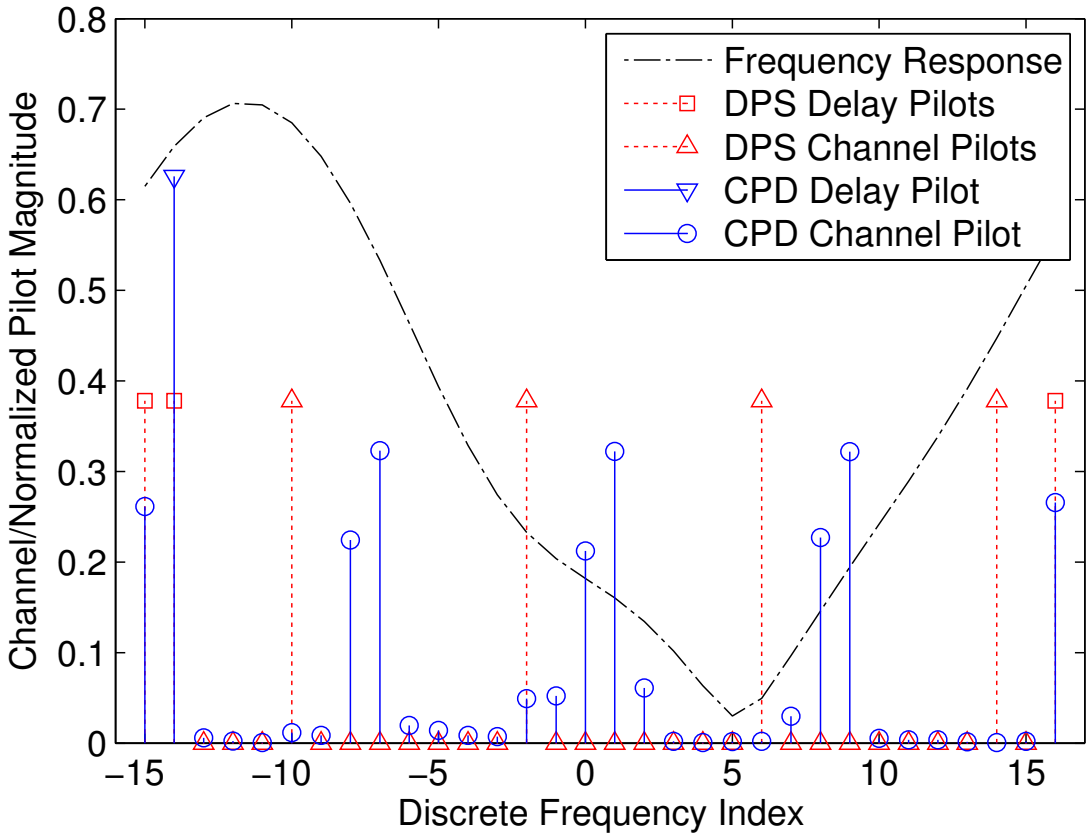


Figure 3.5: Example of DPS and CPD pilot design using the CRB-FOP method with constraint $\frac{\sigma_n^2}{2}\lambda_{\tau_d}^{-1} \leq 3 \times 10^{-5} \times (T_s/(2\pi))^2$; the target achieved using both DPS and CPD methods with $\frac{\sigma_n^2}{2}\lambda_{\tau_d}^{-1} = 2.8 \times 10^{-5} \times (T_s/(2\pi))^2$ and $\mathrm{tr}[\mathbf{CRB}_{\boldsymbol{\Theta},22}]/L = 0.117$ for the DPS case and $\frac{\sigma_n^2}{2}\lambda_{\tau_d}^{-1} = 3.0 \times 10^{-5} \times (T_s/(2\pi))^2$ and $\mathrm{tr}[\mathbf{CRB}_{\boldsymbol{\Theta},22}]/L = 0.099$ for the CPD case.

munications system. In particular, the chapter addressed the problem finding the pilot locations and power levels which minimize the channel impulse response estimation error variance while meeting a constraint on the error variance of the time delay estimate. Following the introduction of the OFDM signaling model, the CRB was derived relative to the both the time delay and channel impulse response. A weighted trace of the CRB was then used to address the pilot design problem. While the optimal pilots could not be found directly, the expression allowed us to determine many of the general properties of the optimal solution. Following this presentation, a second CRB was derived as a function of intermediate parameters. This CRB, while similar to the first, led to a simplified method for finding optimal pilot configurations. The numerical simulations demonstrated the use of these CRBs

in solving the pilot design problem and supported the conclusions of the paper concerning the structure of the optimal pilots.

Chapter 4

Performance Bounds for MIMO-OFDM Channel Estimation

4.1 Introduction

In this chapter, we study the theoretical performance of pilot-based channel interpolation and prediction for frequency-selective, time-fading, wireless MIMO-OFDM channels via bounds for the interpolation and prediction error of the channel. Our analysis of these bounds demonstrates that (1) better estimation and prediction performance can be obtained using MIMO systems, (2) parametric channel modeling is advantageous in terms of estimation and prediction performance, but (3) the presence of modeling errors quickly degrades the performance of parametric approaches based on calibrated arrays and necessitates the use of more robust models. The analysis is based on extended ray-based parametric channel models for SISO and MIMO channels. The lower bounds on interpolation and prediction error are derived using vector formulations of the CRB for biased and unbiased estimators and for functions of parameters, in a manner similar to [59, 60]. We consider bounds for several different types of channel parameterizations. The first model employs directions of departure (DOD) and directions of arrival (DOA) at the transmit and receive arrays, respectively, and includes the effects of imperfect array calibration on estimation performance. We also consider bounds for models that are formulated using more robust spatial-signature representations of the channel instead of DOD and DOA information. We show that while angle-based models enjoy a performance advantage for perfectly calibrated arrays, the advantage is quickly lost when modeling errors are taken into account. The derived bounds allow one to compute the size of the calibration error required before spatial-signature models yield better results. Note that earlier versions of this work are presented in [91–93].

The remainder of the chapter is organized as follows. Section 4.2 introduces the DOD/DOA and spatial-signature-based channel models. The performance bounds on the

interpolation and prediction error are derived in Section 4.3, and numerical evaluations of the bounds are examined in Section 4.4. Finally, concluding remarks are given in Section 4.5.

4.2 Channel Models

We consider wideband ray-based channel models, where the signal at the receiver is represented by a sum of a finite number of copies of the signal sent by the transmitter, each copy experiencing its own attenuation, delay, and Doppler. The resulting channel model is variable in both time (t) and frequency (ω), and can be expressed as the following $M_r \times M_t$ matrix function

$$\mathbf{H}(\omega, t) = \sum_{l=1}^{L} \alpha_l \mathbf{a}_{r,l}(\omega) \mathbf{a}_{t,l}^T(\omega) e^{j((\omega_c - \omega)\tau_l - \omega_{d,l} t)} , \tag{4.1}$$

where L denotes the total number of propagation paths, ω_c is the center or reference frequency of the frequency band of interest, and associated with each path l are the following parameters:

- complex gain α_l
- Doppler frequency $\omega_{d,l}$
- delay τ_l
- $M_t \times 1$ transmit antenna array response $\mathbf{a}_{t,l}(\omega)$
- $M_r \times 1$ receive antenna array response $\mathbf{a}_{r,l}(\omega)$.

We use the above model over time intervals where the relative positions of the transmitter and receiver change by at most a few tens of wavelengths, and thus we assume that the given physical channel parameters are time-invariant. The time-varying phase due to the Doppler induces the multipath fading effect. Narrowband (frequency flat) versions of this model have been used in several channel measurement campaigns [86–88] and form the basis for the derivations and analysis of [60]. The wideband version of the model can be derived directly as a time-varying multipath channel kernel [89]. As described in the sections that

follow, we will consider several different ways of parameterizing the spatial components of the channel $\mathbf{a}_{t,l}(\omega)$ and $\mathbf{a}_{r,l}(\omega)$.

4.2.1 DOD/DOA Model

For the DOD/DOA model, we assume that the array response vectors $\mathbf{a}_{t,l}(\omega)$ and $\mathbf{a}_{r,l}(\omega)$ in (4.1) are functions of the DOD and DOA, respectively, of signal path l (while we use a scalar direction parameter to describe the DOD or DOA, the approach below is easily extended to cases where the array response vectors depend on multiple parameters, including azimuth and elevation angles, polarization states, etc). This model is valid for any array geometry; for example, a uniform linear array at the transmitter may be described using the Vandermonde structure

$$\mathbf{a}_{t,l}^T(\Omega_{t,1},\omega) = \left[\begin{array}{cccc} 1 & e^{-j\Omega_{t,l}(\omega)} & \cdots & e^{-j(M_t-1)\Omega_{t,l}(\omega)} \end{array}\right], \tag{4.2}$$

where $\Omega_{t,l}(\omega) = k(\omega)d_t \sin\phi_{t,l}$ is the solid angle of path l, $k(\omega)$ is the frequency-dependent wave number, d_t is the separation between antenna elements, and $\phi_{t,l}$ is the direction of departure (DOD) of path l.

We may also express (4.1) for the DOD/DOA model in matrix form as

$$\mathbf{H}(\omega,t) = \mathbf{A}_r(\mathbf{\Omega_r},\omega)\mathbf{X}\mathbf{W}(\omega,t)\mathbf{A}_t(\mathbf{\Omega_t},\omega)^T \tag{4.3}$$

with

$$\mathbf{A}_r(\mathbf{\Omega_r},\omega) = \left[\begin{array}{cccc} \mathbf{a}_{r,1}(\Omega_{r,1},\omega) & \cdots & \mathbf{a}_{r,L}(\Omega_{r,L},\omega) \end{array}\right], \tag{4.4}$$

$$\mathbf{A}_t(\mathbf{\Omega_t},\omega) = \left[\begin{array}{cccc} \mathbf{a}_{t,1}(\Omega_{t,1},\omega) & \cdots & \mathbf{a}_{t,L}(\Omega_{t,L},\omega) \end{array}\right], \tag{4.5}$$

$$\mathbf{X} = \mathrm{diag}(\alpha_1,\alpha_2,\ldots,\alpha_L), \tag{4.6}$$

and

$$\mathbf{W}(\omega,t) = \mathrm{diag}(e^{j\gamma_1(\omega,t)},\ldots,e^{j\gamma_L(\omega,t)}), \tag{4.7}$$

where $\gamma_l(\omega,t) = e^{j((\omega_c-\omega)\tau_l-\omega_{d,l}t)}$. In (4.3), the dependence of the array responses in the DOD/DOA model on frequency and the solid angles of departure and arrival is explicitly

shown. For notational simplicity, we omit this dependence from the notation in what follows, though this dependence is implied when discussing the DOD/DOA model. It is convenient for our analysis to represent (4.3) in a vectorized form:

$$\mathbf{h}(\omega, t) = (\mathbf{A}_t \otimes \mathbf{A}_r\mathbf{X})\text{vec}(\mathbf{W}(\omega, t)), \tag{4.8}$$

where $\otimes$ is the Kronecker product and $\text{vec}(\mathbf{A})$, the vectorization operator, stacks the columns of $\mathbf{A}$. Thus, $\mathbf{h}(\omega, t)$ is an $M_t M_r \times 1$ vector.

The channel model in (4.8) is parameterized by the L-length vector parameters $\boldsymbol{\alpha}$, $\boldsymbol{\tau}$, $\boldsymbol{\omega}$, $\boldsymbol{\Omega}_t$, and $\boldsymbol{\Omega}_r$, which are of the form $\boldsymbol{\alpha}^T = [\alpha_1, \alpha_2, \ldots, \alpha_L], \boldsymbol{\omega}^T = [\omega_{d,1}, \omega_{d,2}, \ldots, \omega_{d,L}]$, and so on. For convenience, we will represent these parameters collectively using a single real-valued vector $\boldsymbol{\Theta} = \left[\text{Re}[\boldsymbol{\alpha}]^T, \text{Im}[\boldsymbol{\alpha}]^T, \boldsymbol{\tau}^T, \boldsymbol{\omega}^T, \boldsymbol{\Omega}_t^T, \boldsymbol{\Omega}_r^T\right]^T$. Note that the number of parameters depends only on the number of paths L, not on the size of the antenna arrays M_t and M_r.

4.2.2 DOD/DOA Model with Calibration Errors

The above DOD/DOA model inherently assumes ideally calibrated arrays, with array responses that are precisely known functions of the DOD or DOA. This is never the case in practice due, for example, to effects such as mutual coupling, imprecisely known antenna positions, RF non-linearities, I/Q imbalance, etc. To more realistically study the performance of parametric models based on array calibration, we introduce a generalized model in which the actual array responses are described as the sum of the nominal ideal response and a perturbation term: $\mathbf{a}_{t,l} + \mathbf{v}_{t,l}$ and $\mathbf{a}_{r,l} + \mathbf{v}_{r,l}$. In matrix form, the DOD/DOA model with calibration errors (CE) becomes

$$\mathbf{H}(\omega, t) = \sum_{l=1}^{L} \alpha_l(\mathbf{a}_{r,l} + \mathbf{v}_{r,l})(\mathbf{a}_{t,l} + \mathbf{v}_{t,l})^T e^{j\gamma_l(\omega,t)} \tag{4.9}$$

$$= (\mathbf{A}_r + \mathbf{V}_r)\mathbf{X}\mathbf{W}(\omega, t)(\mathbf{A}_t + \mathbf{V}_t)^T, \tag{4.10}$$

where $\mathbf{V}_t$ and $\mathbf{V}_r$ are defined in a similar manner as $\mathbf{A}_t$ and $\mathbf{A}_r$. In vectorized form, the model becomes

$$\mathbf{h}(\omega, t) = ((\mathbf{A}_t + \mathbf{V}_t) \otimes (\mathbf{A}_r + \mathbf{V}_r)\mathbf{X})\text{vec}(\mathbf{W}(\omega, t)). \tag{4.11}$$

As with (4.8), this model is parameterized by $\boldsymbol{\alpha}$, $\boldsymbol{\tau}$, $\boldsymbol{\omega}$, $\boldsymbol{\Omega}_t$, and $\boldsymbol{\Omega}_r$. While it is possible to treat the calibration error variables $\mathbf{V}_t$ and $\mathbf{V}_r$ as additional model parameters to be estimated, in this paper we will focus on analyzing the performance of methods that ignore the presence of such perturbations. In particular, we show how to augment the resulting performance bound to account for the bias that results from mismodeling the array response.

4.2.3 Vector Spatial Signature Model

The DOD/DOA model assumes specific array configurations that depend on the parameters $\boldsymbol{\Omega}_t$ and $\boldsymbol{\Omega}_r$. The estimation of these parameters can be difficult and performance is often very sensitive to calibration errors. Such problems can be avoided with the use of a more general model in which the path gains and the angle- and position-dependent array responses of the DOD/DOA model are replaced by unstructured vectors, termed *spatial signatures*. In this case, the model of (4.1) becomes simply

$$\mathbf{H}(\omega,t) = \sum_{l=1}^{L} \mathbf{a}_{r,l}\mathbf{a}_{t,l}^T e^{j((\omega_c-\omega)\tau_l-\omega_{d,l}t)} \tag{4.12}$$

$$= \mathbf{A}_r\mathbf{W}(\omega,t)\mathbf{A}_t^T, \tag{4.13}$$

or, in vectorized form,

$$\mathbf{h}(\omega,t) = (\mathbf{A}_t \otimes \mathbf{A}_r)\mathrm{vec}(\mathbf{W}(\omega,t)). \tag{4.14}$$

The vectors $\mathbf{a}_{t,l}$ and $\mathbf{a}_{r,l}$ are not explicit functions of DOD or DOA, but instead abstractly represent the transmit and receive array responses for path l with delay τ_l and Doppler $\omega_{d,l}$. Note that while simpler to estimate and insensitive to calibration errors, this vector spatial signature (VSS) model approximates the array response vectors as being frequency independent, which is not true when using physical antenna arrays with wideband signals. The VSS model is parameterized by $\boldsymbol{\tau}$, $\boldsymbol{\omega}$, the LM_t-element vectors $\mathrm{Re}[\mathbf{a}_t], \mathrm{Im}[\mathbf{a}_t]$, and the LM_r-element vectors $\mathrm{Re}[\mathbf{a}_r], \mathrm{Im}[\mathbf{a}_r]$, where $\mathbf{a}_t = \mathrm{vec}(\mathbf{A}_t)$ and $\mathbf{a}_r = \mathrm{vec}(\mathbf{A}_r)$. Unlike the DOD/DOA model, the number of parameters depends on M_t and M_r, as well as L.

4.2.4 Matrix Spatial Signature Model

In this section, we generalize (4.12) one step further by replacing the vector spatial signatures $\mathbf{a}_{t,l}$ and $\mathbf{a}_{r,l}$ of path l by a matrix spatial signature $\mathbf{A}_l$ so that

$$\mathbf{H}(\omega, t) = \sum_{l=1}^{L} \mathbf{A}_l e^{j((\omega_c - \omega)\tau_l - \omega_{d,l} t)}. \tag{4.15}$$

As in the VSS model, the matrices $\mathbf{A}_l$ are not assumed to be explicit functions of DOD or DOA, instead they abstractly represent the channel characteristics for a particular path with associated delay τ_l and Doppler $\omega_{d,l}$. Since no restriction is placed on the elements of the matrices $\mathbf{A}_l$, they may be of arbitrary rank, in contrast to the rank-one matrices formed by the outer products $\mathbf{a}_{r,l}\mathbf{a}_{t,l}^T$ in the VSS model. The consequences of this fact will be seen later. Note that as in the VSS case, this model assumes the matrix spatial signatures are also frequency independent. This matrix spatial signature (MSS) model is essentially a two-dimensional filter in ω and t with matrix filter taps. At a fixed time t, it is similar to the multi-channel finite impulse response (FIR) channel models frequently used in the literature, although the taps are not evenly spaced in time.

As with the other models, the summation in (4.15) can be expressed using matrix operations as

$$\begin{aligned}
\mathbf{H}(\omega, t) &= \sum_{l=1}^{L} \mathbf{A}_l w_l(\omega, t) \\
&= \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_L \end{bmatrix} \begin{bmatrix} w_1 \mathbf{I}_{M_t} \\ w_2 \mathbf{I}_{M_t} \\ \vdots \\ w_L \mathbf{I}_{M_t} \end{bmatrix} \\
&= \mathbf{A}(\mathbf{w} \otimes \mathbf{I}_{M_t}),
\end{aligned} \tag{4.16}$$

where

$$w_l(\omega, t) = e^{j((\omega_c - \omega)\tau_l - \omega_{d,l} t)}, \tag{4.17}$$

Table 4.1: Number of Real-Valued Model Parameters

Model	No. of Parameters
DOD/DOA	$6L$
VSS	$2L(M_t + M_r + 1)$
MSS	$2L(M_t M_r + 1)$

$$\mathbf{A} = \left[\begin{array}{cccc} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_L \end{array} \right], \tag{4.18}$$

and

$$\mathbf{w}^T = \left[\begin{array}{cccc} w_1(\omega, t) & w_2(\omega, t) & \cdots & w_L(\omega, t) \end{array} \right]. \tag{4.19}$$

In vector form,

$$\mathbf{h}(\omega, t) = (\mathbf{I}_{M_t} \otimes \mathbf{A})\text{vec}(\mathbf{w} \otimes \mathbf{I}_{M_t}). \tag{4.20}$$

The MSS model is parameterized by $\boldsymbol{\tau}$, $\boldsymbol{\omega}$ (through $\mathbf{w}$), and the $M_r \times LM_t$ real-valued matrices Re[$\mathbf{A}$] and Im[$\mathbf{A}$]. Thus, the $2L(M_t M_r + 1)$ parameter vector $\boldsymbol{\Theta}$ depends on both the number of paths and the array sizes. Table 4.1 summarizes the number of parameters for each model.

4.3 Lower Bound on Estimation/Prediction Error

In what follows, we will derive Cramér-Rao bounds (CRBs) for the performance of channel estimation as functions of the parameters of the models presented above. We make the following two assumptions common to most MIMO-OFDM systems: (1) we assume that bursts of training or pilot data are available at several points in time or at several frequencies or both, and (2) we assume that these bursts are short enough in time and narrow enough in frequency that the channel can be assumed to be constant over their duration and bandwidth. In most MIMO-OFDM applications, each burst of training data would be used to produce a local "snapshot" of the channel, and the various local estimates would then be interpolated or extrapolated in order to find estimates of the channel at other times and frequencies. An alternative approach would be to use all of the pilot data directly to find channel estimates at all times and frequencies of interest, rather than breaking the process into two-steps involving intermediate channel snapshots. We will focus on the former approach, but we

discuss analysis of the latter approach in Section 4.3.1. In particular, we show that under certain standard assumptions, the two approaches yield identical bounds.

To derive the CRB here, we assume that pilot symbols are used to obtain a series of N_M channel measurements $\tilde{\mathbf{h}}(\omega_n, t_n)$ at time-frequency pairs $(\omega_1, t_1), \cdots, (\omega_{N_M}, t_{N_M})$. These measurements are imperfect due, for example, to noise and interference present along with the training data. Thus, we model the channel measurements as a sum of the true channel $\mathbf{h}(\omega, t)$ and a Gaussian noise term due to estimation error, so that

$$\tilde{\mathbf{h}}(\omega_n, t_n) = \mathbf{h}(\omega_n, t_n) + \mathbf{n}(\omega_n, t_n), \tag{4.21}$$

where the $M_t M_r \times 1$ Gaussian noise term is distributed as $\mathbf{n}(\omega, t) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{C}_n)$. The N_M samples form the measurement segment, which is used to interpolate or extrapolate the channel to other times and frequencies of interest. For convenience, we stack the N_M channel measurements into an $N_M M_t M_r$-length vector as follows:

$$\tilde{\mathbf{h}} = \begin{bmatrix} \tilde{\mathbf{h}}(\omega_1, t_1) \\ \tilde{\mathbf{h}}(\omega_2, t_2) \\ \vdots \\ \tilde{\mathbf{h}}(\omega_{N_M}, t_{N_M}) \end{bmatrix} = \mathbf{h} + \mathbf{n}, \tag{4.22}$$

where $\mathbf{h}$ is parameterized by any of the models described in Section 4.2 and $\mathbf{n}$ is a $N_M M_t M_r$-length stacked estimation error vector distributed jointly as $\mathbf{n} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{C})$. Our analysis is general enough to accommodate an arbitrary covariance matrix $\mathbf{C}$. However, for simplicity in the presentation of the CRB derivations and expressions, we will assume the channel measurement error to be spatially and temporally white, so that $\mathbf{C} = \sigma\mathbf{I}$ where $\mathbf{I}$ is an $M_t M_r \times M_t M_r$ identity matrix and σ is the variance. This is a common model in the literature that holds when unitary training symbols are used to form the channel measurements. As will be seen shortly, the derivatives with respect to the covariance and the model are decoupled, allowing the results derived using this assumption to be generalized in a straight-forward manner. An expression for σ in terms of the signal-to-noise ratio and the number of training symbols can be found in [60], although in computing the channel estimation bounds we will

assume that σ is an unknown parameter that also must be estimated. Thus, σ must be added as a parameter to each of the models described by $\mathbf{\Theta}$ in the previous section.

At any particular time and frequency (ω, t), the channel estimate is written as $\hat{\mathbf{h}}(\omega, t)$ and the channel estimation error is given by

$$\mathbf{e}(\omega, t; \tilde{\mathbf{h}}, \mathbf{\Theta}) = \hat{\mathbf{h}}(\omega, t; \tilde{\mathbf{h}}) - \mathbf{h}(\omega, t; \mathbf{\Theta}). \tag{4.23}$$

For clarity, we have explicitly included the $\tilde{\mathbf{h}}$ and $\mathbf{\Theta}$ dependence in (4.23), but in what follows we omit these dependencies for simplicity and write the channel estimate as $\hat{\mathbf{h}}(\omega, t)$ and the estimation error as $\mathbf{e}(\omega, t)$. Our goal is to find a lower bound on the covariance matrix of any unbiased estimator $\hat{\mathbf{h}}$ using the Cramér-Rao bound (CRB). Using a vector formulation of the CRB [78], the bound may be written as

$$E\left[\left(\hat{\mathbf{h}}(\omega, t) - \mathbf{h}(\omega, t)\right)\left(\hat{\mathbf{h}}(\omega, t) - \mathbf{h}(\omega, t)\right)^H\right] \geq \mathbf{H}'\mathbf{B}\mathbf{H}'^H, \tag{4.24}$$

where the matrix inequality $\mathbf{F} \geq \mathbf{G}$ indicates that the matrix difference $\mathbf{F} - \mathbf{G}$ is positive semi-definite, $\mathbf{B}$ is the CRB matrix with respect to the parameters $\mathbf{\Theta}$, and $\mathbf{H}'$ is the Jacobian matrix

$$\mathbf{H}' = \left[\begin{array}{cccc} \frac{\partial \mathbf{h}(\omega,t)}{\partial \theta_1} & \frac{\partial \mathbf{h}(\omega,t)}{\partial \theta_2} & \cdots & \frac{\partial \mathbf{h}(\omega,t)}{\partial \theta_P} \end{array}\right], \tag{4.25}$$

where P is the number of model parameters, e.g., $P = 6L + 1$ for the DOD/DOA model, $P = 2L(M_t + M_r + 1) + 1$ for the VSS model, and $P = 2L(M_t M_r + 1) + 1$ for the MSS model. Matrix $\mathbf{B}$, the CRB with respect to $\mathbf{\Theta}$, can be calculated using Bangs formula [40]

$$\begin{aligned} \left[\mathbf{B}^{-1}\right]_{ij} &= \mathrm{tr}\left[\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \theta_i}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \theta_j}\right] + 2\mathrm{Re}\left[\frac{\partial \mathbf{h}^H}{\partial \theta_i}\mathbf{C}^{-1}\frac{\partial \mathbf{h}}{\partial \theta_j}\right] \\ &= \mathbf{J}(\mathbf{\Theta})_{ij}, \end{aligned} \tag{4.26}$$

where $\mathbf{J}(\mathbf{\Theta})$ is the Fisher information matrix (FIM).

When considering the form of the CRB expression in (4.24), it may be instructive to assume a scenario in which the channel parameters $\mathbf{\Theta}$ are estimated using the N_M available channel measurements, and then the channel is estimated at time/frequency locations out-

side the measurement segment using the parameter estimates $\hat{\mathbf{\Theta}}$ and the appropriate channel model. In practice, such a method may be difficult and computationally burdensome, particularly when L is large. However, the CRB formulation above does not depend upon such an estimation scheme. Nor is our goal to suggest a particular algorithm for channel interpolation and extrapolation, but is instead to bound the best possible performance that could be achieved given the assumed channel models.

Once the CRB is known, the sum of variances of the elements of the estimation error vector $\mathbf{e}(\omega, t)$ may be bounded as

$$E\left[\|\mathbf{e}(\omega,t)\|_2^2\right] \geq \mathrm{tr}\left[\mathbf{H}'\mathbf{B}\mathbf{H}'^H\right], \tag{4.27}$$

where $\|\cdot\|_2$ denotes the Euclidean norm. Note that even though $\mathbf{B}$ depends on the N_M channel measurements, this expression is valid for any (ω, t) pair through the transformation $\mathbf{H}'$, not just those in the measurement segment. That is, once the CRB for the model parameters $\mathbf{\Theta}$ is found, the bound for $\mathbf{h}$ may be calculated for any (ω, t).

4.3.1 Direct vs. Indirect Use of Pilots

As mentioned above, we focus in this paper on the two-step procedure where channel measurements are obtained independently at several time-frequency pairs, and then used to find channel estimates elsewhere. However, the one-step approach, where the data used to generate the independent channel measurements is pooled together and used directly to generate channel estimates at all other times and frequencies, can also be analyzed using our framework. In the direct approach, instead of channel measurements, we have N_M bursts of pilot data:

$$\mathbf{Z}(\omega_n, t_n) = \mathbf{H}(\omega_n, t_n)\mathbf{S}_n + \tilde{\mathbf{N}}(\omega_n, t_n)\ ,\ n = 1, \cdots, N_M\ , \tag{4.28}$$

where $\mathbf{S}_n$ is an $M_t \times N_T$ matrix of pilot data and $\tilde{\mathbf{N}}(\omega_n, t_n)$ is the noise present during training. Note that Equation (4.28) implicitly assumes narrowband pilot data of short duration so that

the channel can be assumed to be approximately constant. Vectorizing (4.28) yields

$$\mathbf{z}(\omega_n, t_n) = (\mathbf{S}_n^T \otimes \mathbf{I})\mathbf{h}(\omega_n, t_n) + \tilde{\mathbf{n}}(\omega_n, t_n) \tag{4.29}$$

$$= \tilde{\mathbf{S}}_n \mathbf{h}(\omega_n, t_n) + \tilde{\mathbf{n}}(\omega_n, t_n) \ , \tag{4.30}$$

where we have defined $\tilde{\mathbf{S}}_n = (\mathbf{S}_n^T \otimes \mathbf{I})$. Stacking all N_M data vectors together as in (4.22), we end up with the model

$$\mathbf{z} = \begin{bmatrix} \tilde{\mathbf{S}}_1 & & \\ & \ddots & \\ & & \tilde{\mathbf{S}}_{N_M} \end{bmatrix} \mathbf{h} + \tilde{\mathbf{n}} = \tilde{\mathbf{S}}\mathbf{h} + \tilde{\mathbf{n}} \ , \tag{4.31}$$

where we assume that the noise term $\tilde{\mathbf{n}}$ is distributed as $\tilde{\mathbf{n}} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \tilde{\mathbf{C}})$.

Apart from the presence of the linear transformation $\tilde{\mathbf{S}}$, the form of Equation (4.31) is identical to that in (4.22), and the bound on the channel estimation error again follows from a straightforward application of the vector CRB and Bangs formula:

$$E\left[\left(\hat{\mathbf{h}}(\omega, t) - \mathbf{h}(\omega, t)\right)\left(\hat{\mathbf{h}}(\omega, t) - \mathbf{h}(\omega, t)\right)^H\right] \geq \mathbf{H}'\tilde{\mathbf{B}}\mathbf{H}'^H \ , \tag{4.32}$$

where

$$\left[\tilde{\mathbf{B}}^{-1}\right]_{ij} = \mathrm{tr}\left[\tilde{\mathbf{C}}^{-1}\frac{\partial \tilde{\mathbf{C}}}{\partial \theta_i}\tilde{\mathbf{C}}^{-1}\frac{\partial \tilde{\mathbf{C}}}{\partial \theta_j}\right] + 2\mathrm{Re}\left[\frac{\partial \mathbf{h}^H}{\partial \theta_i}\tilde{\mathbf{S}}^H\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{S}}\frac{\partial \mathbf{h}}{\partial \theta_j}\right] \ . \tag{4.33}$$

Thus, analysis of the one-step estimation approach follows directly from analysis of the two-step procedure presented below.

Under standard assumptions about the training data, noise, and channel estimation procedure, we can further show that the one- and two-step procedures yield exactly the same CRB. Assume $N_T \geq M_t$ and that the training data matrices $\mathbf{S}_n$ have full row rank. If the measurement noise is uncorrelated at different time-frequency pairs (ω_n, t_n), the FIM with respect to the channel parameters becomes

$$\left[\tilde{\mathbf{B}}^{-1}\right]_{ij} = \sum_{n=1}^{N_M} 2\mathrm{Re}\left[\frac{\partial \mathbf{h}_n^H}{\partial \theta_i}(\mathbf{S}_n^T \otimes \mathbf{I})^H \tilde{\mathbf{C}}_n^{-1}(\mathbf{S}_n^T \otimes \mathbf{I})\frac{\partial \mathbf{h}_n}{\partial \theta_j}\right], \tag{4.34}$$

where $\tilde{\mathbf{C}}_n = E[\tilde{\mathbf{n}}(\omega_n, t_n)\tilde{\mathbf{n}}(\omega_n, t_n)^H]$. If we also assume that $\tilde{\mathbf{C}}_n = \tilde{\sigma}_n \mathbf{I}$, then (4.34) simplifies further to

$$\left[\tilde{\mathbf{B}}^{-1}\right]_{ij} = \sum_{n=1}^{N_M} \frac{2}{\tilde{\sigma}_n} \mathrm{Re}\left[\frac{\partial \mathbf{h}_n^H}{\partial \theta_i}(\mathbf{S}_n^* \mathbf{S}_n^T \otimes \mathbf{I})\frac{\partial \mathbf{h}_n}{\partial \theta_j}\right]. \tag{4.35}$$

A standard approach to flat-fading channel estimation with training data is to apply the pseudo-inverse of $\mathbf{S}_n^\dagger$ to the right of $\mathbf{Z}(\omega_n, t_n)$ in (4.28), which results in the channel estimate

$$\hat{\mathbf{H}}(\omega_n, t_n) = \mathbf{Z}(\omega_n, t_n)\mathbf{S}_n^\dagger = \mathbf{H}(\omega_n, t_n) + \tilde{\mathbf{N}}(\omega_n, t_n)\mathbf{S}_n^\dagger. \tag{4.36}$$

In this case, the covariance of the two-step measurement noise becomes

$$\hat{\mathbf{C}}_n = (\mathbf{S}_n^{\dagger T} \otimes \mathbf{I})\tilde{\mathbf{C}}_n(\mathbf{S}_n^{\dagger *} \otimes \mathbf{I}) = \tilde{\sigma}_n(\mathbf{S}_n^{\dagger T}\mathbf{S}_n^{\dagger *} \otimes \mathbf{I}) \ , \tag{4.37}$$

and its inverse is given by

$$\hat{\mathbf{C}}_n^{-1} = \frac{1}{\tilde{\sigma}_n}(\mathbf{S}_n^* \mathbf{S}_n^T \otimes \mathbf{I}). \tag{4.38}$$

Plugging this into (4.26) and again assuming the measurement noise is uncorrelated from measurement to measurement, we obtain an FIM identical to (4.35). Thus, for the given assumptions, the one- and two-step procedures have identical CRBs. Note that if $N_T = M_t$, the same result may be shown for more general $\tilde{\mathbf{C}}_n$.

4.3.2 DOD/DOA Model

Using the CRB tools given by (4.24)-(4.26), we derive a lower bound for the estimation error using the DOD/DOA model without calibration errors. In order to evaluate (4.26), we explicitly write out $\mathbf{h}$, the sampled channel vector of (4.22). For the DOD/DOA model, $\mathbf{h}$ may be written as

$$\mathbf{h} = (\mathbf{I}_{N_M} \otimes (\mathbf{A}_t \otimes \mathbf{A}_r\mathbf{X}))\mathcal{W}, \tag{4.39}$$

where

$$\mathcal{W} = \begin{bmatrix} \text{vec}(\mathbf{W}(\omega_1, t_1)) \\ \text{vec}(\mathbf{W}(\omega_2, t_2)) \\ \vdots \\ \text{vec}(\mathbf{W}(\omega_{N_M}, t_{N_M})) \end{bmatrix}. \tag{4.40}$$

The evaluation of the derivatives in (4.25) and (4.26) is straightforward but tedious, and the derivations are left to Appendices B.1 and B.2. The form of the resulting FIM with respect to $\mathbf{\Theta}$ for the DOD/DOA model can be expressed as a collection of submatrices as follows:

$$\mathbf{J}(\mathbf{\Theta}) = \begin{bmatrix} \frac{M_t M_r N_M}{\sigma^2} & \mathbf{0}^T \\ \mathbf{0} & \mathbf{J}_{22} \end{bmatrix}, \tag{4.41}$$

where

$$\mathbf{J}_{22} = \begin{bmatrix} \text{Re}[\mathbf{P}_1] & -\text{Im}[\mathbf{P}_1] & \text{Re}[\mathbf{P}_2] & \text{Re}[\mathbf{P}_3] & \text{Re}[\mathbf{P}_4] & \text{Re}[\mathbf{P}_5] \\ \text{Im}[\mathbf{P}_1] & \text{Re}[\mathbf{P}_1] & \text{Im}[\mathbf{P}_2] & \text{Im}[\mathbf{P}_3] & \text{Im}[\mathbf{P}_4] & \text{Im}[\mathbf{P}_5] \\ \text{Re}[\mathbf{P}_2^T] & \text{Im}[\mathbf{P}_2^T] & \text{Re}[\mathbf{P}_6] & \text{Re}[\mathbf{P}_7] & \text{Re}[\mathbf{P}_8] & \text{Re}[\mathbf{P}_9] \\ \text{Re}[\mathbf{P}_3^T] & \text{Im}[\mathbf{P}_3^T] & \text{Re}[\mathbf{P}_7^T] & \text{Re}[\mathbf{P}_{10}] & \text{Re}[\mathbf{P}_{11}] & \text{Re}[\mathbf{P}_{12}] \\ \text{Re}[\mathbf{P}_4^T] & \text{Im}[\mathbf{P}_4^T] & \text{Re}[\mathbf{P}_8^T] & \text{Re}[\mathbf{P}_{11}^T] & \text{Re}[\mathbf{P}_{13}] & \text{Re}[\mathbf{P}_{14}] \\ \text{Re}[\mathbf{P}_5^T] & \text{Im}[\mathbf{P}_5^T] & \text{Re}[\mathbf{P}_9^T] & \text{Re}[\mathbf{P}_{12}^T] & \text{Re}[\mathbf{P}_{14}^T] & \text{Re}[\mathbf{P}_{15}] \end{bmatrix}. \tag{4.42}$$

Note that σ has been added as the first element of $\mathbf{\Theta}$ in this and all subsequent models. The individual blocks in the above expression have a simple form; for example, $\mathbf{P}_1$ is given by

$$\mathbf{P}_1 = \frac{2}{\sigma} \sum_{n=1}^{N_M} \mathbf{W}_n^H (\mathbf{A}_t^H \mathbf{A}_t) \odot (\mathbf{A}_r^H \mathbf{A}_r) \mathbf{W}_n, \tag{4.43}$$

where $\odot$ denotes an element-wise matrix (Hadamard) product and $\mathbf{W}_n = \mathbf{W}(\omega_n, t_n)$. The remainder of the CRB submatrices are given in Appendix B.3.1. These subblocks are easily calculated for any specific channel or array configuration. Numerical examples of the resulting bounds are presented later in Section 4.4.

4.3.3 DOD/DOA Model with Calibration Errors

For the DOD/DOA model with calibration errors, the measurement vector $\mathbf{h}$ is given by

$$\mathbf{h} = (\mathbf{I}_{N_M} \otimes ((\mathbf{A}_t + \mathbf{V}_t) \otimes (\mathbf{A}_r + \mathbf{V}_r)\mathbf{X}))\mathcal{W}. \tag{4.44}$$

In this paper, the calibration errors $\mathbf{V}_t$ and $\mathbf{V}_r$ are assumed to be ignored during the channel estimation process, resulting in a biased channel estimator. Therefore, the bound given by the CRB formula in (4.24) will not apply. However, we may still find a corresponding bound for the biased case. To do this, we assume

$$\begin{aligned} E\left[\hat{\mathbf{h}}(\omega,t)\right] &= \mathbf{h}(\omega,t) + \mathbf{b}(\omega,t) \\ &= (\mathbf{A}_t \otimes \mathbf{A}_r\mathbf{X})\mathrm{vec}(\mathbf{W}(\omega,t)), \end{aligned} \tag{4.45}$$

where $\mathbf{h}(\omega,t)$ is the DOD/DOA channel model with calibration errors and $\mathbf{b}(\omega,t)$ is the estimation bias given by

$$\mathbf{b}(\omega,t) = (\mathbf{A}_t \otimes \mathbf{A}_r\mathbf{X})\mathrm{vec}(\mathbf{W}(\omega,t)) - \mathbf{h}(\omega,t). \tag{4.46}$$

Then the CRB may be found as

$$E\left[\mathbf{e}(\omega,t)\mathbf{e}(\omega,t)^H\right] \geq \mathbf{H}'\mathbf{B}\mathbf{H}'^H + \mathbf{b}\mathbf{b}^H, \tag{4.47}$$

where $\mathbf{B}$ is the CRB matrix with respect to the parameters and $\mathbf{H}'$ is the Jacobian matrix

$$\mathbf{H}' = \left[\begin{array}{ccc} \frac{\partial(\mathbf{h}(\omega,t)+\mathbf{b}(\omega,t))}{\partial\theta_1} & \cdots & \frac{\partial(\mathbf{h}(\omega,t)+\mathbf{b}(\omega,t))}{\partial\theta_P} \end{array}\right]. \tag{4.48}$$

The time and frequency dependent bias term accounts for the effects of the calibration errors on the bound. Matrix $\mathbf{B}$, the CRB for $\boldsymbol{\Theta}$, may be calculated by applying Bangs formula to the sampled channel model $\mathbf{h}$ of (4.44).

If we further assume that the calibration errors are random variables drawn from a particular distribution, e.g., $\mathrm{vec}(\mathbf{V}_t) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \sigma_t\mathbf{I})$ and $\mathrm{vec}(\mathbf{V}_r) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \sigma_r\mathbf{I})$, then we may

remove specific calibration error terms from the CRB, obtaining

$$E\left[\mathbf{e}(\omega,t)\mathbf{e}(\omega,t)^H\right] \geq \mathbf{H}'\mathbf{B}\mathbf{H}'^H + E\left[\mathbf{b}\mathbf{b}^H\right], \tag{4.49}$$

where $\mathbf{B}$, the CRB with respect to the parameters, becomes

$$\mathbf{B}^{-1} = E[\mathbf{J}(\boldsymbol{\Theta})]. \tag{4.50}$$

The expectations in (4.49) and (4.50) are with respect to the calibration and measurement error distributions. A derivation of this modified CRB is available in Appendix A. The CRB expression found for the calibration error model using the above biased CRB formula is given in Appendix B.3.2. Note that the bound of (4.49) is not a Bayesian or posterior CRB since the calibration errors are not considered channel parameters in this scenario. The expectation over the calibration errors removes the dependence of the bound on specific calibration error realizations, but does not qualitatively alter the bound of (4.49) relative to (4.47).

4.3.4 VSS CRB

The VSS measurement vector $\mathbf{h}$ is given by

$$\mathbf{h} = (\mathbf{I}_{N_M} \otimes (\mathbf{A}_t \otimes \mathbf{A}_r))\mathcal{W}. \tag{4.51}$$

The results of applying (4.24)-(4.26) to this model are detailed in Appendix B.3.3.

4.3.5 MSS CRB

The MSS measurement vector $\mathbf{h}$ is given by

$$\begin{aligned}\mathbf{h} &= (\mathbf{I}_{N_M} \otimes (\mathbf{I}_{M_t} \otimes \mathbf{A}))\mathcal{B} \\ &= (\mathbf{I}_{N_M M_t} \otimes \mathbf{A})\mathcal{B},\end{aligned} \tag{4.52}$$

where $\mathbf{A}$ is given in (4.18),

$$\mathcal{B} = \begin{bmatrix} \text{vec}(\mathbf{w}(\omega_1, t_1)) \\ \text{vec}(\mathbf{w}(\omega_2, t_2)) \\ \vdots \\ \text{vec}(\mathbf{w}(\omega_{N_M}, t_{N_M})) \end{bmatrix}, \tag{4.53}$$

and $\mathbf{w}(\omega_n, t_n)$ is given in (4.19). Application of (4.24)-(4.26) to this model is presented in Appendix B.3.4.

4.4 Numerical Simulations

In this section, we explore the limiting performance of MIMO-OFDM channel estimation and prediction by numerically evaluating the derived bounds for several scenarios. In examining the estimation error performance via the CRB, a natural measure is the Root Mean Square Error

$$\text{RMSE}(\omega, t) = \sqrt{\frac{E[\|\mathbf{e}(\omega, t)\|_2^2]}{E[\|\mathbf{H}(\omega, t)\|_F^2]}}, \tag{4.54}$$

where (ω, t) is the time and frequency at which the channel is estimated or predicted. Unlike previous expressions, the expected values in (4.54) are over not only the measurement error, but the channel realizations as well. Thus, this measure represents a normalized average (over the channel and the measurement noise) error of the elements of $\mathbf{h}(\omega, t)$, and it allows for a direct and fair comparison of the error performances of MIMO systems with various array sizes, including the limiting SISO case. Substituting the bound of (4.27) into (4.54), we obtain the normalized RMSE bound

$$\text{RMSE}(\omega, t) \geq \sqrt{\frac{E\left[\text{tr}[\text{CRB}(\omega, t)]\right]}{E[\|\mathbf{H}(\omega, t)\|_F^2]}}, \tag{4.55}$$

where we use $\text{CRB}(\omega, t)$ to denote the CRB for any of the models described above.

In the simulations that follow, the expected value in the numerator of (4.55) is computed by averaging the estimation error bound over 500 independent channel realizations. The channel realizations are generated using the DOD/DOA channel model. To obtain these realizations, the channel parameters for each realization are selected as follows. We begin

by assuming that the different channel parameters composing $\boldsymbol{\Theta}$ are independent, a justifiable approximation when considering the channel model. The scattering parameters $\boldsymbol{\alpha}$ are chosen as independent circular-symmetric complex Gaussian random variables distributed as $\alpha_l \sim \mathbb{CN}(0,1)$. The path delays $\boldsymbol{\tau}$ are selected from an exponential distribution such that approximately 98% of the τ_l fall in the delay range of $(100-1000)\lambda/c$, where λ is the wavelength at $\omega_c = 2\pi(2.4\text{ GHz})$ and c is the speed of light in free space, i.e., 98% of the τ_l fall between $0.042\mu s$ and $0.42\mu s$. The physical DODs and DOAs for the antenna arrays are drawn from a uniform distribution so that $\phi_{t,i}, \phi_{r,i} \sim U[0, 2\pi)$, and the solid angles are given by the formula $\Omega_{\cdot,l} = k(\omega)d_{\cdot} \sin\phi_{\cdot,l}$ with $d_{\cdot} = \lambda_c/2$, where λ_c is the wavelength at ω_c. Note that $k(\omega)$, and therefore $\boldsymbol{\Omega}_t$ and $\boldsymbol{\Omega}_r$, are frequency dependent. The Doppler frequency of path l is derived from a physical viewpoint to be $\omega_{d,l} = k(\omega)\frac{\Delta_s}{T_s}\sin\phi_{d,l}$, where $\phi_{d,l}$ is the angle between the propagation path l and the direction of array motion, and Δ_s and T_s are the distance and time separating consecutive channel measurements, respectively. We assume $\phi_{d,l} \sim U[0, 2\pi)$. Since $\frac{\Delta_s}{T_s}$ is the rate of motion of the antenna array, we may choose either the sample spacings and sample rate or the rate of array motion to complete the specification of $\omega_{d,l}$. The simulations use a rate of motion of 5m/s. Using this, the sample rate may be determined from the measurement spacing discussed in the following. Like $\boldsymbol{\Omega}_t$ and $\boldsymbol{\Omega}_r$, the values chosen for ω_d are frequency dependent. Finally, we assume that the calibration errors are Gaussian distributed with $\text{vec}(V_t) \sim \mathbb{CN}(\mathbf{0}, \sigma_t\mathbf{I})$ and $\text{vec}(V_r) \sim \mathbb{CN}(\mathbf{0}, \sigma_r\mathbf{I})$.

With the assumptions above, the denominator of (4.55) simplifies to $LM_tM_r(1+\sigma_t)(1+\sigma_r)$. To avoid the need of adjusting the channel-to-noise power ratio (CNR), which is defined from (4.21) to be

$$CNR = \frac{E\left[\|\mathbf{h}(\omega,t)\|_2^2\right]}{E\left[\|\mathbf{n}(\omega,t)\|_2^2\right]}, \tag{4.56}$$

for different levels of calibration error, we normalize the average power for each of the elements in the array vectors (array plus calibration error) to unity. Now we may express the normalized error bound as

$$\text{RMSE}(\omega,t) \geq \sqrt{\frac{E\left[\text{tr}[\text{CRB}(\omega,t)]\right]}{LM_tM_r}}. \tag{4.57}$$

The CNR reduces to

$$CNR = \frac{L}{\sigma}. \tag{4.58}$$

A realistic channel model often consists of several multipath components. For example, outdoor measurements taken in downtown Austin, TX, were generally well described by 3-8 paths [53]. Unless otherwise stated, the numerical simulations in this section use $L = 6$ paths in the channel models. In the simulation results, the time variable will be given in terms of wavelengths (position or distance) to allow for direct comparisons with previous work and to abstract the results away from mobile velocity and towards the physical environment. The results may be easily converted into units of time by assuming a particular mobile velocity and channel sampling rate. For most of the simulations, the N_M pilot-based channel measurements are selected or sampled evenly over space and frequency: 16 measurements across a 20 MHz frequency bandwidth and 32 in space between -10 and 0 wavelengths for a total of $N_M = 512$ channel measurements. Thus, results for wavelengths less than zero correspond to channel interpolation performance, while those for wavelengths greater than zero correspond to channel prediction. The measurement noise power is -20dB relative to unity per receive antenna.

4.4.1 The Channel Models

Figure 4.1 is an example of the 2-dimensional (2D) normalized error bound using a 3×3 MIMO antenna configuration for DOD/DOA and VSS channel models where the upper surface is the VSS bound. For the sake of clarity, the remaining results will be presented as position (time) and frequency slices, the position slices at $\omega - \omega_c = 0$ MHz, the midpoint of the frequency measurements, and the frequency slices at -5λ, the mid-point of the time measurements. A representative frequency slice of the normalized error bounds for a 2×2 MIMO configuration is shown in Fig. 4.2. In the figure, we see that the ideal DOD/DOA model provides the best performance in terms of the error bound, which is expected since it provides the most parsimonious parametrization of the channel. The addition of -60 dB of calibration error (value of $\sigma_r = \sigma_t$ relative to the gain of the ULA elements) results in a slight increase in the bound. However, a rapid increase in the bound is observed as the calibration

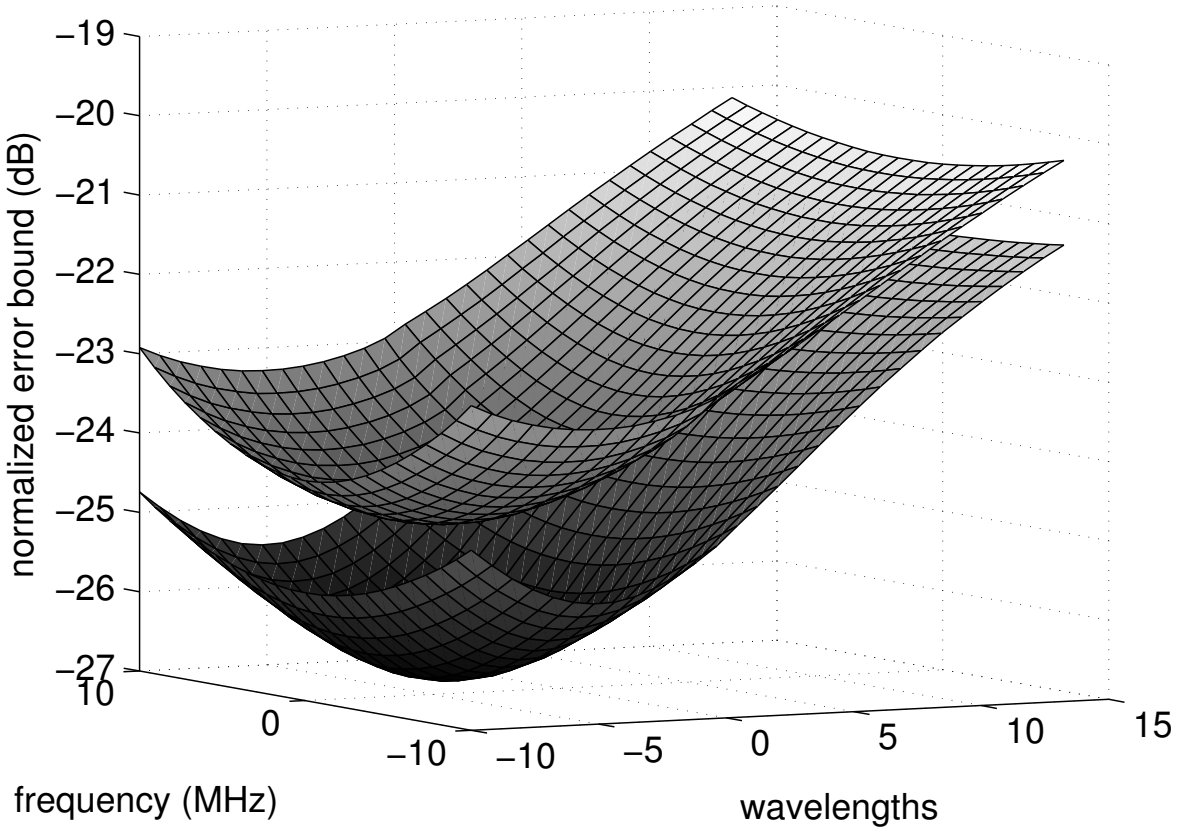


Figure 4.1: Example of 2D normalized error bound results for the DOD/DOA and VSS models.

error is increased further. A calibration error level of -50 dB is enough to result in the poorest performance for any of the bounds. The bound for all models increases towards the band edge since fewer nearby channel measurements are available for interpolation. The VSS and MSS bounds fall between the two calibration error cases, with the VSS model having a performance edge over MSS in the center of the bandwidth with decreasing performance towards the band-edges. Recall that the ULA array responses are frequency dependent via the wavenumber k, which explains why the VSS performance degrades more quickly away from the band center. On the other hand, the MSS bound is higher than VSS at the band center because it does not exploit the rank-one nature of the channel taps, but it is less sensitive to the frequency dependence of the channel because of the additional modeling degrees of freedom it has available.

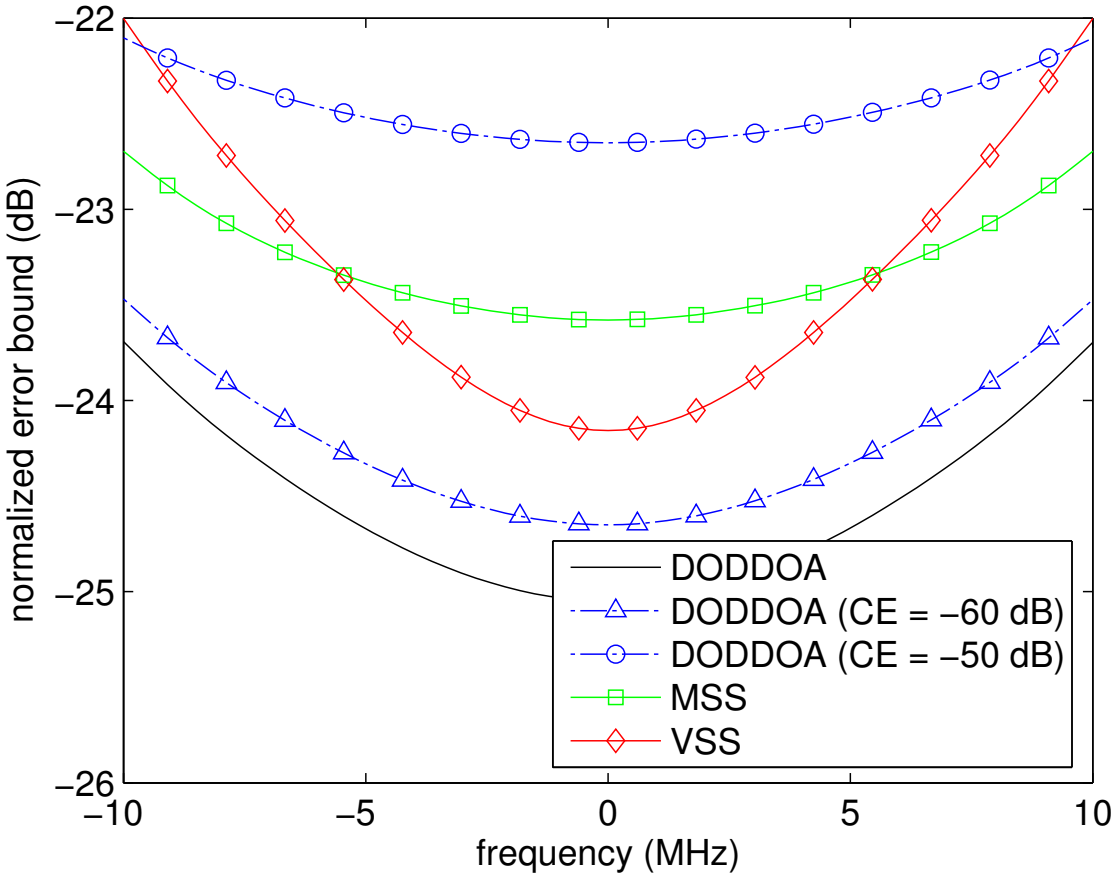


Figure 4.2: Comparison of bounds for all discussed models in frequency.

Figure 4.3 is a position (time) slice corresponding to the same overall 2D results used to produce Fig. 4.2. In the estimation region of the plot, the appearance of the bound slices are very similar to their frequency counterparts. This similarity is due to the duality in the form of the time and frequency dependence in (4.1). However, in this case the VSS bound remains lower than the MSS bound as a function of position since the array responses are assumed to be time-invariant. As expected, in the prediction region, the bound for the DOD/DOA model without calibration errors is lower than the others.

4.4.2 Calibration Error

The results in Figs. 4.2 and 4.3 suggest that the DOD/DOA model has an extreme sensitivity to array calibration errors. The normalized error bounds are used to explore this

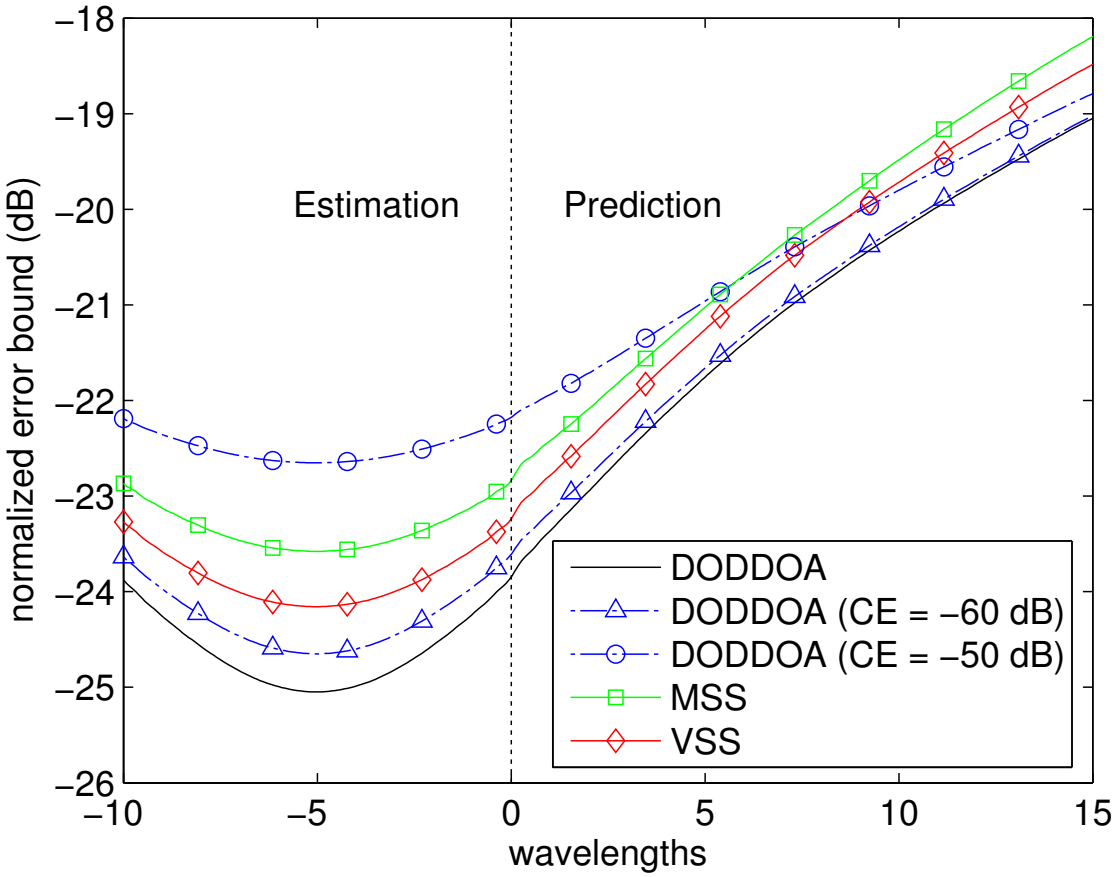


Figure 4.3: Comparison of bounds for all discussed models as a function of wavelength.

issue further in Fig. 4.4. Each point in the curves in this plot represent the lowest bound point from the 3D error bound surfaces for a 2×2 MIMO configuration, which occur at (0 MHz, -5λ). The results in the figure demonstrate the robustness of the VSS model with respect to calibration errors; around ω_c the VSS model performs equally well regardless of the underlying array structure. This is a significant advantage of the VSS model, particularly in situations when the array structure may be in doubt or calibration errors are present. The DOD/DOA model, on the other hand, is shown to be extremely sensitive to even small amounts of calibration error. This suggests that unless the calibration errors in a system can be accurately accounted for prior to channel estimation, the DOD/DOA model should be abandoned in favor of the VSS model for parametric channel estimation. Also included

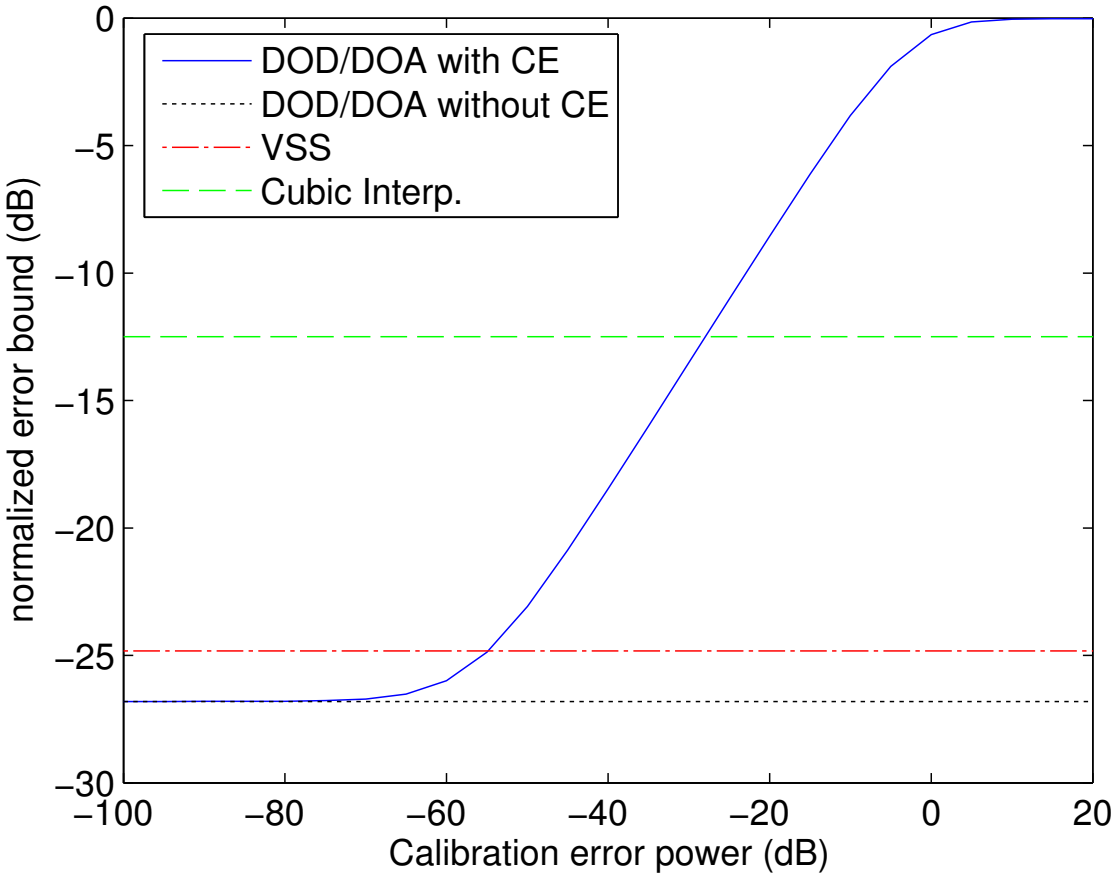


Figure 4.4: Examination of the normalized error bound as a function of calibration error.

in the plot as a reference is the average normalized error performance achieved when using cubic interpolation to estimate the channel from the channel measurements.

Recall that these simulations assume a measurement noise power of -20 dB. Though not plotted here, we note that changing the measurement noise power level results in the raising or lowering of the calibration-error-free bound levels, without significantly altering the slope and position of the calibration-error-dependent portions of the plot. Thus, channels with lower measurement noise are more sensitive to small amounts of calibration error, while the converse is true for higher measurement noise power.

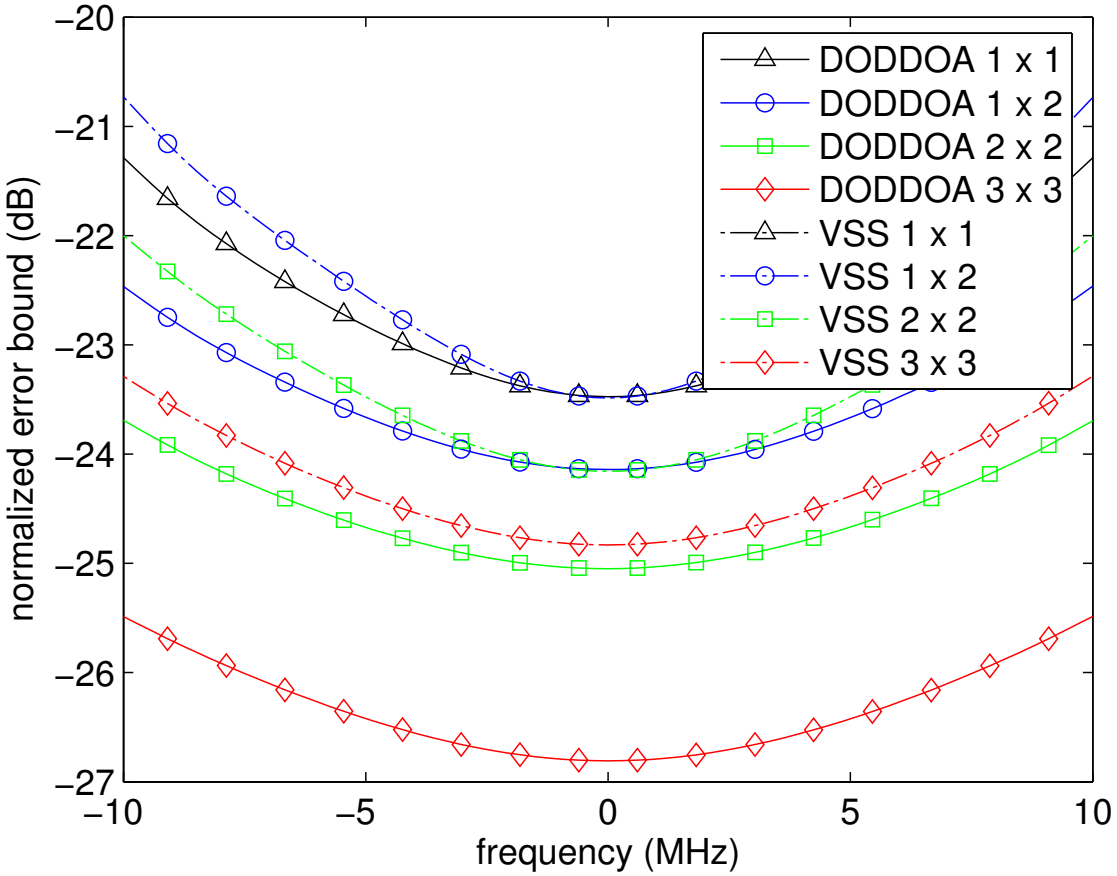


Figure 4.5: Plot of the normalized error bound in frequency for various array sizes.

4.4.3 Array Sizes

We now examine the impact of the array sizes M_t and M_r on the normalized error bounds. Figure 4.5 displays a frequency slice of the DOD/DOA and VSS error bounds for a SISO, a 1×2 single input multiple output (SIMO), and 2×2 and 3×3 MIMO configurations. Note that DOD/DOA and VSS models are not uniquely identifiable for the SISO case since the array parameters cannot be estimated. Practically, however, the DOD/DOA and VSS models reduce to the same identifiable SISO channel, and their performance is identical for this scenario. It is clear that significant gains in channel estimation performance may be achieved through the use of an increased number of antennas at the transmitter and receiver. The one exception to this in the plot is the 1×2 SIMO VSS system, whose bound is higher than for the SISO system as a result of the extra and unnecessary intermediate step of

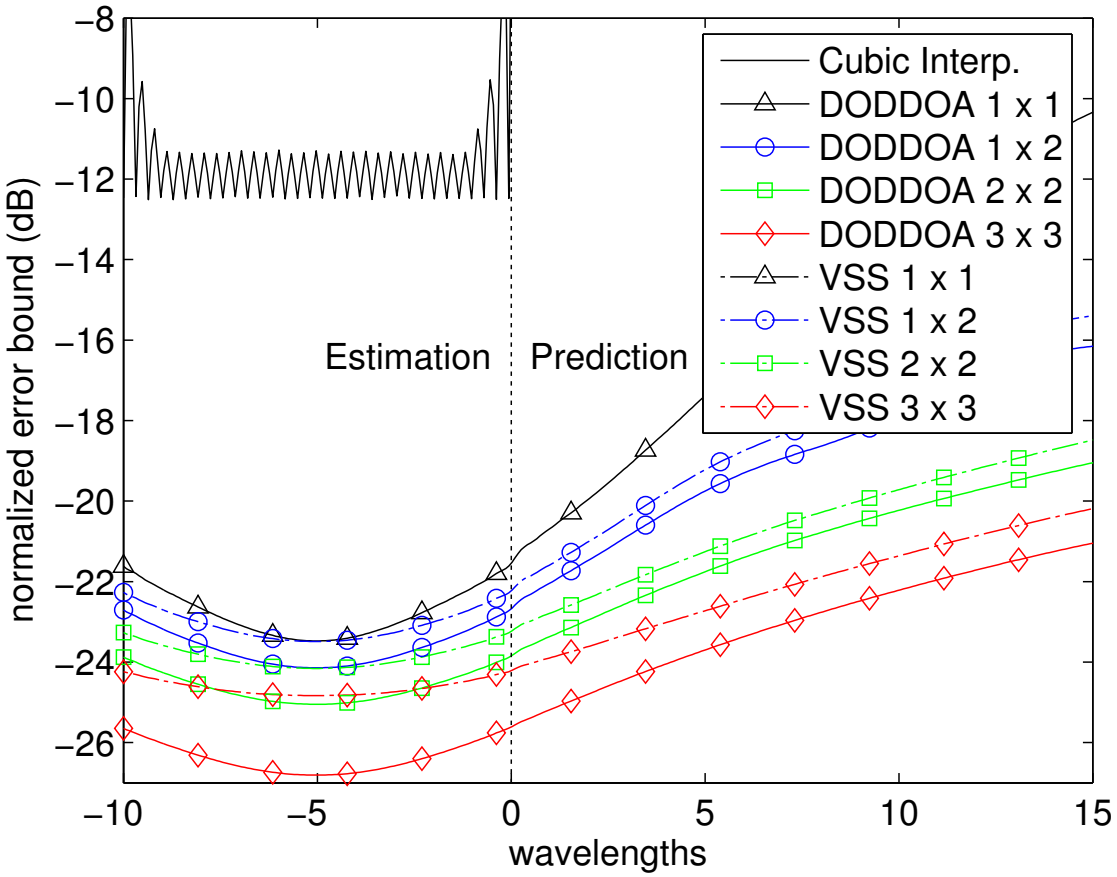


Figure 4.6: Plot of the normalized error bound versus wavelength for various array sizes.

estimating the transmit array element. The SIMO DOD/DOA bound was formulated to omit this extra step, and therefore does not suffer the same penalty. Overall, these results are in harmony with those obtained with the wideband time-invariant bounds developed in [91].

Nearly identical results are seen in the estimation portion of the position slice in Fig. 4.6. Even greater benefits due to MIMO arrays are seen in the prediction portion of this plot. The results indicate that both the DOD/DOA and VSS MIMO systems may be predicted much farther into the future than the corresponding SISO and SIMO systems. As was suggested in [60], this increase in performance is intuitively explained by noting that the larger arrays reveal more of the underlying channel structure, allowing for a better characterization of the channel parameters. This advantage is maintained even when the

number of channel measurements N_M is adjusted to be proportional to $1/M_t$, allowing for a fairer comparison for the given receive CNR of -20 dB. Also included in this plot is an example of the average normalized error performance when 2D cubic interpolation is used to estimate the channel from the measurement segment. The low points in the curve correspond to the locations of the channel measurements in time. It is clear that estimation of the channel through a parametric approach offers dramatic gains over simple unstructured interpolation schemes.

The improvement in channel estimation performance with increasing array size suggests that it may be possible to use fewer pilot channels in MIMO-OFDM systems compared with SISO systems for the same performance level. To illustrate how the number of pilot subcarriers affect the normalized error bounds, Fig. 4.7 shows VSS bounds corresponding to a 1×2 SIMO configuration with the number of pilot subcarriers (frequency measurements) ranging from 4 to 32. The markers in the plot indicate the frequencies at which direct channel estimates are obtained using the pilots. Also displayed on the plot are the coherence bandwidths measured as the frequency range for which the coherence function is either above 0.9 or 0.5. As expected, having sufficiently dense pilot channels is vital to channel interpolation performance. In Fig. 4.8, the number of pilot subcarriers are reduced from 16 to 8 (there are now $N_M = 8 * 32 = 256$ channel measurements), and we plot the normalized error bounds for various array sizes. When smaller arrays are used, the number of samples is insufficient to achieve interpolated channel estimates of similar quality to those obtained by the pilots. However for larger arrays, this frequency sampling density is adequate.

4.4.4 Number of Multipath Components

In all of the numerical simulations presented so far, the number of multipath components of the channel was chosen to be $L = 6$. Our final results explore the dependence of the normalized error bounds on the number of paths. For the following results, the noise power is adjusted as L is varied to maintain a constant CNR equivalent to that for the -20 dB noise, $L = 6$ case used in the previous simulations. Figures 4.9 and 4.10 are frequency and position slice plots, respectively, of the normalized error bounds for 2×2 DOD/DOA and VSS models with L ranging from 2 to 10 paths. The results in the figures indicate that as

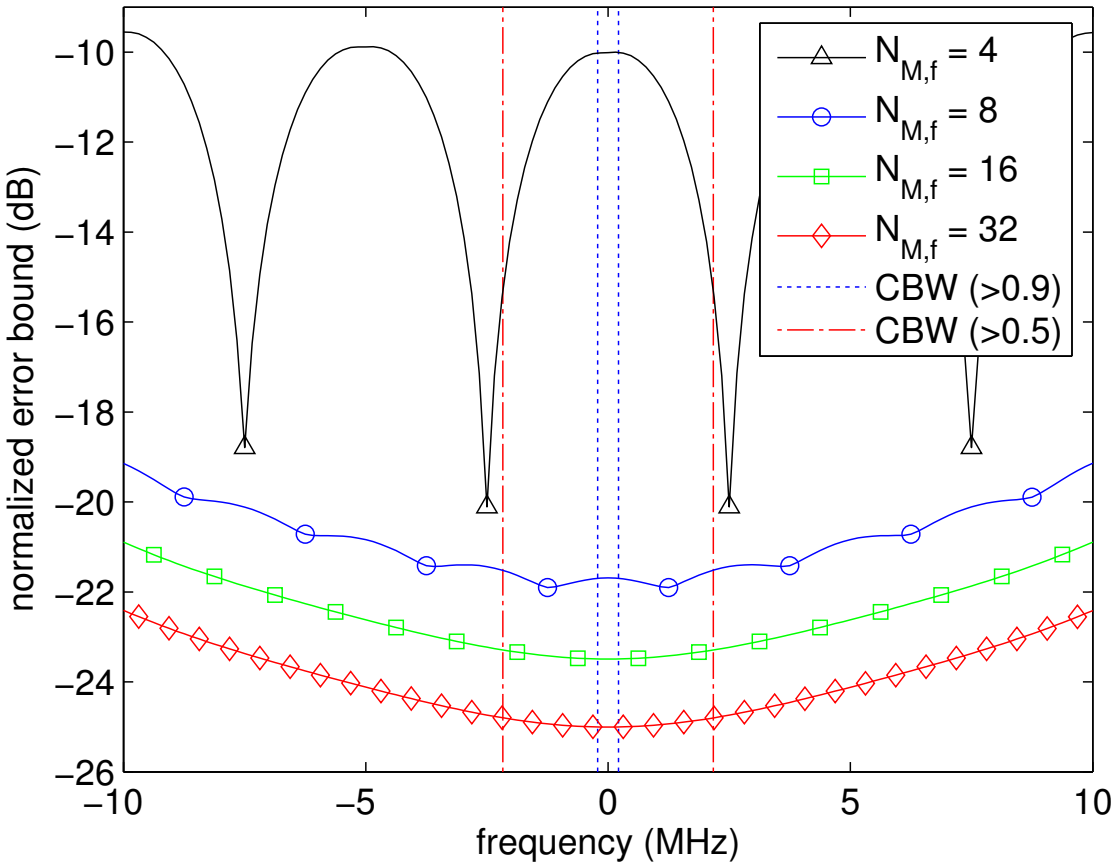


Figure 4.7: The effect of frequency measurement spacing on performance.

the number of multipath components of a channel increases, the possible channel estimation performance is reduced. It is interesting to note that while increasing the number of paths raises the level of the normalized error bound, it does not significantly alter the bound slope. Thus, increasing the number of channel paths in the model does not severely impact the predictability of MIMO-OFDM channels.

4.5 Conclusions

In this chapter, we have presented lower bounds for several parametric mobile wideband MIMO-OFDM channel models. These bounds were derived using vector formulations for the Cramér-Rao lower bound for functions of parameters for both unbiased and biased estimators. The Fisher information matrices in the CRB formulations were noted to have a

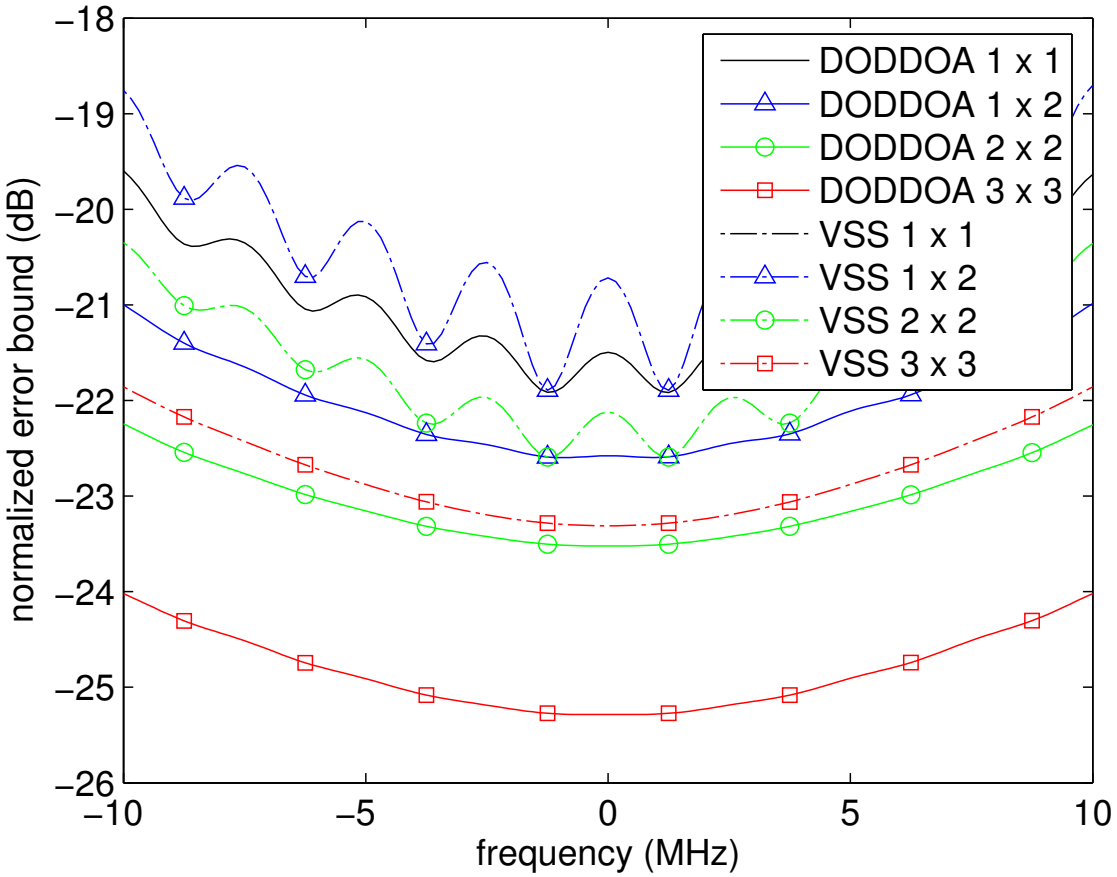


Figure 4.8: The effect array sizes and frequency measurement spacing on performance.

simple block structure allowing for convenient representation and calculation of the bounds. Numerical evaluations of these bounds demonstrated some interesting features regarding the estimation and prediction of MIMO-OFDM channels. First, our analysis indicates that better channel estimation and prediction performance can be achieved through the use of antenna arrays at the transmitter and receiver. A consequence of this fact is that a MIMO-OFDM system could function with a lower time and frequency pilot density than a corresponding SISO implementation, even taking into account the fact that more pilot symbols are required to estimate a MIMO channel. Second, our results suggest that when suitable, the use of parametric channel modeling provides a significant advantage in estimation and prediction performance, particularly when compared with simple unstructured interpolation schemes for estimating the channel. Finally, we have shown that channel estimation methods based

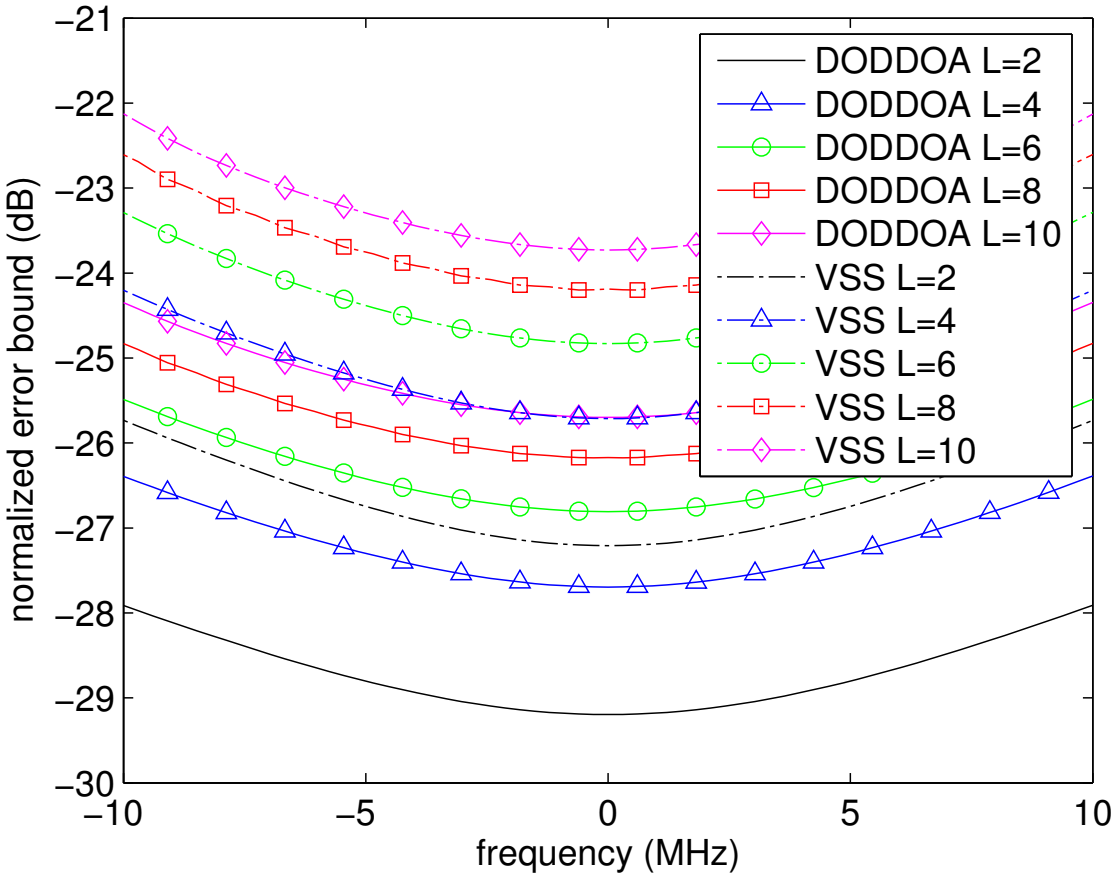


Figure 4.9: The normalized error bound for various path numbers L in frequency.

on DOD/DOA parameterizations are extremely sensitive to array calibration errors, and that approaches based on either vector or matrix spatial signatures are significantly more robust despite the imprecise way in which they characterize the frequency dependence of the channel. In particular, the VSS channel model appears to be a reasonable compromise in terms of complexity, performance and robustness.

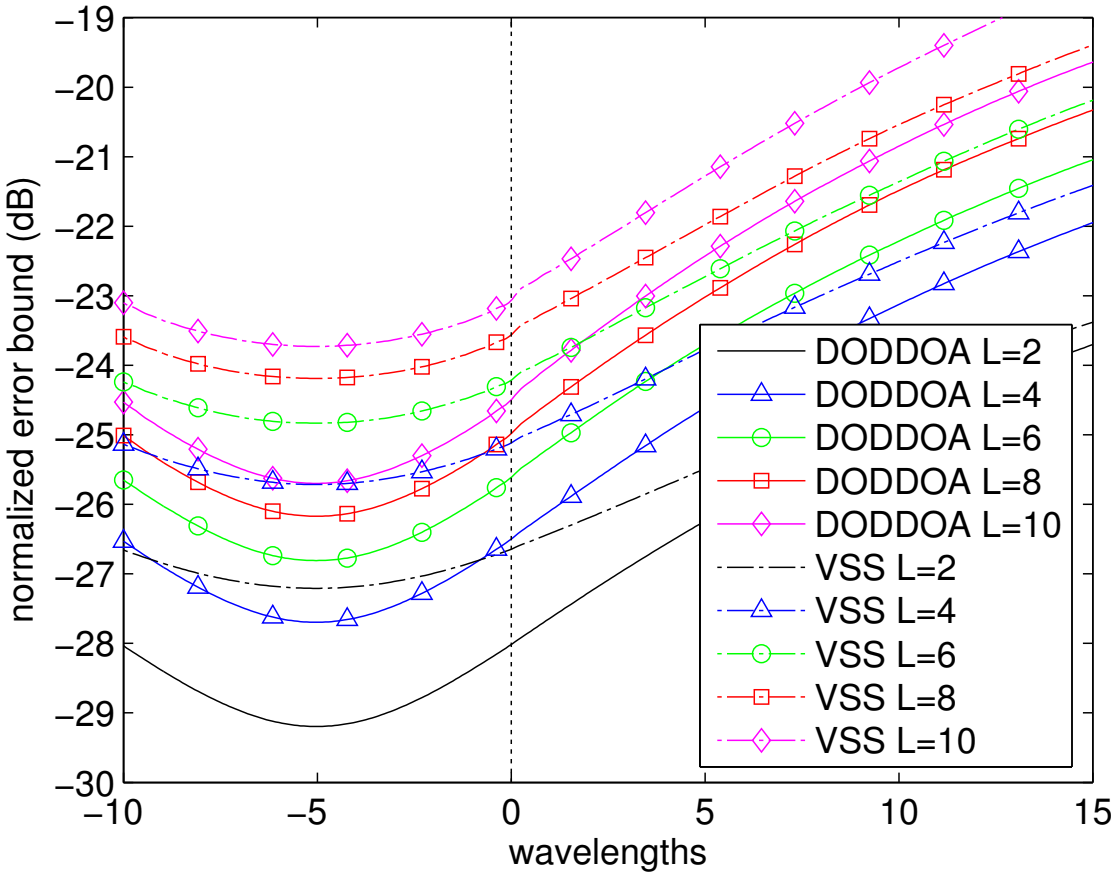


Figure 4.10: The normalized error bound for various path numbers L versus wavelengths.

Chapter 5

Power Allocation and Bit Loading for Spatial Multiplexing with Imperfect CSI

5.1 Introduction

In this chapter, we develop a closed-form expression for the subchannel SINR when using imperfect CSI in SVD-based signaling schemes. This result generalizes the SINR derivation of [64] by allowing for non-uniform power allocation and correcting an overestimation of the interference power due to CSI inaccuracies. In addition, this new expression allows for different levels of CSI error at the transmitter and receiver, whereas previous methods typically assume the CSI is identical at both ends of the communications link or that perfect CSI is available at the receiver. Our SINR expressions are developed with the aid of the first-order SVD perturbation analysis of [77] in a manner similar to the previous work of [94]. Note that the first-order analysis of [77], unlike previous SVD perturbation studies [95–99], includes in its analysis the effects of the subchannel interferences between singular vectors within the signal subspace, making the results directly applicable to the present problem. Once obtained, our SINR expressions allow us to (1) analyze the impact of imperfect CSI on the MIMO SVD subchannels; (2) perform robust bit-loading design through the systematic selection of the subchannel power levels; and (3) find thresholds for the amount of channel uncertainty and measurement noise above which beamforming should be used instead of spatial multiplexing over the SVD-based subchannels. For the special case of identical CSI at both ends of the link, the SINR expression allows us to derive approximately optimal subchannel power levels for M-QAM signaling under a subchannel SER constraint, a result directly applicable to robust bit-loading design.

The remainder of this chapter will proceed as follows. In Section 5.2, the MIMO channel and signaling model are presented and subchannel SINR expressions are derived. These

expressions are then used in Sections 5.3 and 5.4 to develop subchannel power allocation schemes and to find thresholds for the channel perturbation levels tolerable for multiplexing. Finally, in Section 5.5, numerical simulations are used to demonstrate the accuracy of the derived expressions and their usefulness when applied to the MIMO SVD-based bit-loading problem.

5.2 Derivation of the Subchannel SINRs

Consider a narrow-band MIMO communications system employing N_t transmit antennas and N_r receive antennas with CSI as discussed in Chapter 2.5, which we now review here. The available CSI is obtained through channel estimation at the receiver, presumably with the aid of training data sent from the transmitter, and is periodically made available to the transmitter through feedback. Imperfections in the CSI arise as a result of both the estimation and feedback processes. Mathematically, we describe this scenario by letting $\mathbf{H}$ be the channel as known at the transmitter, $\mathbf{H} + \Delta\mathbf{H}_1$ the channel as known at the receiver, and $\mathbf{H} + \Delta\mathbf{H}_1 + \Delta\mathbf{H}_2$ the actual MIMO channel, where the preceding matrices all have dimension $N_r \times N_t$. In this context, the matrix $\Delta\mathbf{H}_1$ represents the discrepancies between the actual channel and the transmit CSI that are unique to the transmitter alone. The source of these imperfections may include quantization effects and errors in the feedback channel as well as variations in the MIMO channel when assuming that the receiver updates its CSI more frequently than the transmitter. Channel variations considered in this context may occur, for example, if the transmitter or receiver moves during the interval between channel estimation and channel feedback. On the other hand, the matrix $\Delta\mathbf{H}_2$ represents CSI errors that are common to both the transmitter and receiver. Such errors are typically introduced at the receiver and are then fed back to the transmitter. Factors contributing to this term include estimation errors and outdated CSI present at the receiver. When $\Delta\mathbf{H}_1 = 0$, the transmitter and receiver possess the same CSI. We refer to this special case as the common CSI (CCSI) case. The general case with $\Delta\mathbf{H}_1 \neq 0$ is referred to as the mismatched CSI (MCSI) case.

The transmitter CSI matrix $\mathbf{H}$ has the following singular value decomposition (SVD):

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \,, \tag{5.1}$$

where $\mathbf{U}$ and $\mathbf{V}$ are respectively $N_r \times F$ and $N_t \times F$, $\mathbf{\Sigma} = \text{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_F\}$, $F = \min\{N_r, N_t\}$, and $(\cdot)^H$ denotes the Hermitian tranpose. The singular values are assumed to all be unique, and are ordered in the normal way: $\sigma_1 > \sigma_2 > \cdots > \sigma_F$. The SVD of the receiver CSI matrix $\mathbf{H} + \Delta\mathbf{H}_1$ is defined to be

$$\mathbf{H} + \Delta\mathbf{H}_1 = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^H \,, \tag{5.2}$$

where the terms in (5.2) are defined analogously to those in (5.1). Viewing the receiver CSI as a perturbed version of $\mathbf{H}$, the matrices on the right-hand side of (5.2) are related to those on the right-hand side of (5.1) by the following perturbation relationships:

$$\mathbf{U}_1 = \mathbf{U} + \Delta\mathbf{U}_1, \tag{5.3}$$

$$\mathbf{\Sigma}_1 = \mathbf{\Sigma} + \Delta\mathbf{\Sigma}_1, \tag{5.4}$$

and

$$\mathbf{V}_1 = \mathbf{V} + \Delta\mathbf{V}_1. \tag{5.5}$$

Given the above CSI and definitions, consider the following narrow-band MIMO SVD-based signaling scheme:

$$\mathbf{y} = \mathbf{U}_1^H \left[(\mathbf{H} + \Delta\mathbf{H}_1 + \Delta\mathbf{H}_2)\mathbf{V}\mathbf{P}\mathbf{x} + \mathbf{n}\right]. \tag{5.6}$$

The matrix $\mathbf{P}$ is a real-valued $F \times F$ diagonal matrix whose entries are the square-roots of the power assigned to the elements of the signal vector $\mathbf{x}$, which is assumed to be normalized and composed of uncorrelated entries so that $E\left[\mathbf{x}\mathbf{x}^H\right] = \mathbf{I}$, where $\mathbf{I}$ is an $F \times F$ identity matrix and $E[\cdot]$ denotes expectation. The vector $\mathbf{n}$ represents additive spatially and temporally

white noise: $E[\mathbf{n}] = 0$, $E\left[\mathbf{n}\mathbf{n}^H\right] = \sigma_n^2\mathbf{I}$. Note that in general the transmitter and receiver make use of singular vectors from different SVDs.

Were perfect CSI available at the transmitter and receiver, i.e., $\Delta\mathbf{H}_1 = \Delta\mathbf{H}_2 = 0$, the MIMO channel would be decomposed into F independent single-input single-output (SISO) subchannels by this signaling scheme. With imperfect CSI, however, the resulting subchannels are coupled and subchannel interference results. In what follows, we derive SINR expressions for the SVD subchannels that account for this interchannel interference. We will assume that in equation (5.6), the signal $\mathbf{x}$, the noise $\mathbf{n}$, and the channel perturbations $\Delta\mathbf{H}_1$ and $\Delta\mathbf{H}_2$ are mutually uncorrelated[1]. The channel perturbation terms $\Delta\mathbf{H}_1$ and $\Delta\mathbf{H}_2$ are assumed to be zero-mean complex proper random variables; *i.e.*, $E[\mathrm{vec}(\Delta\mathbf{H}_k)] = 0$, $E[\mathrm{vec}(\Delta\mathbf{H}_k)\mathrm{vec}(\Delta\mathbf{H}_k)^H] = \mathbf{C}_{\Delta\mathbf{H}_k}$, $E[\mathrm{vec}(\Delta\mathbf{H}_k)\mathrm{vec}(\Delta\mathbf{H}_k)^T] = \mathbf{0}$, where the vectorization operator $\mathrm{vec}(\mathbf{A})$ stacks the columns of the matrix $\mathbf{A}$. While the derivations that follow may be generalized in a straightforward manner to the case where $\Delta\mathbf{H}_1$ and $\Delta\mathbf{H}_2$ are correlated, for simplicity we will not make this assumption.

5.2.1 General Case: MCSI

As mentioned above, an analysis for the general MCSI case has not appeared in the literature; only the CCSI case has been addressed. The results of our analysis can be specialized to the CCSI case by simply setting $\Delta\mathbf{H}_1 = 0$ (or equivalently, by setting $\mathbf{C}_{\Delta\mathbf{H}_1} = 0$). Using the relationships in (5.1)-(5.5), equation (5.6) may be written as follows:

$$\mathbf{y} = \mathbf{U}_1^H\left[(\mathbf{H} + \Delta\mathbf{H}_1 + \Delta\mathbf{H}_2)\mathbf{V}\mathbf{P}\mathbf{x} + \mathbf{n}\right] \tag{5.7}$$

$$= \left[\mathbf{\Sigma} + \Delta\mathbf{\Sigma}_1 + \Sigma\Delta\mathbf{V}_1^H\mathbf{V} + \mathbf{U}^H\Delta\mathbf{H}_2\mathbf{V} + \Delta\mathbf{\Sigma}_1\Delta\mathbf{V}_1^H\mathbf{V} + \Delta\mathbf{U}_1^H\Delta\mathbf{H}_2\mathbf{V}\right]\mathbf{P}\mathbf{x} + \mathbf{U}_1^H\mathbf{n}. \tag{5.8}$$

For relatively small perturbations, the first-order analysis of [77] can be used to approximate this expression. Using the results of [77] and dropping all second-order and higher terms, we obtain

$$\mathbf{y} \approx \left[\mathbf{\Sigma} + \Delta\mathbf{\Sigma}_1 + \mathbf{\Sigma}\mathbf{Q}_1^H + \mathbf{U}^H\Delta\mathbf{H}_2\mathbf{V}\right]\mathbf{P}\mathbf{x} + \mathbf{U}_1^H\mathbf{n} \tag{5.9}$$

[1]Although noise is partially responsible for the CSI errors, we assume that the current realization $\mathbf{n}$ of the temporally white noise was not used in calculating the receiver CSI.

with

$$\mathbf{Q}_1 = \mathbf{D} \odot (\mathbf{\Sigma}\mathbf{U}^H \Delta\mathbf{H}_1 \mathbf{V} + \mathbf{V}^H \Delta\mathbf{H}_1^H \mathbf{U}\mathbf{\Sigma}), \tag{5.10}$$

$$[\mathbf{D}]_{fg} = \begin{cases} (\sigma_g^2 - \sigma_f^2)^{-1}, & f \neq g \\ 0, & f = g \end{cases}, \tag{5.11}$$

$$\Delta\mathbf{\Sigma}_1 = \mathrm{diag}(\Delta\sigma_1, \Delta\sigma_2, \ldots, \Delta\sigma_F), \tag{5.12}$$

and

$$\Delta\sigma_f = \frac{1}{2}(\mathbf{u}_f^H \Delta\mathbf{H}_1 \mathbf{v}_f + \mathbf{v}_f^H \Delta\mathbf{H}_1^H \mathbf{u}_f), \tag{5.13}$$

where $\odot$ is the element-wise, or Hadamard, matrix product, $\Delta\mathbf{u}_{1,i}$ and $\mathbf{u}_i$ are the i-th columns of $\Delta\mathbf{U}_1$ and $\mathbf{U}$, respectively, and $\Delta\mathbf{v}_{1,i}$ and $\mathbf{v}_i$ are the i-th columns of $\Delta\mathbf{V}_1$ and $\mathbf{V}$, respectively.

In (5.9), the diagonal term $\mathbf{\Sigma} + \Delta\mathbf{\Sigma}_1$, along with the diagonal of $\mathbf{U}^H \Delta\mathbf{H}_2 \mathbf{V}$, represent the contribution of the desired signals to the received data, while the term $\mathbf{\Sigma}\mathbf{Q}_1^H$, which has a diagonal of all zeros, plus the off-diagonal elements of $\mathbf{U}^H \Delta\mathbf{H}_2 \mathbf{V}$ correspond to multiplexing interference. Thus, the signal portion of $\mathbf{y}$ at the i-th receive antenna, $y_{s,i}$, may be written as

$$y_{s,i} = (\sigma_i + \Delta\sigma_i + \mathbf{u}_i^H \Delta\mathbf{H}_2 \mathbf{v}_i)[\mathbf{P}]_{ii} x_i \ , \tag{5.14}$$

and the interference-plus-noise term is given by

$$y_{IN,i} = \sigma_i [\mathbf{Q}_1^H \mathbf{P}\mathbf{x}]_i + \sum_{\substack{f=1 \\ f \neq i}}^{F} \mathbf{u}_i^H \Delta\mathbf{H}_2 \mathbf{v}_f [\mathbf{P}]_{ff} x_f + \mathbf{u}_{1,i}^H \mathbf{n}, \tag{5.15}$$

where $[\mathbf{z}]_i$ is the i-th element of the vector $\mathbf{z}$ and $[\mathbf{A}]_{ij}$ is the ij-th element of the matrix $\mathbf{A}$.

Using the statistical assumptions made above, we may write the average signal power of the i-th data stream as

$$\begin{aligned} P_{s,i} &= E[y_{s,i}y_{s,i}^*] \\ &= E[|(\sigma_i + \Delta\sigma_i + \mathbf{u}_i^H \Delta\mathbf{H}_2 \mathbf{v}_i)[\mathbf{P}]_{ii} x_i|^2] && (5.16) \\ &= \left(\sigma_i^2 + E[|\Delta\sigma_i + \mathbf{u}_i^H \Delta\mathbf{H}_2 \mathbf{v}_i|^2]\right) E[x_i x_i^*][\mathbf{P}]_{ii}^2 && (5.17) \\ &= \left(\sigma_i^2 + E[|\Delta\sigma_i|^2] + E[|\mathbf{u}_i^H \Delta\mathbf{H}_2 \mathbf{v}_i|^2]\right)[\mathbf{P}]_{ii}^2, && (5.18) \end{aligned}$$

where $(\cdot)^*$ denotes conjugation and the equality in (5.18) is due to the assumed independence of $\Delta\mathbf{H}_1$ and $\Delta\mathbf{H}_2$. The third term in (5.18) can be further manipulated using the well-known result

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B}), \tag{5.19}$$

where $\otimes$ is the matrix Kronecker product, to obtain

$$E[|\mathbf{u}_i^H \Delta\mathbf{H}_2 \mathbf{v}_i|^2] = (\mathbf{v}_i^T \otimes \mathbf{u}_i^H)\mathbf{C}_{\Delta\mathbf{H}_2}(\mathbf{v}_i^* \otimes \mathbf{u}_i). \tag{5.20}$$

Similarly, the second term can be written as

$$\begin{aligned} E[|\Delta\sigma_i|^2] &= \frac{1}{4} E[|\mathbf{u}_i^H \Delta\mathbf{H}_1 \mathbf{v}_i + \mathbf{v}_i^H \Delta\mathbf{H}_1^H \mathbf{u}_i|^2] && (5.21) \\ &= \frac{1}{4} E[2|\mathbf{u}_i^H \Delta\mathbf{H}_1 \mathbf{v}_i|^2] && (5.22) \\ &= \frac{1}{2}(\mathbf{v}_i^T \otimes \mathbf{u}_i^H)\mathbf{C}_{\Delta\mathbf{H}_1}(\mathbf{v}_i^* \otimes \mathbf{u}_i), && (5.23) \end{aligned}$$

where (5.22) results from the complex proper assumption on $\Delta\mathbf{H}_1$ and (5.23) follows directly from the derivation of (5.20). Thus, the average signal power of the i-th data stream may be expressed as

$$P_{s,i} = \left(\sigma_i^2 + (\mathbf{v}_i^* \otimes \mathbf{u}_i)^H \left(\mathbf{C}_{\Delta\mathbf{H}_2} + \frac{1}{2}\mathbf{C}_{\Delta\mathbf{H}_1}\right)(\mathbf{v}_i^* \otimes \mathbf{u}_i)\right)[\mathbf{P}]_{ii}^2. \tag{5.24}$$

If the channel perturbations are assumed to be spatially white, *i.e.,* $\mathbf{C}_{\Delta\mathbf{H}_1} = \sigma_{H,1}^2\mathbf{I}$ and $\mathbf{C}_{\Delta\mathbf{H}_2} = \sigma_{H,2}^2\mathbf{I}$, equation (5.24) simplifies to

$$P_{s,i} = \left(\sigma_i^2 + \frac{\sigma_{H,1}^2}{2} + \sigma_{H,2}^2\right)[\mathbf{P}]_{ii}^2. \tag{5.25}$$

The interference and noise variance for the i-th subchannel is derived in a like manner, albeit with slightly more complexity due to the multiple co-channel sources of interference. The average interference-plus-noise variance for the i-th subchannel is given by

$$\begin{aligned} P_{IN,i} &= E[y_{IN,i}y_{IN,i}^*] \\ &= \sum_{\substack{f=1\\f\neq i}}^{F}\sum_{\substack{k=1\\k\neq i}}^{F}[\mathbf{P}]_{ff}[\mathbf{P}]_{kk}E\left[|\mathbf{u}_i^H\Delta\mathbf{H}_2\mathbf{v}_f|^2\right]E\left[x_fx_k^*\right] \\ &\quad + \sigma_i^2E[|\mathbf{e}_i^H\mathbf{Q}_1^H\mathbf{P}\mathbf{x}|^2] + E[\mathbf{u}_{1,i}^H\mathbf{n}\mathbf{n}^H\mathbf{u}_{1,i}] \qquad (5.26) \\ &= \sum_{\substack{f=1\\f\neq i}}^{F}[\mathbf{P}]_{ff}^2(\mathbf{v}_f^T\otimes\mathbf{u}_i^H)\mathbf{C}_{\Delta\mathbf{H}_2}(\mathbf{v}_f^*\otimes\mathbf{u}_i) + \sigma_i^2E[|\mathbf{e}_i^H\mathbf{Q}_1^H\mathbf{P}\mathbf{x}|^2] + \sigma_n^2, \qquad (5.27) \end{aligned}$$

where $\mathbf{e}_i$ is a length-N_r vector with one in the i-th position and zeros elsewhere. Using the fact that $\mathbf{Q}_1$ is skew symmetric, the complex proper assumption on $\Delta\mathbf{H}_1$, and properties of the trace operator, we find after some manipulation that the second term of (5.27) becomes

$$\begin{aligned} \sigma_i^2E[|\mathbf{e}_i^H\mathbf{Q}_1^H\mathbf{P}\mathbf{x}|^2] &= \sigma_i^4\mathrm{tr}\left[\mathbf{D}_i^2\mathbf{P}^2(\mathbf{V}^H\otimes\mathbf{u}_i^T)\mathbf{C}_{\Delta\mathbf{H}_1}^T(\mathbf{V}\otimes\mathbf{u}^*)\right] \\ &\quad + \sigma_i^2\mathrm{tr}\left[\mathbf{D}_i^2\mathbf{\Sigma}^2\mathbf{P}^2(\mathbf{v}_i^T\otimes\mathbf{U}^H)\mathbf{C}_{\Delta\mathbf{H}_1}(\mathbf{v}_i^*\otimes\mathbf{U})\right], \end{aligned} \tag{5.28}$$

where

$$[\mathbf{D}_i]_{fg} = \begin{cases} (\sigma_i^2-\sigma_f^2)^{-1}, & f=g \text{ and } f\neq i \\ 0, & f\neq g \text{ or } f=i \end{cases}. \tag{5.29}$$

As in the signal power case, if we assume that $\mathbf{C}_{\Delta\mathbf{H}_1} = \sigma_{H,1}^2\mathbf{I}$, $\mathbf{C}_{\Delta\mathbf{H}_2} = \sigma_{H,2}^2\mathbf{I}$, then (5.27) reduces to

$$P_{IN,i} = \sigma_{H,1}^2\sum_{\substack{f=1\\f\neq i}}^{F}\frac{\sigma_i^2(\sigma_f^2+\sigma_i^2)[\mathbf{P}]_{ff}^2}{(\sigma_f^2-\sigma_i^2)^2} + \sigma_{H,2}^2\sum_{\substack{f=1\\f\neq i}}^{F}[\mathbf{P}]_{ff}^2 + \sigma_n^2. \tag{5.30}$$

Finally, if the average total transmit power is constrained to be equal to P_T, i.e.,

$$\sum_{i=1}^{F} [\mathbf{P}]_{ii}^2 = P_T, \tag{5.31}$$

then we may further simplify (5.30) to

$$P_{IN,i} = \sigma_{H,1}^2 \sum_{\substack{f=1 \\ f \neq i}}^{F} \frac{\sigma_i^2(\sigma_f^2 + \sigma_i^2)[\mathbf{P}]_{ff}^2}{(\sigma_f^2 - \sigma_i^2)^2} + \sigma_{H,2}^2(P_T - [\mathbf{P}]_{ii}^2) + \sigma_n^2. \tag{5.32}$$

The SINR for the MCSI case is given by the ratio of (5.25) and (5.32).

5.2.2 Special Case: CCSI

When $\Delta\mathbf{H}_1 = \mathbf{0}$, the transmitter and receiver possess the same, albeit incorrect, CSI. In this case, referred to as the CCSI scenario, the average or expected signal power at the i-th receive antenna is given by

$$P_{s,i} = (\sigma_i^2 + \sigma_{H,2}^2)[\mathbf{P}]_{ii}^2, \tag{5.33}$$

and the interference-plus-noise variance in the i-th subchannel is given by

$$P_{IN,i} = \sigma_{H,2}^2(P_T - [\mathbf{P}]_{ii}^2) + \sigma_n^2. \tag{5.34}$$

Using (5.33) and (5.34), we find the per-subchannel SINR for the CCSI signaling model to be

$$\mathrm{SINR}_{\mathrm{CCSI},i} = \frac{(\sigma_i^2 + \sigma_{H,2}^2)[\mathbf{P}]_{ii}^2}{\sigma_{H,2}^2(P_T - [\mathbf{P}]_{ii}^2) + \sigma_n^2}. \tag{5.35}$$

Two things of significance should be noted about this special case. First, all first-order approximations used in deriving the signal and interference-plus-noise expressions of (5.25) and (5.32) are due to the presence of $\Delta\mathbf{H}_1$. Since (5.33) and (5.34) are computed for the case where $\Delta\mathbf{H}_1 = 0$, these expressions are exact. Second, note that the SINR expression for the i-th subchannal given in (5.35) is decoupled from the other subchannels. This property will

be used later in Section 5.3, allowing us to optimize the subchannel power allocation levels when using M-QAM.

5.2.3 Effect of Channel Estimation Errors

The channel perturbation variance $\sigma^2_{H,2}$ (or, correspondingly, $\mathbf{C}_{\Delta\mathbf{H}_2}$) is the result of two primary sources of error: imprecise channel estimation due to noise, and temporal variations in the channel since the receiver last updated its channel estimate. In many situations, it is possible to quantify the channel estimation errors, allowing $\sigma^2_{H,2}$ to be more precisely described in terms of system parameters. In particular, if we make the reasonable assumption that the errors due to channel estimation and time variation are uncorrelated (*i.e.*, the channel is relatively stationary over the short duration of the training symbols), then we can decompose $\sigma^2_{H,2}$ into two terms:

$$\sigma^2_{H,2} = \sigma^2_{H,o} + \sigma^2_{H,n}, \tag{5.36}$$

where $\sigma^2_{H,n}$ is the component due to estimation error and $\sigma^2_{H,o}$ accounts for all other error sources. The size of $\sigma^2_{H,n}$ depends on a number of factors including, for example, the number of pilot symbols $N_P \geq N_t$ used to estimate the channel and the power P'_T allocated to the pilots. For example, when unitary training data is used to estimate the channel, it is easily shown that if the elements of $\Delta\mathbf{H}_{2,n}$ are i.i.d, then $\sigma^2_{H,n} = K\sigma^2_n$ where $K = N_t/(P'_T N_P)$ [100].

5.2.4 Effect of Channel Mobility

There are a number of ways to relate the component of the channel perturbation due to mobility to the given physical scenario in question. As a simple example, consider a Gauss-Markov fading channel model [101], in which the current channel at time t, $\mathbf{H}_t$, is related to a reference channel $\mathbf{H}_{t-r}$ in the past at time $t-r$ according to the following first-order auto-regressive (AR) model:

$$\mathbf{H}_t = \sqrt{\alpha_r}\mathbf{H}_{t-r} + \sqrt{1-\alpha_r}\mathbf{E}_t \ , \tag{5.37}$$

where $\mathbf{E}_t$ is independent of $\mathbf{H}_{t-r}$, $\text{vec}(\mathbf{H}_{t-r}) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{C}_H)$, $\text{vec}(\mathbf{E}_t) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{C}_H)$, $\mathbf{C}_H$ is the covariance of the channel matrix and $0 \leq \alpha_r \leq 1$. Note that under this model, $\mathbf{H}_t$ maintains the same distribution as $\mathbf{H}_{t-r}$. If this model is employed in the signaling scheme of (5.6), the average received signal power remains constant as the channel changes, unlike the additive perturbation model $\mathbf{H} + \Delta\mathbf{H}$. The parameter α_r is tuned according to the level of mobility assumed in the scenario under consideration. For example, it can be shown [101] that the choice

$$\alpha_r = J_0(2\pi r f)^2 \tag{5.38}$$

yields second-order channel statistics that match those of Jakes' fading model [102], where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind, $f = f_d T_s$, f_d is the maximum Doppler frequency in the fading environment, and T_s is the sampling period.

The preceding SINR analysis can be easily adapted to the above mobility model. For example, the subchannel power and interference expressions for the CCSI case under (5.37) are given by

$$P_{s,i} = \left(\alpha_r \sigma_i^2 + 1 - \alpha_r + \sigma_{H,\mathbf{n}}^2\right) [\mathbf{P}]_{ii}^2 \tag{5.39}$$

and

$$P_{IN,i} = \left(1 - \alpha_r + \sigma_{H,\mathbf{n}}^2\right) (P_T - [\mathbf{P}]_{ii}^2) + \sigma_n^2, \tag{5.40}$$

assuming the noise model in (5.36) with array motion and measurement noise as the primary contributors to the channel perturbation.

5.2.5 SINR for Specified Outage Rates

The SINR expressions developed thus far are average SINR expressions; that is, they are defined as the ratio of the average signal power and the average interference-plus-noise variance. These average expressions are simple and may often be easily applied to problems involving the SINR. In some applications, however, such average expressions may not be sufficient. For example, consider the problem of transmitting using subchannel power bit-loading levels chosen based on the expressions of (5.33) and (5.34). An outage is considered to occur anytime the instantaneous SER exceeds the target value $\mathcal{E}$. Bit-loading design based upon these average expressions may lead to unacceptably high outage rates as the

level of channel perturbation grows. A common approach to solving this problem is to use a back-off factor, decreasing the target SER (or equivalently, increasing the target SINR) to achieve a more conservative, lower-throughput operating point. While this approach is simple, it is not obvious how to choose the back-off value in order to achieve an acceptable outage probability.

An alternative approach would be to derive a more conservative SINR expression that lies below the instantaneous SINR for some specified probability near one and choose bit-loading levels based upon this more conservative analysis. Such an SINR confidence level could be directly found were the probability distribution function (pdf) of the SINR known. Unfortunately, the exact SINR distribution for the problem we consider is not easily derived, and confidence interval thresholds based on it would likely have to be determined numerically. However, using some simplifying assumptions, it is possible to derive approximate SINR confidence levels.

To facilitate the derivation of the SINR confidence levels, we assume that the channel perturbation and measurement noise distributions are zero-mean, spatially uncorrelated complex Gaussian random variables, i.e., $\text{vec}(\Delta\mathbf{H}_1) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \sigma_H^2\mathbf{I})$, $\text{vec}(\Delta\mathbf{H}_2) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \sigma_H^2\mathbf{I})$, and $\mathbf{n} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I})$. In particular, note that the Gaussian assumption was not necessary in the previous SINR derivation, but is necessary for what follows. Now consider again the signaling scheme of (5.6), which we may approximate as

$$\mathbf{y} = (\mathbf{U} + \Delta\mathbf{U}_1)^H \left[(\mathbf{H} + \Delta\mathbf{H}_1 + \Delta\mathbf{H}_2)\mathbf{V}\mathbf{P}\mathbf{x} + \mathbf{n}\right] \tag{5.41}$$

$$\approx (\Sigma + \Delta\mathbf{U}_1^H\mathbf{H}\mathbf{V} + \mathbf{U}^H\Delta\mathbf{H}_1\mathbf{V} + \mathbf{U}^H\Delta\mathbf{H}_2\mathbf{V})\mathbf{P}\mathbf{x} + (\mathbf{U} + \Delta\mathbf{U}_1)^H\mathbf{n}, \tag{5.42}$$

where (5.42), like (5.9), is a first-order approximation relative to the channel perturbation terms. However, the approximation made here is different from that made in (5.9) and, coupled with the Gaussian assumptions, allows for a simpler derivation. From (5.42), we find that the squared norm of the received signal at the i-th receive antenna may be written as

$$|y_{s,i}|^2 = |\sigma_i + \Delta\mathbf{u}_i^H\mathbf{H}\mathbf{v}_i + \mathbf{u}_i^H\Delta\mathbf{H}_1\mathbf{v}_i + \mathbf{u}_i^H\Delta\mathbf{H}_2\mathbf{v}_i|^2[\mathbf{P}]_{ii}^2|x_i|^2. \tag{5.43}$$

Using the first-order analysis of [77], we find that

$$\Delta \mathbf{u}_i^H \mathbf{H} \mathbf{v}_j = \begin{cases} \frac{\sigma_i \sigma_j}{\sigma_i^2 - \sigma_j^2} \mathbf{v}_i^H \Delta \mathbf{H}_1^H \mathbf{u}_j + \frac{\sigma_j^2}{\sigma_i^2 - \sigma_j^2} \mathbf{u}_i^H \Delta \mathbf{H}_1 \mathbf{v}_j, & i \neq j \\ 0, & i = j \end{cases} . \tag{5.44}$$

This simplifies (5.43) to

$$|y_{s,i}|^2 = |\sigma_i + \mathbf{u}_i^H \Delta \mathbf{H}_1 \mathbf{v}_i + \mathbf{u}_i^H \Delta \mathbf{H}_2 \mathbf{v}_i|^2 [\mathbf{P}]_{ii}^2 |x_i|^2. \tag{5.45}$$

In Section 5.2.1, the average signal power was found by taking the expected value of $|y_{s,i}|^2$ over the channel perturbations $\Delta \mathbf{H}_1$ and $\Delta \mathbf{H}_2$ and the signal x_i. This time, we take the expectation over x_i only, obtaining

$$E_{\mathbf{x}}[|y_{s,i}|^2] = |\sigma_i + \mathbf{u}_i^H \Delta \mathbf{H}_1 \mathbf{v}_i + \mathbf{u}_i^H \Delta \mathbf{H}_2 \mathbf{v}_i|^2 [\mathbf{P}]_{ii}^2. \tag{5.46}$$

Note that $\mathbf{u}_i^H \Delta \mathbf{H}_1 \mathbf{v}_i + \mathbf{u}_i^H \Delta \mathbf{H}_2 \mathbf{v}_i$ is a zero-mean Gaussian random variable whose real and imaginary parts have variances $(\sigma_{H,1}^2 + \sigma_{H,2}^2)/2$ independent of i. As a result, (5.46) may be expressed as a function of a single random variable as

$$E_{\mathbf{x}}[|y_{s,i}|^2] = \frac{\sigma_{H,1}^2 + \sigma_{H,2}^2}{2} [\mathbf{P}]_{ii}^2 Z_s, \tag{5.47}$$

where the random variable $Z_{s,i} \sim \chi_{nc}^2(k, \lambda)$ is noncentral chi-square with $k = 2$ degrees of freedom and noncentrality parameter $\lambda = 2\sigma_i^2/(\sigma_{H,1}^2 + \sigma_{H,2}^2)$ [103].

Following a similar procedure and utilizing (5.44), the interference and noise term averaged over $\mathbf{x}$ is given by

$$\begin{aligned} E_{\mathbf{x}}[|y_{IN,i}|^2] = & \sum_{\substack{f=1 \\ f \neq i}}^{F} \left| \frac{\sigma_i \sigma_f}{\sigma_i^2 - \sigma_f^2} \mathbf{v}_i^H \Delta \mathbf{H}_1^H \mathbf{u}_f + \frac{\sigma_i^2}{\sigma_i^2 - \sigma_f^2} \mathbf{u}_i^H \Delta \mathbf{H}_1 \mathbf{v}_f + \mathbf{u}_i^H \Delta \mathbf{H}_2 \mathbf{v}_f \right|^2 [\mathbf{P}]_{ff}^2 \\ & + |\mathbf{U}_1^H \mathbf{n}|^2. \end{aligned} \tag{5.48}$$

It is straightforward to show that the random variables $\mathbf{v}_i^H \Delta \mathbf{H}_1^H \mathbf{u}_f$, $\mathbf{u}_i^H \Delta \mathbf{H}_1 \mathbf{v}_f$, and $\mathbf{u}_i^H \Delta \mathbf{H}_2 \mathbf{v}_f$ are mutually independent zero-mean Gaussian random variables with variances $\sigma_{H,1}^2$, $\sigma_{H,1}^2$,

and $\sigma_{H,2}^2$, respectively, irrespective of i and f. Thus, each term in the sum of (5.48) is itself a zero-mean complex Gaussian random variable. It is also easy to show that the terms in the sum of (5.48) are independent one from another. Were the measurement noise and each term in the sum to have the same variance, then (5.48) would be a scaled realization of a chi-square random variable. Unfortunately, this is not the case, and the distribution of (5.48) is not known. However, we may obtain an approximate result by assuming all of the terms in (5.48) share an average variance for the purposes of selecting a chi-square distribution, and then substituting the resulting random variable into the equation. Following this procedure, we have

$$E_{\mathbf{x}}[|y_{IN,i}|^2] \approx \frac{Z_{IN}}{2(M-1)} \left[\sum_{\substack{f=1 \\ f \neq i}}^{M} \left(\frac{\sigma_i^2(\sigma_f^2 + \sigma_i^2)}{(\sigma_f^2 - \sigma_i^2)^2} \sigma_{H,1}^2 + \sigma_{H,2}^2 \right) [\mathbf{P}]_{ff}^2 + \sigma_n^2 \right], \tag{5.49}$$

where $Z_{IN} \sim \chi^2(2(M-1))$ and M is the number of active subchannels, or, equivalently, the number of nonzero diagonal entries in $\mathbf{P}$. Note that the result in (5.49) implicitly assumes that $M > 1$. If $M = 1$, there is no interchannel interference and $E_{\mathbf{x}}[|y_{IN,i}|^2]$ is easily found to be $Z'_{IN}\sigma_n^2/2$ with $Z'_{IN} \sim \chi^2(2)$.

Using (5.47) and (5.49), we may now write an approximate SINR confidence level expression. Let $X_{nc}^{-1}(p,k,\lambda)$ and $X^{-1}(p,k)$ be the inverse cumulative distribution functions for the noncentral and central chi-square distributions, respectively, i.e., $z = X_{nc}^{-1}(p,k,\lambda) = \{z : \Pr(Z \leq z|k,\lambda) = p\}$ and $z = X^{-1}(p,k) = \{z : \Pr(Z \leq z|k,\lambda) = p\}$ where Z is a noncentral or central chi-square random variable, respectively. Then we may select signal power and interference-plus-noise confidence levels as

$$P_{s,i} = \frac{\sigma_{H,1}^2 + \sigma_{H,2}^2}{2} [\mathbf{P}]_{ii}^2 X_{nc}^{-1} \left(p_s, 2, \frac{2\sigma_i^2}{\sigma_{H,1}^2 + \sigma_{H,2}^2} \right) \tag{5.50}$$

and

$$P_{IN,i} = \frac{X^{-1}\left(p_{IN}, 2(M-1)\right)}{2(M-1)} \left[\sum_{\substack{f=1 \\ f \neq i}}^{M} \left(\frac{\sigma_i^2(\sigma_f^2 + \sigma_i^2)}{(\sigma_f^2 - \sigma_i^2)^2} \sigma_{H,1}^2 + \sigma_{H,2}^2 \right) [\mathbf{P}]_{ff}^2 + \sigma_n^2 \right]. \tag{5.51}$$

Now define $\text{SINR}_{p,i}$ as the subchannel SINR confidence level such that $\Pr(\text{SINR}_i \leq \text{SINR}_{p,i}) = p$. In order to obtain $\text{SINR}_{p,i}$, we must combine the confidence levels of (5.50) and (5.51). Motivated by the observation that the random variables (5.47) and (5.49) are independent, we will combine the confidence levels for the signal and interference-plus-noise powers by simply taking their ratios. That is,

$$\text{SINR}_{p,i} \approx \begin{cases} \dfrac{\frac{\sigma_{H,1}^2+\sigma_{H,2}^2}{2}[\mathbf{P}]_{ii}^2 X_{nc}^{-1}\left(p,2,\frac{2\sigma_i^2}{\sigma_{H,1}^2+\sigma_{H,2}^2}\right)}{\frac{X^{-1}(1-p,2(M-1))}{2(M-1)}\left[\sum_{\substack{f=1\\f\neq i}}^{M}\left(\frac{\sigma_i^2(\sigma_f^2+\sigma_i^2)}{(\sigma_f^2-\sigma_i^2)^2}\sigma_{H,1}^2+\sigma_{H,2}^2\right)[\mathbf{P}]_{ff}^2+\sigma_n^2\right]} & M \geq 2 \\ \dfrac{\frac{\sigma_{H,1}^2+\sigma_{H,2}^2}{2}[\mathbf{P}]_{ii}^2 X_{nc}^{-1}\left(p,2,\frac{2\sigma_i^2}{\sigma_{H,1}^2+\sigma_{H,2}^2}\right)}{\frac{\sigma_n^2}{2}X^{-1}(1-p,2)} & M = 1 \end{cases}, \tag{5.52}$$

where values for $X_{nc}^{-1}(p,k,\lambda)$ and $X^{-1}(p,k)$ may be found in distribution tables or calculated using statistical software. While we could combine the confidence levels in a more sophisticated manner, the additional effort involved is not justified considering the level of approximation already present in the confidence levels. We expect the above SINR confidence level expression to be most accurate when the interference term closely resembles a scaled chi-square random variable. This is most likely to occur in the CCSI case with uniform subchannel power levels when $\sigma_n^2 \approx \sigma_{H,2}^2$.

Note that the inverse noncentral chi-square function depends on the subchannel singular values and the channel perturbation variance. As a result, its precomputation may pose an additional design challenge. For a moderate number of transmit and receive antennas, we have found that we may further approximate the above SINR confidence level expression to remove this function from the numerator for $M \geq 2$. The resulting approximation is given by

$$\text{SINR}_{p,i} \approx \begin{cases} \dfrac{\sigma_i^2[\mathbf{P}]_{ii}^2}{\frac{X^{-1}(1-0.5p,2(M-1))}{2(M-1)}\left[\sum_{\substack{f=1\\f\neq i}}^{M}\left(\frac{\sigma_i^2(\sigma_f^2+\sigma_i^2)}{(\sigma_f^2-\sigma_i^2)^2}\sigma_{H,1}^2+\sigma_{H,2}^2\right)[\mathbf{P}]_{ff}^2+\sigma_n^2\right]} & M \geq 2 \\ \dfrac{\frac{\sigma_{H,1}^2+\sigma_{H,2}^2}{2}[\mathbf{P}]_{ii}^2 X_{nc}^{-1}\left(p,2,\frac{2\sigma_i^2}{\sigma_{H,1}^2+\sigma_{H,2}^2}\right)}{\sigma_n^2} & M = 1 \end{cases}, \tag{5.53}$$

where we have replaced the numerator with a nominal value, which we chose in this case to be the received signal power in the absence of channel perturbations, and the denominator probability level was then modified to compensate for the change in the numerator. An example of the performance of the approximate confidence interval will be given below.

5.2.6 Numerical Examination of the SINR

In this section, we provide Monte Carlo simulations to detail some of the characteristics of the derived SINR expressions. We assume a 4×4 MIMO channel $\mathbf{H}$ selected such that $\text{vec}(\mathbf{H}) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{I}_{16})$, a measurement noise variance of -30 dB relative to unity, and a total transmit power of unity with power allocated uniformly over all of the SVD subchannels. The channel perturbations for the Monte Carlo simulations are generated with $\text{vec}(\Delta\mathbf{H}_1) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \sigma_{H,1}^2\mathbf{I})$ and $\text{vec}(\Delta\mathbf{H}_2) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \sigma_{H,2}^2\mathbf{I})$ for various values of $\sigma_{H,1}^2$ and $\sigma_{H,2}^2$. Figure 5.1 compares the SINRs generated using the MCSI SINR expression with Monte Carlo results for the first and fourth subchannels and for two special cases, $\sigma_{H,1}^2 = 0$ and $\sigma_{H,2}^2 = 0$. The plot indicates excellent agreement between the simulations and the first-order analysis, even for values of $\sigma_{H,1}^2 \to 1$, where the perturbations are of the same "size" as the channel coefficients themselves (excellent agreement for $\sigma_{H,1}^2 = 0$ and $\sigma_{H,2}^2 \to 1$ is guaranteed since the analysis is exact in this case). SINR results when both $\sigma_{H,1}^2$ and $\sigma_{H,2}^2$ are nonzero may be inferred from the special cases given above.

In some cases, the receiver may be aware of the value of $\mathbf{H}$ known at the transmitter. This situation will arise, for example, if the error term $\Delta\mathbf{H}_1$ is due to quantization effects, which are known at the receiver prior to feedback. In such situations, the receiver would have the option of creating a CCSI scenario by ignoring the $\Delta\mathbf{H}_1$ term, and using $\mathbf{U}$ instead of $\mathbf{U}_1$ in (5.6). In effect, this amounts to assigning $\Delta\mathbf{H}_1$ (or the portion of it due to quantization errors) to $\Delta\mathbf{H}_2$. Using the SINR analysis approach above, one can show that it is preferable for the transmitter and receiver to use common CSI rather than mismatched CSI on the i-th subchannel if

$$\sum_{\substack{f=1 \\ f \neq i}}^{F} \frac{\sigma_i^2(\sigma_f^2 + \sigma_i^2)[\mathbf{P}]_{ff}^2}{(\sigma_f^2 - \sigma_i^2)^2} > P_T - [\mathbf{P}]_{ii}^2 \,, \tag{5.54}$$

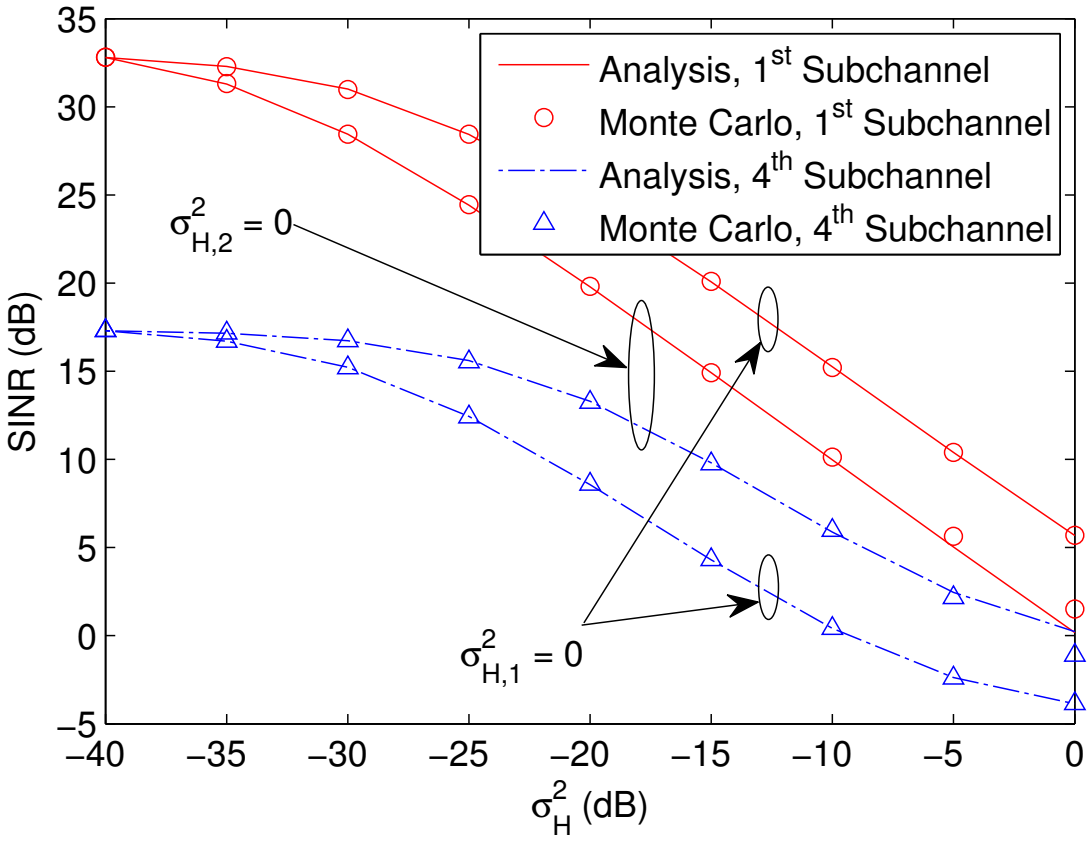


Figure 5.1: Comparison of the MCSI SINR expression results with Monte Carlo results for the first and fourth subchannels.

which, interestingly, will often be the case. The inequality of (5.54) will most likely be violated when the singular value of subchannel i is significantly smaller than the larger singular values (in which case it may not be assigned any transmit power anyway). This observation is borne out by the following simulation example. In this example, we fix the total channel perturbation variance to $\sigma_T^2 = \sigma_{H,1}^2 + \sigma_{H,2}^2 = -15$ dB and we let $\sigma_{H,1}^2 = \rho\sigma_T^2$ and $\sigma_{H,2}^2 = (1-\rho)\sigma_T^2$ for $0 \leq \rho \leq 1$. Thus, ρ is the fraction of the total channel perturbation variance due to $\sigma_{H,1}^2$; $\rho = 0$ means we assign all of $\Delta\mathbf{H}_1$ to $\Delta\mathbf{H}_2$ and create a CCSI scenario. Increasing ρ towards one increases the mismatch in CSI between the transmitter and receiver, but reduces the power of $\Delta\mathbf{H}_2$. Figure 5.2 shows the resulting SINRs for the four subchannels of a 4×4 Rayleigh distributed channel. As predicted, the simulation shows that using CCSI instead of MCSI is preferable except for the weakest subchannel over all values of ρ.

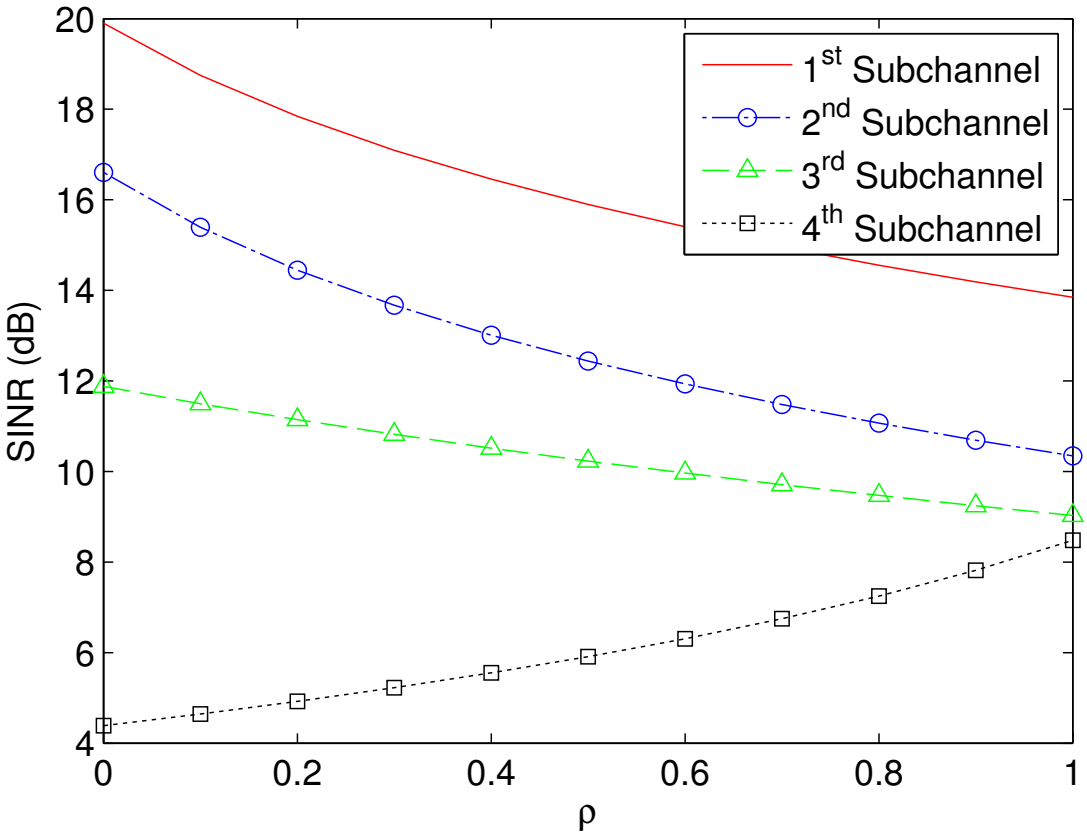


Figure 5.2: A comparison of the MCSI and CCSI cases via an examination of the per-subchannel SINRs as a function of ρ where $\sigma_{H,1}^2 = \rho\sigma_T^2$ and $\sigma_T^2 = \sigma_{H,1}^2 + \sigma_{H,2}^2 = 15$ dB.

In the last example, we compare the derived SINR expressions with those from previous CCSI SINR analyses. We again assume a 4×4 MIMO channel $\mathbf{H}$ selected such that $\text{vec}(\mathbf{H}) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{I}_{16})$. Using uniform power loading over all of the subchannels with a total transmit power of one, and assuming a estimation error variance of -20 dB, we find SINR values for the first MIMO subchannel for several different channel perturbation powers and compare these results to those obtained through Monte Carlo simulation assuming $\text{vec}(\Delta\mathbf{H}_2) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \sigma_H^2\mathbf{I})$. The results for the other subchannels are similar to those presented here. Figure 5.3 shows the resulting SINR averaged over 100 channel realizations with 10,000 independent perturbations per realization. The upper four curves shown in the figure correspond to average SINR expressions. These include our CCSI results (equation (5.35), labeled "CCSI"), the first-order SINR expression from [94] (labeled "1st-order"), the SINR

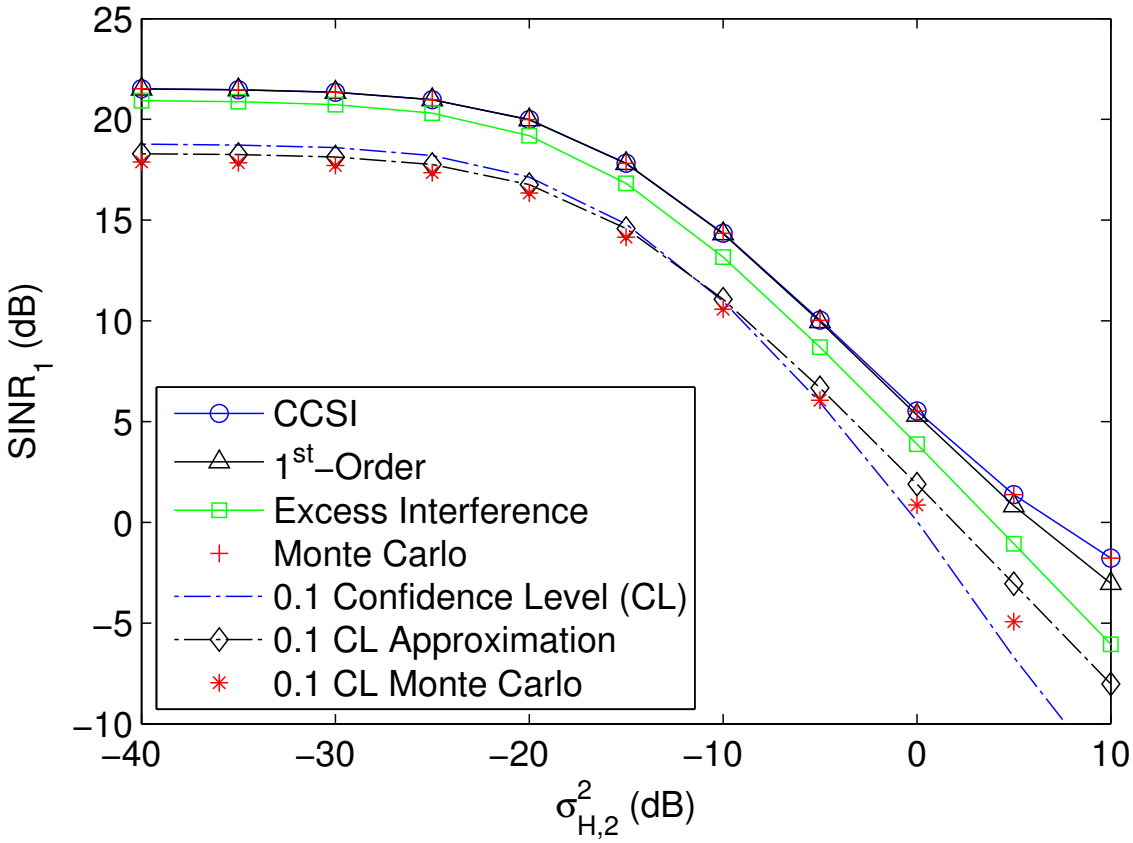


Figure 5.3: A comparison of the per-subchannel SINR expressions with Monte Carlo simulation.

expression from [64] (labeled "Excess Interference"), and Monte Carlo simulation results. Equation (5.35) matches the Monte Carlo results exactly through the entire range of channel perturbation powers, while the other SINR expressions differ from the simulations to varying degrees. In particular, the result from [64] tends to underestimate the subchannel SINR over a wide range of $\sigma^2_{H,2}$ values. The lower three curves in Fig. 5.3 are results for the SINR confidence level expressions. Both the original SINR confidence level expression of (5.52) and its further approximation of (5.53) agree well with the Monte Carlo results over a wide range of useful channel perturbation variance levels.

5.3 Subchannel Power Allocation

Assume a communications system in which the transmission over each subchannel is required to achieve as high a rate as possible while meeting a particular symbol error rate (SER) threshold, $\mathcal{E}$. The various available subchannels may be active or inactive depending on the chosen approach. When only one subchannel is in use, the scheme is referred to as beamforming; on the other hand, transmission over multiple subchannels results in spatial multiplexing. Using the SINR expressions derived above, we wish to determine whether beamforming or multiplexing is the optimal approach in terms of total throughput. If multiplexing is chosen, one must determine how to allocate power to the various subchannels in order to maximize the resulting transmission rate. We will consider this second problem first, for both the MCSI and CCSI cases.

In order to proceed, we must first relate the SER, the number of bits per transmission symbol, and the SINR. This is done for M-QAM signals using two approximations. The first relates the SER to the SINR and the minimum distance between constellation points for a constellation with unit average energy [104]:

$$\text{SER} \approx 4Q\left(\sqrt{d_{min}^2 \text{SINR}}\right), \tag{5.55}$$

where Q is the well-known Q-function for Gaussian distributions and d_{min} is the minimum constellation distance. This approximation is valid for general M-QAM constellations. Other approximations can be used for certain special cases [64, 104].

The second approximation relates the minimum distance between M-QAM constellation points to bits per M-QAM symbol by fitting the following first-order-in-the-exponent curve to the constellation points:

$$d_{min}^2 = e^{-\alpha b + \beta}, \tag{5.56}$$

where b is the number of bits per symbol. Assuming symmetric constellations, a least squares fit of α and β to (5.56) results in the following parameter values $\alpha = 0.717, \beta = 0.604$. Using (5.55) and (5.56), we may then approximate the number of bits per transmission as a function

of the SER and SINR:

$$b \approx \frac{1}{\alpha}\left(\beta - \ln\left(\frac{(Q^{-1}(\mathrm{SER}/4))^2}{\mathrm{SINR}}\right)\right), \tag{5.57}$$

where $Q^{-1}(\cdot)$ is the inverse Q-function. Note that any suboptimality in the approximately optimal power allocation developed below is a result of the approximations in (5.57).

5.3.1 MCSI Subchannel Power Allocation

For the MCSI case, we consider two ad-hoc methods which employ the signal power and interference-plus-noise expressions in (5.25) and (5.32) (or the SINR confidence level expressions of (5.52) or (5.53)) to adjust the power levels suggested by the waterfilling solution in order to obtain more robust bit-loading levels. The first, which is the simpler of the two methods, takes a bottom-up approach to the subchannels. (Note that the subchannels are assumed to be ordered from largest to smallest according to the size of their associated singular values.) We begin by examining the smallest subchannel allocated power by the waterfilling method by using (5.57) to determine the number of bits to be allocated to the subchannel. If (5.57) indicates that the subchannel cannot support any bits, the subchannel's power is reallocated up to the next subchannel. This process is repeated in an ascending manner for the remaining subchannels.

The second method is essentially the ad-hoc method of [73], though modified to use our SINR expressions and the SER and bit expressions of (5.55) and (5.57). This method employs an iterative top-down approach for allocating power. We begin the process by using the waterfilling method to allocate power among the subchannels assuming perfect CSI. Next, (5.57) is used to determine the number of bits that may be allocated to the strongest subchannel using the available power. Given this bit allocation, we then determine the minimum amount of power required to support that number of bits at the required SER target level. Since this cannot be determined directly through our expressions, we employ a numerical search to find this minimum power. The excess power is reallocated to the next strongest subchannel. The above steps are then repeated for the next subchannel and so on until all of the remaining subchannels have been considered. Once the total bit allocation for all of the subchannels has been determined and recorded, the process is carried out again,

this time removing the smallest subchannel and reallocating its power before repeating the process. This overall process is repeated until the limiting beamforming case. The total bit allocations for each case are compared and the one that results in the largest number of bits is chosen.

Note that in [73], excess power from the smallest subchannel could also be reallocated from one temporal transmission to the next. Though this modification is not difficult to implement, it is not used in the simulations of Section 5.5 as it would unnecessarily complicate comparisons with other methods.

5.3.2 CCSI Subchannel Power Allocation

For the CCSI case, our SINR expression allows for a more systematic optimization than in the MCSI case. In this section, we use the subchannel CCSI SINR expression in (5.35) to find approximately optimal subchannel power levels. Consider the following constrained optimization problem, where for a given number of active subchannels $M \leq F$, a diagonal power loading matrix $\mathbf{P}$ with positive, real-valued elements is chosen to maximize the number of M-QAM bits per transmission while guaranteeing a minimum SER of $\mathcal{E}$:

$$\mathbf{P}^* = \arg \max_{\substack{\mathbf{P}:\sum [\mathbf{P}]_{ff}^2 = P_T \\ \text{SER} \leq \mathcal{E}}} \sum_{f=1}^{M} b_f. \tag{5.58}$$

The general optimization problem in (5.58) is difficult to solve, but the approximation obtained by substituting (5.57) into (5.58) can be solved in closed form. The resulting problem may be expressed as

$$\mathbf{P}^* \approx \arg \min_{\sum [\mathbf{P}]_{ff}^2 = P_T} \sum_{f=1}^{M} \ln \left(\frac{P_{IN,f}}{P_{s,f}} \right), \tag{5.59}$$

which we solve by using Lagrange multipliers:

$$J = \sum_{f=1}^{M} \ln \left(\frac{P_{IN,f}}{P_{s,f}} \right) - \gamma \left(\sum_{f=1}^{M} [\mathbf{P}]_{ff}^2 - P_T \right). \tag{5.60}$$

Equating the derivative of J with respect to $[\mathbf{P}]_{ii}^2$ to zero, we find after some manipulation that

$$[\mathbf{P}]_{ii}^2 = \frac{\gamma B \pm \sqrt{\gamma^2 B^2 + 4\gamma B \sigma_{H,2}^2}}{2\gamma\sigma_{H,2}^2}, \tag{5.61}$$

$$\gamma = \frac{-B}{[\mathbf{P}]_{ii}^2 (B - [\mathbf{P}]_{ii}^2 \sigma_{H,2}^2)}, \tag{5.62}$$

and

$$B = P_T \sigma_{H,2}^2 + \sigma_n^2. \tag{5.63}$$

The second derivative of J with respect to $[\mathbf{P}]_{ii}^2$ is strictly greater than zero when

$$[\mathbf{P}]_{ii}^2 \leq \frac{P_T}{2} + \frac{\sigma_n^2}{2\sigma_{H,2}^2}, \tag{5.64}$$

in which case (5.61) will not correspond to a saddle point. A few things must be noted here. First, with γ as given in (5.62), the powers given by (5.61) are always real and positive as we require. Second, while it may appear otherwise, the solution for γ in (5.62) must be independent of i. Thus, (5.61) and (5.62) imply that the optimal power loading scheme when using $M \leq F$ active subchannels is to spread the power uniformly over the subchannels, i.e.,

$$[\mathbf{P}]_{ii}^{2*} = \frac{P_T}{M}. \tag{5.65}$$

This solution, which corresponds to the larger root of (5.61) ($-$ from $\pm$), always satisfies the second derivative condition of (5.64). The power level in (5.65) is for a particular $M \leq F$. The optimal M may be selected as the value of M which results in the largest value for

$$J_2 = \sum_{f=1}^{M} \ln\left(\frac{P_{s,f}}{P_{IN,f}}\right), \tag{5.66}$$

with the power allocated evenly over the M active subchannels. Since F is typically not large, the cost of this search is small.

5.4 Beamforming vs. Multiplexing Thresholds

In this section, using the power allocation methods discussed above, we derive thresholds for the CSI error and noise level above which beamforming rather than spatial multiplexing is optimal in terms of bits per transmission.

5.4.1 CCSI Thresholds

We begin with the CCSI case, for which we will assume the optimality of uniform power allocation. Multiplexing over M subchannels is preferable to beamforming in terms of bits per transmission when

$$b_{BF} \leq \sum_{m=1}^{M} b_m \ , \tag{5.67}$$

where b_{BF} is the number of bits supported when beamforming and $b_1, b_2, \ldots, b_M$ are the number of bits supported when multiplexing using the first M subchannels, where the subchannels are ordered by decreasing singular values. Using (5.57), we may re-express the relationship of (5.67) at the threshold as

$$\frac{\beta}{\alpha} - \frac{1}{\alpha} \log \left[\frac{(Q^{-1}(\mathcal{E}/4))^2 \sigma_n^2}{P_{s,BF}} \right] \leq \sum_{m=1}^{M} \left(\frac{\beta}{\alpha} - \frac{1}{\alpha} \log \left[\frac{(Q^{-1}(\mathcal{E}/4))^2 P_{IN,m}}{P_{s,m}} \right] \right), \tag{5.68}$$

where $P_{s,BF}$ is the signal power when beamforming. With the uniform power allocation assumption, $P_{IN,m}$ is the same for all m, and the above expression may be simplified to

$$(Q^{-1}(\mathcal{E}/4))^{2(M-1)} (P_{IN,m})^M P_{s,BF} \leq \left[\prod_{m=1}^{M} P_{s,m} \right] \sigma_n^2 e^{(M-1)\beta}. \tag{5.69}$$

Plugging (5.33) and (5.34) into (5.69), we find after some rearrangement that

$$\sigma_{H,2}^2 \leq \frac{\left[\left(\frac{P_T e^{\beta}}{(Q^{-1}(\mathcal{E}/4))^2} \right)^{M-1} \left(\prod_{m=2}^{M} (\sigma_m^2 + \sigma_{H,2}^2) \right) \sigma_n^2 \right]^{\frac{1}{M}} - M\sigma_n^2}{(M-1)P_T}. \tag{5.70}$$

Finally, if we assume that $\sigma_m^2 \gg \sigma_{H,2}^2$ near the threshold for the M active subchannels, we obtain the following threshold for the channel perturbation power:

$$\sigma_{H,2,M}^{2*} = \frac{\left[\left(\frac{P_T e^{\beta}}{(Q^{-1}(\mathcal{E}/4))^2}\right)^{M-1}\left(\prod_{m=2}^{M}\sigma_m^2\right)\sigma_n^2\right]^{\frac{1}{M}} - M\sigma_n^2}{(M-1)P_T}, \tag{5.71}$$

where multiplexing is optimal for $\sigma_{H,2}^2 \leq \sigma_{H,2,M}^{2*}$ and beamforming is optimal otherwise. The above threshold describes the optimal transition point between multiplexing over M subchannels and beamforming. An overall threshold may be found by evaluating (5.71) for possible values of M and selecting the maximum, i.e,

$$\sigma_{H,2}^{2*} = \max_{M} \sigma_{H,2,M}^{2*}. \tag{5.72}$$

Since $1 < M \leq F$ and F is typically not large, the cost of this search is small.

Note that the threshold of (5.72) must be positive in order for multiplexing to be useful for any value of $\sigma_{H,2}^2$. Using this observation, we obtain the following threshold on the measurement noise above which beamforming is always preferable:

$$\sigma_n^{2*} = \left(\frac{P_T e^{\beta}}{M^*(Q^{-1}(\mathcal{E}/4))^2}\right)\left(\frac{\prod_{m=2}^{M^*}\sigma_m^2}{M^*}\right)^{\frac{1}{M^*-1}}, \tag{5.73}$$

where M^* is the maximizing M from (5.72).

5.4.2 CCSI Thresholds for Specified Outage Rate

The threshold of (5.72) was derived using average signal and interference-plus-noise powers, and thus, due to outages, the effective multiplexing/beamforming threshold may be somewhat lower than that predicted by (5.72). The SINR confidence intervals derived earlier can be used to address this issue if we use (5.33) instead of (5.50) to obtain an expression for $P_{s,BF}$ in (5.69). This approximation is accurate since $\sigma_1^2 \gg \sigma_{H,2}^{2*}$ for values near the threshold. The other power terms in (5.69), $P_{s,m}$ and $P_{IN,m}$, are found using the numerator and denominator of (5.52) for $M > 1$. The derivation proceeds in the same manner as before,

and we find

$$\sigma^{2*}_{H,2,M} = \frac{\left[\left(\frac{P_T e^{\beta}}{(Q^{-1}(\mathcal{E}/4))^2}\right)^{M-1}\left(\prod_{m=2}^{M}\sigma_m^2\right)\sigma_n^2\right]^{\frac{1}{M}} - M\sigma_n^2}{X^{-1}(1-2p, 2(M-1))P_T/2}, \tag{5.74}$$

the maximum of which over M gives the overall threshold. Note that the numerator of (5.74) is identical to that of (5.71). Thus, the threshold on σ_n^2 is still given by (5.73).

5.4.3 Separate Estimation Error and the Gauss-Markov Model

We may modify the previously derived thresholds to accommodate (5.36) with $\sigma^2_{H,n} = K\sigma_n^2$ and the Gauss-Markov channel mobility model in (5.37). Assuming that α_r is sufficiently close to one, a similar assumption to that made in Section 5.4.1 with regard to the size of the channel perturbation near the threshold, and following the previous derivations, we find thresholds on $1-\alpha_r$ and σ_n^2 which are given by

$$1-\alpha_r^* = \frac{\left[\left(\frac{P_T e^{\beta}}{(Q^{-1}(\mathcal{E}/4))^2}\right)^{M^*-1}\left(\prod_{m=2}^{M^*}\sigma_m^2\right)\sigma_n^2\right]^{\frac{1}{M^*}} - M^*\sigma_n^2 - (M^*-1)P_T\sigma^2_{H,n}}{(M^*-1)P_T}, \tag{5.75}$$

and

$$\sigma_n^{2*} = \left(\frac{P_T e^{\beta}}{(Q^{-1}(\mathcal{E}/4))^2}\right)\left(\frac{\prod_{m=2}^{M^*}\sigma_m^2}{((P_T K+1)M^* - P_T K)^{M^*}}\right)^{\frac{1}{M^*-1}}. \tag{5.76}$$

Note that the influence of the assumption on α_r acts in a manner counter to the effect of the earlier assumption $\sigma_m^2 \gg \sigma^2_{H,o}$.

5.4.4 MCSI Thresholds

The techniques used to find the thresholds for the CCSI case with uniform power allocation do not hold for the MCSI case. However, we can find alternative thresholds for the MCSI case if we make the additional assumption that the transition from multiplexing to beamforming occurs as the transition from using the two subchannels associated with the two largest singular values to beamforming along the single subchannel associated with the largest singular value:

$$b_{BF} \leq b_1 + b_2\ , \tag{5.77}$$

where b_{BF} is the number of bits supported when beamforming and b_1 and b_2 are the number of bits supported on the first and second subchannels, respectively, when multiplexing. The threshold can be obtained by considering the case where $b_2 = 1$ and b_{BF} and b_1 are as large as possible given the available power and the SER target $\mathcal{E}$. Using (5.57), we express (5.77) at the threshold as

$$-\frac{1}{\alpha}\ln\left[\frac{\sigma_n^2}{P_{s,BF}}\left(Q^{-1}\left(\frac{\mathcal{E}}{4}\right)\right)^2\right] \leq -\frac{1}{\alpha}\ln\left[\frac{P_{IN,1}}{P_{s,1}}\left(Q^{-1}\left(\frac{\mathcal{E}}{4}\right)\right)^2\right] + 1, \tag{5.78}$$

which reduces to

$$\left(\frac{P_{s,1}}{P_{IN,1}}\right) \geq \left(\frac{P_{s,BF}}{\sigma_n^2}\right) e^{-\alpha}. \tag{5.79}$$

The interference power $P_{IN,1}$ is a function of the transmit power allocated to the second subchannel, $[\mathbf{P}]_{22}^2$. Using the exact relationship

$$\text{SER} = Q\left(\sqrt{2\,\text{SINR}}\right) \tag{5.80}$$

for the single bit-per-symbol case, the SINR expression in (5.32) for the MCSI case, and the two-subchannel power relationship $[\mathbf{P}]_{11}^2 = P_T - [\mathbf{P}]_{22}^2$, we find after some manipulation that at the threshold SER$= \mathcal{E}$,

$$[\mathbf{P}]_{22}^2 = \frac{(P_T(\lambda_{21}\sigma_{H,1}^2 + \sigma_{H,2}^2) + \sigma_n^2)(Q^{-1}(\mathcal{E}))^2}{(\lambda_{21}\sigma_{H,1}^2 + \sigma_{H,2}^2)(Q^{-1}(\mathcal{E}))^2 + \sigma_{H,1}^2 + 2(\sigma_{H,2}^2 + \sigma_2^2)} \tag{5.81}$$

in order to support a single bit, where

$$\lambda_{ij} = \frac{\sigma_i^2(\sigma_j^2 + \sigma_i^2)}{(\sigma_j^2 - \sigma_i^2)^2}. \tag{5.82}$$

Plugging (5.81) into (5.79) via (5.25) and (5.32) and collecting $\sigma_{H,1}^2$ terms, we obtain a quadratic inequality in $\sigma_{H,1}^2$,

$$a_2\sigma_{H,1}^4 + a_1\sigma_{H,1}^2 + a_0 \leq 0, \tag{5.83}$$

where

$$a_2 = P_T^2(Q^{-1}(\mathcal{E}))^2\lambda_{12}\lambda_{21}, \tag{5.84}$$

$$a_1 = P_T^2(Q^{-1}(\mathcal{E}))^2(\lambda_{21}+\lambda_{12})\sigma_{H,2}^2 - P_T(e^\alpha - 1)\sigma_n^2 + P_T(Q^{-1}(\mathcal{E}))^2(\lambda_{21}+\lambda_{12})\sigma_n^2, \tag{5.85}$$

and

$$a_0 = (Q^{-1}(\mathcal{E}))^2(P_T^2\sigma_{H,2}^4 + e^\alpha\sigma_n^4) - 2P_T(e^\alpha - 1)\sigma_2^2\sigma_n^2 + 2P_T((Q^{-1}(\mathcal{E}))^2 - e^\alpha + 1)\sigma_n^2\sigma_{H,2}^2. \tag{5.86}$$

Since $\sigma_{H,1}^2$ must be real and nonnegative, and since $a_1, a_2 > 0$ when $\mathcal{E} \leq 0.15$ (we assume the desired SER threshold will be less than this value), the resulting threshold on $\sigma_{H,1}^2$ is given by

$$\sigma_{H,1}^{2*}(\sigma_{H,2}^2, \sigma_n^2) = \frac{-a_1 + \sqrt{a_1^2 - 4a_2a_0}}{2a_2}, \tag{5.87}$$

where multiplexing is beneficial when $\sigma_{H,1}^2 < \sigma_{H,1}^{2*}$ and beamforming is optimal otherwise.

If the above threshold is negative, then beamforming is always throughput optimal. Necessary conditions on $\sigma_{H,2}^2$ and σ_n^2 for multiplexing can be found by examining the condition that guarantees the non-negativity of $\sigma_{H,1}^{2*}$, which is $a_1^2 - 4a_2a_0 \geq a_1^2$. Since $a_2 > 0$, this condition is equivalent to $a_0 \leq 0$. Setting (5.86) less than or equal to zero yields the following quadratic inequality for $\sigma_{H,2}^2$:

$$b_2\sigma_{H,2}^4 + b_1\sigma_{H,2}^2 + b_0 \leq 0, \tag{5.88}$$

where

$$b_2 = P_T^2(Q^{-1}(\mathcal{E}))^2, \tag{5.89}$$

$$b_1 = 2P_T((Q^{-1}(\mathcal{E}))^2 - e^\alpha + 1)\sigma_n^2, \tag{5.90}$$

and

$$b_0 = (Q^{-1}(\mathcal{E}))^2e^\alpha\sigma_n^4 - 2P_T(e^\alpha - 1)\sigma_2^2\sigma_n^2. \tag{5.91}$$

Since $\sigma_{H,2}^2$ must be real and nonnegative, and since $b_2 > 0$, and $b_1 > 0$ for $\mathcal{E} \leq 0.15$, the resulting threshold is given by

$$\sigma_{H,2}^{2*}(\sigma_n^2) = \frac{-b_1 + \sqrt{b_1^2 - 4b_2 b_0}}{2b_2}\ , \tag{5.92}$$

where a necessary condition for multiplexing to be beneficial is $\sigma_{H,2}^2 < \sigma_{H,2}^{2*}$; otherwise, beamforming is optimal.

Once again, non-negativity requirements come into play, this time for $\sigma_{H,2}^2$. The non-negativity of $\sigma_{H,2}^2$ requires that $b_0 \leq 0$, or equivalently,

$$(Q^{-1}(\mathcal{E}))^2 e^{\alpha} \sigma_n^4 - 2P_T(e^{\alpha} - 1)\sigma_2^2 \sigma_n^2 \leq 0. \tag{5.93}$$

Ignoring the trivial solution of $\sigma_n^2 = 0$, (5.93) may be rewritten to form a threshold for the measurement noise:

$$\sigma_n^{2*} = \frac{2P_T(e^{\alpha} - 1)\sigma_2^2}{(Q^{-1}(\mathcal{E}))^2 e^{\alpha}}, \tag{5.94}$$

where for $\sigma_n^2 > \sigma_n^{2*}$, beamforming will always be optimal. Note that when $\sigma_{H,2}^2 = 0$, this same threshold on σ_n^2 will be obtained from the $a_0 \leq 0$ condition above.

Figure 5.4 depicts the multiplexing region as a function of $\sigma_{H,1}^2$ and $\sigma_{H,2}^2$ for a particular 4×4 channel realization, assuming a noise variance of $\sigma_n^2 = -20$ dB. The shaded region in the figure represents the combinations of $\sigma_{H,1}^2$ and $\sigma_{H,2}^2$ for which multiplexing is advantageous.

Thresholds using the noise model of (5.36), the Gauss-Markov model, or the confidence level SINR expressions may be derived in a similar manner. See Appendix C for some additional examples.

5.5 Numerical Simulations

In this section, we examine the usefulness of the SINR expressions derived in Section 5.2 and the related power allocation schemes presented in Section 5.3. Additionally, we look at the performance of the multiplexing/beamforming thresholds presented in Section

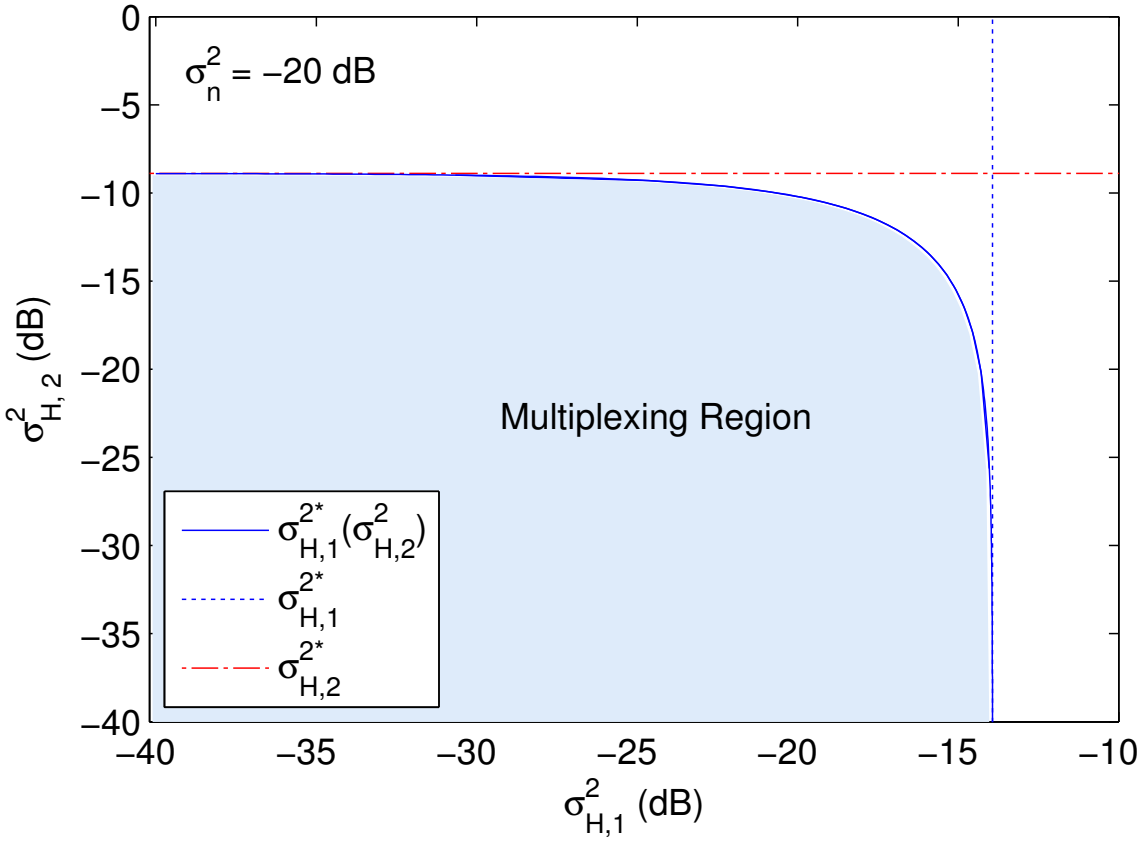


Figure 5.4: Example of multiplexing/beamforming thresholds on $\sigma_{H,1}^2$ and $\sigma_{H,2}^2$.

5.4 by noting how well they predict the transition from multiplexing to beamforming in the simulated power allocation schemes.

Five subchannel power allocation and bit-loading methods will be employed in the simulations. The first uses conventional waterfilling to determine subchannel power levels and bit loading is performed using (5.57) assuming perfect CSI. Since this method does not incorporate knowledge about the channel perturbations, we will refer to this method as waterfilling with uniformed bit loading (WFUBL). The next two methods are the ad-hoc methods discussed in Section 5.3.1. We refer to the first ad-hoc technique in this section as waterfilling with informed bit loading (WFIBL) since knowledge of the channel perturbation variances is used to determine the bit-loading levels and, in some cases, to adjust the subchannel power levels. We refer to the second ad-hoc method of Section 5.3.1 as power trimming with informed bit loading (PTIBL), its name referring to the algorithmic trim-

ming of excess power from each subchannel for use in other subchannels. The fourth power allocation method is uniform power allocation with informed bit loading (UPIBL), which was shown to have optimality properties in Section 5.3.2 for the CCSI scenario. We will, however, also employ this power allocation scheme for the MCSI scenario considered. The final power allocation method is beamforming, which we couple with informed bit loading and refer to as beamforming with informed bit loading (BFIBL).

In the simulations, the MIMO channel is 4×4 and selected such that $\text{vec}(\mathbf{H}) \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{I}_{16})$. The Gauss-Markov model is used to prevent power growth in the channel, thus producing simulations with more intuitive results. Note, however, that the results will be plotted versus the channel perturbation power to better match the discussion of the paper. The autoregressive parameter α_r is related $\sigma_{H,2}^2$ by

$$\alpha_r = \frac{1}{\sigma_{H,2}^2 + 1}. \tag{5.95}$$

At the value $\sigma_{H,2}^2 = 0$ dB ($\alpha_r = 0.5$), the channel perturbation is of the same variance, on average, as the channel coefficients themselves. Unless otherwise mentioned, $P_T = 1$, $\sigma_n^2 = -20$ dB at each receive antenna, and the measurement noise contributes to the channel perturbations with $K = 1$; in other words, $\sigma_{H,n}^2 = \sigma_n^2$. The target SER $\mathcal{E}$ is set at 0.01. The results of the simulations are averaged over 100 channel realizations with 100 Monte Carlo perturbations per channel realization.

We first consider simulation results for the CCSI channel perturbation scenario. Figure 5.5(a) displays, as a function of the channel perturbation variance $\sigma_{H,o}^2$, the sum of the bits allocated to each subchannel by the various power allocation and bit-loading methods. (Note that the results are plotted relative to $\sigma_{H,o}^2$ instead of the full $\sigma_{H,2}^2$ since the effect of measurement noise on the channel perturbations is a fixed effect given a specific K.) Note that due to outages, the average number of bits successfully sent per transmission will be less than the levels shown in this figure. In Fig. 5.5(b), we plot the average effective transmission rates, assuming that bits corresponding to SER outages are lost.

WFUBL allocates the same number of bits per transmission (BPT) regardless of $\sigma_{H,o}^2$ since it does not take CSI errors into account. Of the informed methods, the PTIBL method

allocates the most BPT, and it remains superior or equivalent in this regard to the other methods over the entire range of considered channel perturbation powers. This performance advantage is expected due to the power trimming, which assures all available power is directed towards bit allocation. For this reason, in the absence of measurement-noise-induced channel perturbations, PTIBL will also allocate more bits than WFUBL for low channel perturbation variances, though we don't show that scenario here. Interestingly, in the measurement-noise-free scenario, the UPIBL method also allocates slightly more BPT than WFUBL at low channel perturbation levels, since unlike UPIBL, WFUBL is not optimized for the case of finite M-QAM constellations.

While allocating a large number of bits, these methods suffer from a large number of outages, drastically reducing their effective throughput. Of the informed methods, PTIBL, in particular, has a poor outage performance at low channel perturbation levels. This is result of its power trimming, which leaves no power margin to protect its transmissions against outages. UPIBL and WFIBL are best in terms of actual throughput.

The vertical lines in Figs. 5.5(a) and 5.5(b) indicate the location of the channel perturbation threshold $\sigma_{H,2}^{2*}$. In Fig. 5.5(a), we see that the threshold accurately determines the point at which the bit allocation switches from multiplexing to beamforming (the point at which the UPIBL allocation equals that of BFIBL). However, when outages are taken into account, the threshold is optimistically high, and it is evident in Fig. 5.5(b) that the switch to beamforming should have occurred earlier. The WFIBL method continues multiplexing beyond the threshold, resulting in significantly lower throughput. Note that overall, despite the throughput loss due to outages, the proposed UPIBL approach still provides valuable throughput robustness as compared with techniques that do not properly account for the channel perturbation.

Figure 5.6 shows simulation results when bit allocations and threshold evaluations are carried out for PTIBL and BFIBL using the SINR confidence level expression of (5.53) with $p = 0.1$. Note, in particular, that the outage probabilities for these two methods are close to 0.1 even though it is the SINR outages that are restricted to the 0.1 level, not the SER outages. It is (5.57), along with the power trimming, that provides the connection between these two types of outages, though the SER outage level will typically be less than that for

SINR since the number of bits chosen for each subchannel will be the floor of the fractional level chosen by (5.57). For UPIBL and WFIBL, this connection between outage types is less tight as a result of the excess power in the subchannels, with the SINR confidence level of (5.53) resulting in overly restrictive bit allocation. To strengthen this connection, albeit in an ad-hoc manner, we have modified the SINR confidence level expression of (5.53) by replacing the $M-1$ dividing the inverse chi-square function with M, and changing the probability level back from $1-0.5p$ to $1-p$. In the simulations, these modifications work well for moderate array sizes, as is demonstrated by the UPIBL results presented in Fig. 5.6. For comparison purposes, the WFIBL results shown are for the average, not confidence level, SINR design. Note that the multiplexing/beamforming threshold shown in Fig. 5.6 now accurately predicts the transition point between multiplexing and beamforming in terms of real BPT (see Fig. 5.6(b)), due to the reduction in outages.

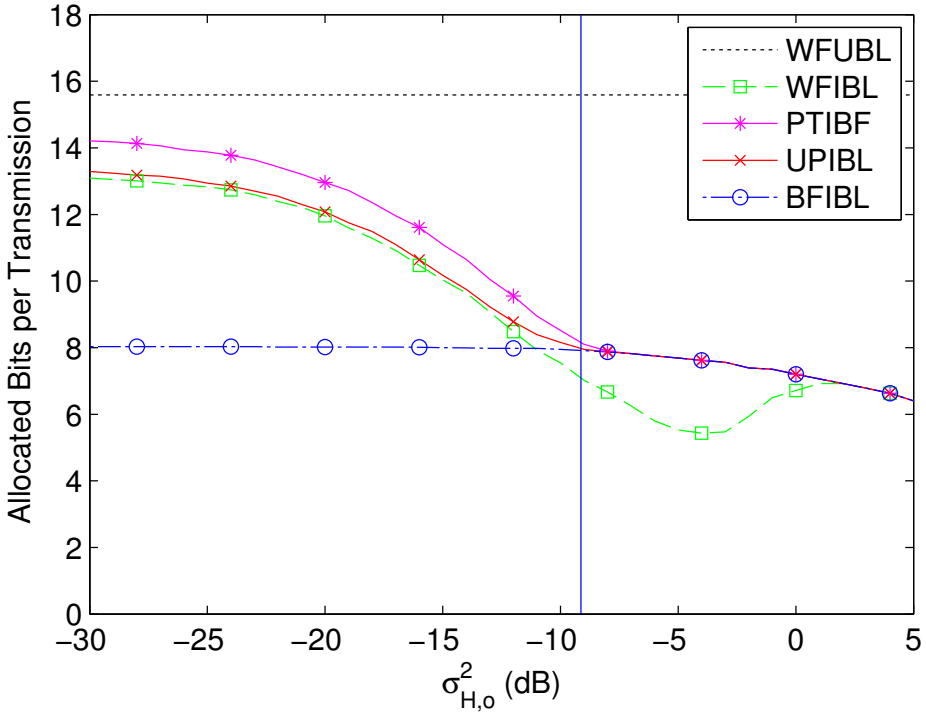
Finally, we consider results for an MCSI channel perturbation scenario in Fig. 5.7. We assume in this case that $\sigma_{H,2}^2 = \sigma_{H,n}^2$, and plot the results versus $\sigma_{H,1}^2$. The results are qualitatively similar to those given for the CCSI case. Note that relative to the comparable CCSI results shown in Figs. 5.5(a) and 5.5(b), the effect of channel perturbation variance on thoughput is noticeably more pronounced (see the discussion in Section 5.2.6). The threshold derived for the MCSI case appears to accurately represent the transition from multiplexing to beamforming when allocating bits, though it suffers from the same outage problem as in the CCSI case. We have found that simulations for the MCSI closely parallel those for the CCSI case, so we do not include additional MCSI results here.

In summary, we note that using the derived SINR expressions to allocate power and bits in MIMO-SVD systems appears to a very effective method for mitigating some of the effects of channel perturbation when perfect CSI is not available. Also note that, overall, the best performing power allocation and bit-loading methods appear to be the PTIBL and UPIBL methods. Though the UPIBL method was only shown to be optimal for bit allocation in the CCSI case when allowing for fractional bits, it is particularly interesting to observe that, despite its simplicity, UPIBL remains quite competitive with the more adaptive and computationally intensive PTIBL method when including the effect of outages, even for the MCSI case. As mentioned previously, much of this competitiveness is a result of the excess

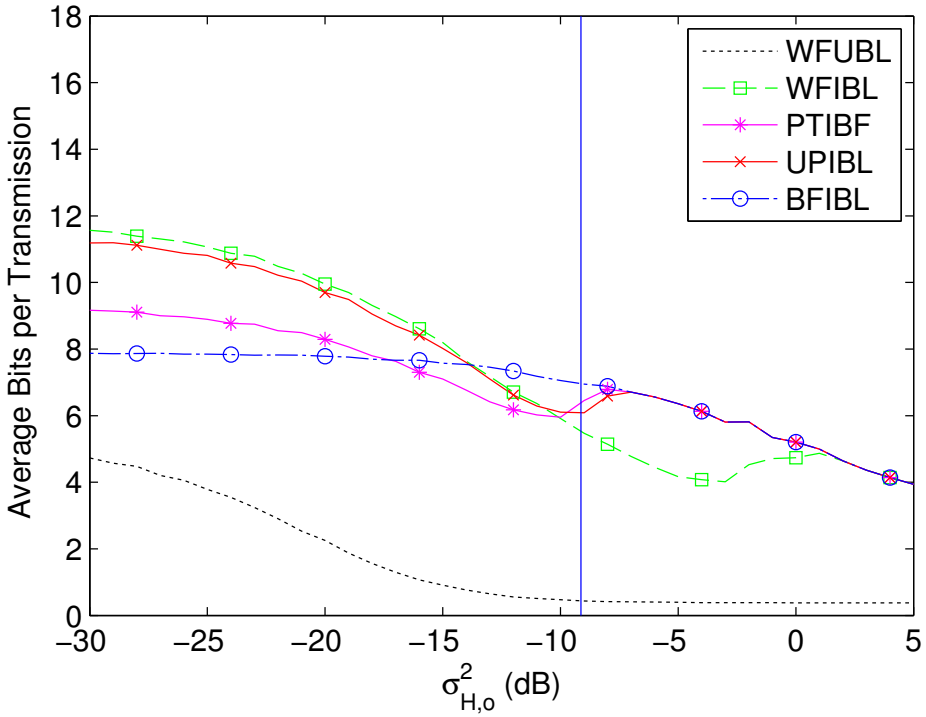
power UPIBL leaves on the subchannels, the very excess PTIBL attempts to remove. It is also interesting to consider that the power reallocation performed in PTIBL and WFIBL may tend to push these methods closer, at least on average, to a uniform power allocation.

5.6 Conclusions

In this chapter, the problem of MIMO multiplexing using SVD-based methods with imperfect CSI was considered. In order to account for the effects of the imperfect CSI, expressions for the per-subchannel signal power and interference-plus-noise power were derived which allow for the computation of a more realistic value for the per-subchannel SINR. These expressions, which allow for different levels of CSI imperfections at the transmitter and receiver, were shown to be valid over a wide range of useful channel perturbation powers. Assuming M-QAM modulation, methods were then presented that use these SINR expressions to address the subchannel power and bit allocation problems given an SER constraint. In particular, for the special case when identical, though imperfect, CSI is available at the transmitter and receiver, it was found that uniform power allocation is the optimal allocation strategy. Additionally, the SINR expressions were also used to derive thresholds on the channel perturbation and noise variances that indicate when to switch between spatial multiplexing and beamforming in order to maximize the number of bits per transmission given a constraint on symbol error rate (SER). Finally, the validity of these derived thresholds and usefulness of the presented power allocation methods were confirmed through numerical simulations.

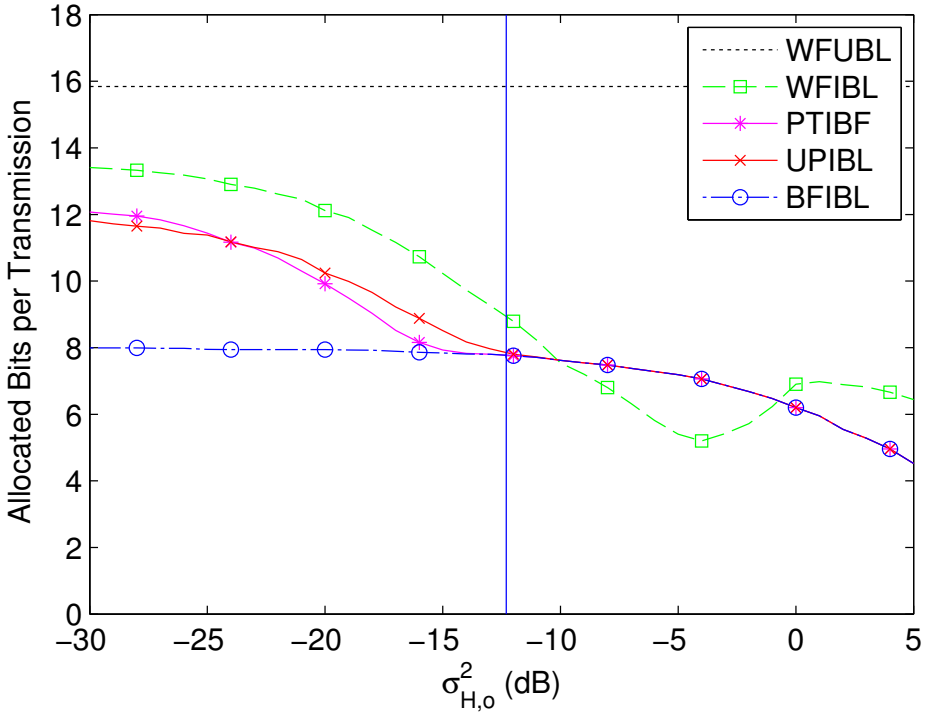


(a)

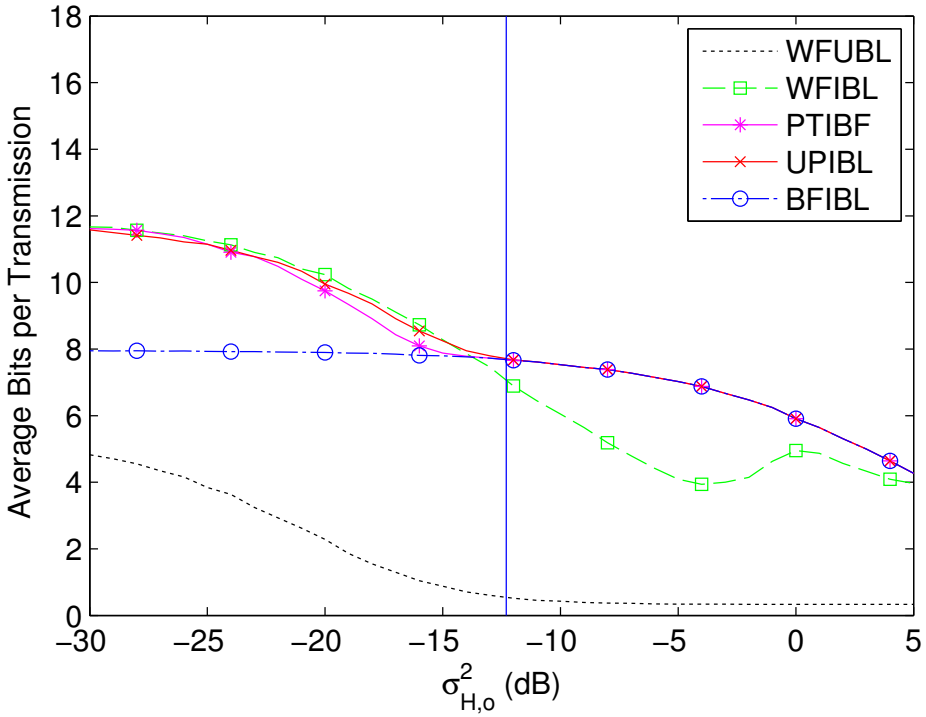


(b)

Figure 5.5: A comparison of (a) designed bit-loading levels and (b) average bits per transmission achieved for uniformed and informed bit-loading schemes.

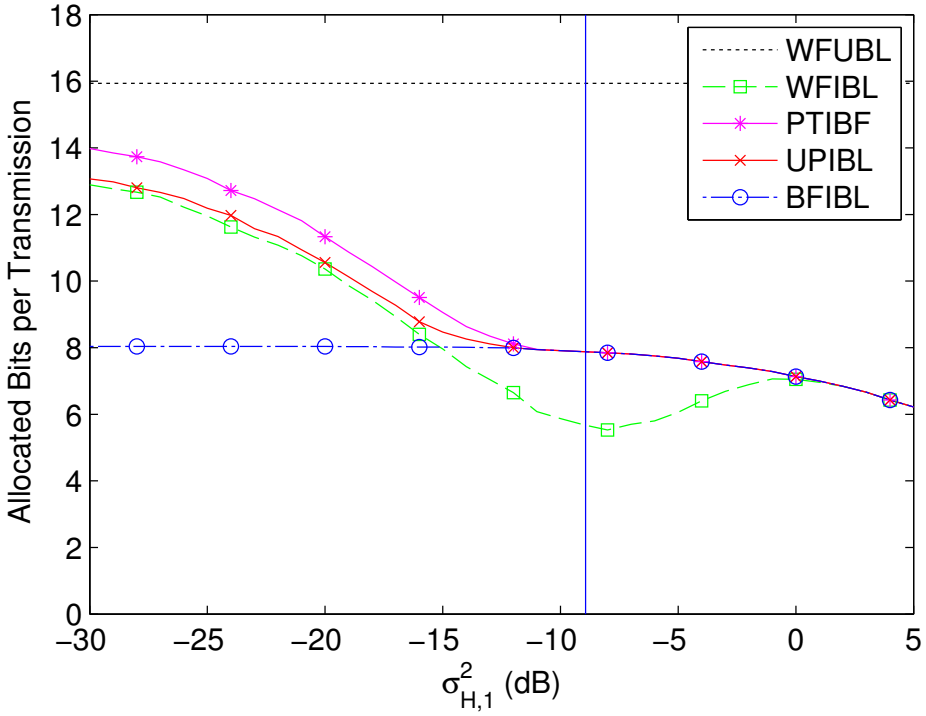


(a)

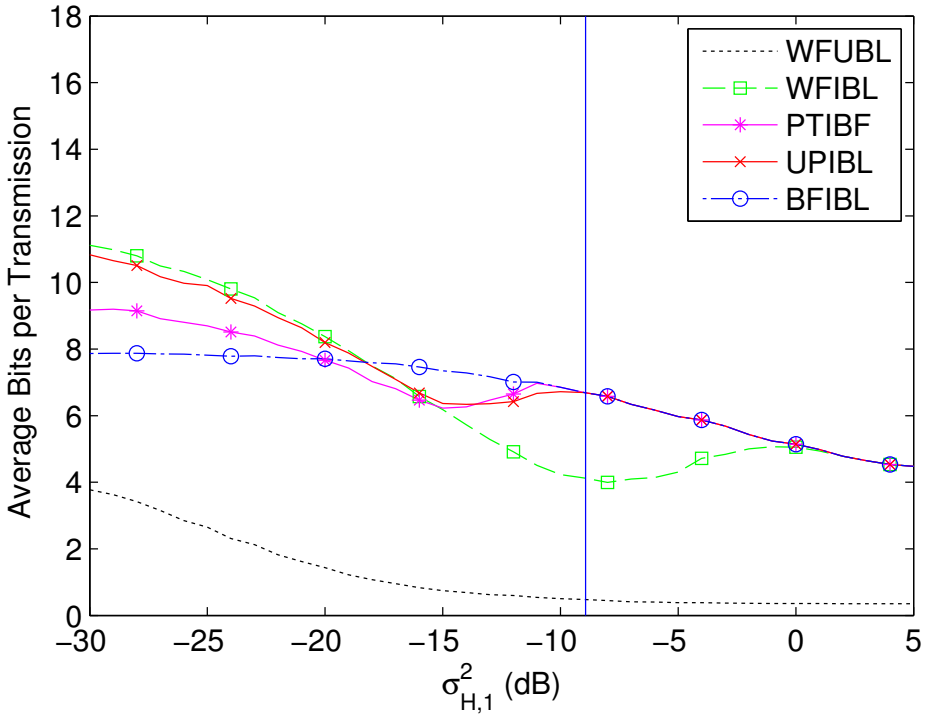


(b)

Figure 5.6: A comparison of (a) designed bit-loading levels and (b) average bits per transmission achieved for uniformed and informed bit-loading schemes with confidence level $p = 0.1$.



(a)



(b)

Figure 5.7: A comparison of (a) designed bit-loading levels and (b) average bits per transmission achieved for the MCSI scenario.

Chapter 6

Conclusions and Open Problems

6.1 Conclusions

Several topics related to the channel parameter estimation problem for OFDM, MIMO, and MIMO-OFDM systems were presented in this dissertation. These topics included discussions on optimal pilot design for channel and delay estimation in OFDM systems, channel estimation performance bounds for several parameter-based MIMO-OFDM channel models, and robust bit-loading and power-allocation design for MIMO-SVD systems with imperfect channel estimates. The contributions made in these areas are summarized below.

Chapter 2 began with a review of notation and background information on areas related to OFDM, MIMO, and the CRB. Following the background material, specific models were presented including an OFDM signaling model, several parameterized MIMO-OFDM models, and a narrowband MIMO perturbation model. These models formed the basis of the discussions to follow.

Chapter 3 considered the design of optimal pilot signals for the joint estimation of time delay and channel impulse response (CIR) parameters in an OFDM communications system. In particular, this chapter addressed the problem of finding the pilot locations and power levels required to minimize the CIR estimation error variance while meeting a constraint on the error variance of the time delay estimate. Using the OFDM signaling model derived in Chapter 2, the standard CRB was derived relative to the both the time delay and channel impulse response. Based on a weighted trace of this CRB, a cost function was presented for the purpose of addressing the OFDM pilot design problem. While the optimal pilots could not be found directly, the terms from the cost expression were used to develop intuition concerning the general structure of the optimal solution.

As an alternative to the using the standard CRB as the basis for pilot selection, a second CRB was derived as a function of intermediate parameters. This CRB, while maintaining a similar structure to the first, led to a simplified method for finding optimal pilot configurations. Numerical simulations demonstrated the use of these CRBs in solving the pilot design problem and supported the conclusions of the paper concerning the structure of the optimal pilots.

Chapter 4 presented lower bounds derived for several parametric time-varying wideband MIMO-OFDM channel models. These bounds were derived using vector formulations for the Cramér-Rao lower bound for functions of parameters for both unbiased and biased estimators. For the case of biased estimators, a new CRB formulation was obtained following the function-of-parameters approach (see Appendix A). This formulation allowed the CRB expressions to be obtained for the biased estimator considered. Numerical evaluations of these bounds illuminated some interesting features regarding the estimation and prediction of MIMO-OFDM channels. First, the numerical analysis indicated that better channel estimation and prediction performance may be realized through the use of antenna arrays at the transmitter and receiver for MIMO-OFDM systems. This result, an extension of a similar result for narrowband MIMO systems [60], illustrated that MIMO-OFDM systems can operate with lower time and frequency pilot densities than corresponding SISO implementations, even after taking into account the larger number of pilot symbols required to estimate a MIMO channel.

Second, the results suggested that when appropriate, full parametric channel modeling provides a significant advantage in estimation and prediction performance when compared to less structured channel models. This advantage was particularly evident when comparing performance results for full parametric models with those obtained using simple unstructured interpolation schemes for estimating the channel.

And third, the simulations demonstrated that channel estimation methods based on DOD/DOA parameterizations are extremely sensitive to array calibration errors. On the other hand, approaches based on either vector or matrix spatial signatures were shown to be significantly more robust despite the imprecise way in which they characterized the frequency

dependence of the channel. In particular, the VSS channel model appeared to be a reasonable compromise in terms of complexity, performance, and robustness.

Finally, Chapter 5 considered the problem of MIMO multiplexing using SVD-based methods when only imperfect channel estimates are available. Expressions for the per-subchannel signal power and interference-plus-noise power were derived which accounted for the effects of the imperfect CSI. The derived expressions allowed more realistic values for the per-subchannel SINRs to be computed. These expressions, which allow for different levels of CSI imperfections at the transmitter and receiver, were shown to be valid over a wide range of useful channel perturbation powers.

Assuming M-QAM modulation, methods were then presented which used these SINR expressions to address the subchannel power and bit allocation problems given an SER constraint. For the special case when identical, though imperfect, CSI was available at the transmitter and receiver, it was found that uniform power allocation is the optimal allocation strategy. Additionally, the SINR expressions were also used to derive thresholds on the channel perturbation and noise variances. These thresholds indicated the channel perturbation and measurement noise variance levels at which the switch between spatial multiplexing and beamforming or vice versa is beneficial in terms of the number of bits per transmission. As with previous chapters, the validity and usefulness of the derived expressions were confirmed through numerical simulations.

6.2 Open Problems

Due to the usefulness and popularity of OFDM and MIMO technologies, continued research into the theory and applications of these systems is warranted. While the number of open problems in these areas is quite large, a few open problems related to the topics covered in this dissertation are worth mentioning here. In Chapter 2, the problem of finding optimal pilots signals for OFDM channel parameter estimation was discussed. Note that while the chapter focused on finding pilots for the joint estimation of the time delay and channel impulse response, there are several other parameters in OFDM systems that bear consideration, such as the carrier frequency and sampling frequency offsets. Such parameters could be considered along with those already studied in terms of the problem addressed in

Chapter 3. The topics of this chapter could be expanded in other ways, as well, addressing the estimation over multiple OFDM symbols, time-varying channels, and MIMO-OFDM systems.

In addition, the CRB is but one metric that could be used for the optimal pilot design problem. Other metrics such as minimum mean-square error and capacity could be used to study the pilot configuration problem. While this has been done for limiting cases in [20, 29] and others, such approaches may be worth extending to this joint parameter problem in some fashion.

In order to realize all of the benefits available through the use of optimal pilots such as those found in Chapter 3, it may be instructive to find estimators that achieve or approach the CRBs used to find these pilots. For the standard CRB formulation, such estimators may be found in a manner similar to the joint channel impulse response and carrier frequency offset estimator discussed by [29] and others. This estimator is not detailed in closed form, but may be found by solving a related ML problem numerically. For the CRB-FOP-based pilots, the problem of a CRB-achieving estimator is still open as it is not immediately clear how to solve the related ML problem.

In Chapter 4, the CRB as a function of parameters was used to analyze the performance of channel estimation relative to parameterized wideband MIMO models. Note that the functions of parameters used in this chapter were nonlinear. As a result, unbiased estimators will only be able to approach, but not achieve, the CRB levels given in this work [41]. Since tighter lower bounds than the CRB exist, an open problem is how to apply some of these other bounds, which are often more difficult to find, to this same problem.

As was the case in Chapter 3, it is also not clear how to find feasible estimators whose performance can approach the bounds in Chapter 4. The joint maximum likelihood estimator for this problem is computationally intractable for practical application and channel prediction efforts have not lived up to their promise to date. Thus, there is a need to find lower-computation alternatives that will allow estimation and prediction performance levels in practice to approach those promised by these bounds.

Also note that the bounds of Chapter 4 depend upon the assumed training structure used to gather CSI. Thus, the bounds of this chapter could potentially be used in the same

manner as those of Chapter 3, that is, in application to the problem of optimal pilot selection. Since this would be a very complex issue to study analytically, a numeric approach might be the most tractable initially.

Finally, there are a great deal of open problems remaining relative to understanding the impact of imperfect CSI on MIMO system performance. The work discussed in Chapter 5 focused on the specific case of MIMO SVD-based transmission, which, while simple, is not optimal in terms of capacity when there are imperfections in the CSI. While solutions exist for the limiting cases of this problem (see [65–68]), a solution for the general case has yet to be found. There are also opportunities for expanding the work detailed in Chapter 5 itself, such as extending the work from narrowband to wideband MIMO channels. As another example, some work has already been done to extend the work of this chapter to multiuser MIMO. At this point, though, the expressions derived, while detailing the subchannel SINR, do not appear to be useful in compensating for imperfections in the channel knowledge due to the additional challenges of multiuser MIMO systems. As of the present, the area of quantifying and compensating for imperfect CSI for multiuser MIMO problem is still relatively open.

Overall, many opportunities still exist to advance the study and performance of OFDM, MIMO, and MIMO-OFDM systems. Such work, whether directly tied to material studied in this dissertation or not, has the potential for meaningful impact as society's need for increasingly high-throughput communication technologies continues.

Bibliography

[1] R. W. Chang, "Synthesis of band limited orthogonal signals for multichannel data transmission," *Bell Systems Technical Journal*, vol. 45, pp. 1775–1796, Dec. 1966.

[2] B. R. Salzberg, "Performance of an efficient parallel data transmission system," *IEEE Transactions on Communications*, vol. COM-15, pp. 805–813, Dec. 1967.

[3] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communcations*. Boston, MA: Artech House, 2000.

[4] R. R. Mosier and R. Clabaugh, "Kineplex, a bandwidth efficient binary transmission system," *AIEE Transactions*, vol. 76, pp. 723–728, Jan 1958.

[5] "Orthogonal frequency division multiplexing," U.S. Patent 3,488,4555, Jan. 6, 1970.

[6] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency division multiplexing using the discrete Fourier transform," *IEEE Transactions on Communications*, vol. COM-19, pp. 628–634, Oct. 1971.

[7] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs," Internal Tech. Memo., June 1995.

[8] ——, "Capacity of multi-antenna Gaussian channels," *Euro. Trans. on Telecommun.*, vol. 10, pp. 585–595, June 1999.

[9] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, Mar. 1998.

[10] *Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Std. 802.16, Oct. 2004.

[11] *Air Interface for Broadband Wireless Access Systems*, IEEE Std. 802.16, Rev. 2/D3, Feb. 2008.

[12] "Mobile WiMax - part I: A technical overview and performance evaluation," WiMax Forum, Feb. 2006.

[13] 3GPP, TR 25.913, "Requirements for evolved UTRA and evolved UTRAN." [Online]. Available: www.3gpp.org

[14] 3GPP, TR 25.813, "E-UTRA and E-UTRAN; radio interface protocol aspects." [Online]. Available: www.3gpp.org

[15] E. Dahlman *et al.*, "The 3G long-term evolution – radio interface concepts and performance evaluation," in *Proc. IEEE Vehicular Technology Conference (VTC 2006–Spring)*, vol. 1, May 2006, pp. 137–141.

[16] H. Yeng, "A road to future broadband wireless access: MIMO-OFDM-based air interface," *IEEE Communications Magazine*, vol. 43, pp. 53–60, Jan. 2005.

[17] G. L. Stuber *et al.*, "Broadband MIMO-OFDM wireless communications," *Proc. of the IEEE*, vol. 92, pp. 271–294, Feb. 2004.

[18] S. Ohno and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *IEEE Transactions on Communcations*, vol. 50, no. 12, pp. 2113–2123, Dec. 2002.

[19] ——, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 2138–2145, Sept. 2004.

[20] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 6, pp. 1615–1624, June 2003.

[21] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," in *IEEE GLOBECOM '04*, vol. 1. Dallas, TX: IEEE, 29 Nov.-3 Dec. 2004, pp. 219–224.

[22] H. Minn, N. Al-Dhahir, and Y. Li, "Optimal training signals for MIMO OFDM channel estimation in the presence of frequency offset and phase noise," *IEEE Transactions on Communcations*, vol. 54, no. 10, pp. 1754–1759, Oct. 2006.

[23] T.-L. Tung, K. Yao, and R. E. Hudson, "Channel estimation and adaptive power allocation for performance and capacity improvement of multiple-antenna ofdm systems," in *Third IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, Taoyuan, Taiwan, Mar. 2001, pp. 82–84.

[24] J.-J. van de Beek, M. Sandell, M. Isaksson, and P. O. Börjesson, "Low-complex frame synchronization in OFDM systems," in *Fourth IEEE International Conference on Universal Personal Communications (ICUPC)*, Tokyo, Japan, Nov. 1995, pp. 982–986.

[25] J.-J. van de Beek, M. Sandell, and P. O. Börjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Transactions on Signal Processing*, vol. 45, no. 7, pp. 1800–1805, July 1997.

[26] A. N. Mody and G. L. Stüber, "Synchronization for MIMO OFDM systems," in *IEEE Global Telecommunications Conference, 2001 (GLOBECOM '01)*, vol. 1, San Antonio, TX, Nov. 2001, pp. 509–513.

[27] E. Zhou, X. Zhang, H. Zhao, and W. Wang, "Synchronization algorithms for MIMO OFDM systems," in *IEEE Wireless Communications and Networking Conference, 2005 (WCNC 2005)*, vol. 1, New Orleans, LA, Mar. 2005, pp. 18–22.

[28] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for ofdm," *IEEE Transactions on Communications*, vol. 45, no. 7, pp. 1613–1621, July 1997.

[29] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1910–1917, Nov. 2003.

[30] A. J. Coulson, "Maximum likelihood synchronization for OFDM using a pilot symbol: Algorithms," *IEEE Journal on Selected Areas of Communications*, vol. 19, no. 12, pp. 2486–2494, Dec. 2001.

[31] H. Minn, M. Zeng, and V. K. Bhargava, "On timing offset estimation for OFDM systems," *IEEE Communications Letters*, vol. 4, no. 7, pp. 242–244, July 2000.

[32] H. Minn, V. K. Bhargava, and K. B. Letaief, "A combined timing and frequency synchronization and channel estimation for OFDM," *IEEE Transactions on Communications*, vol. 54, no. 3, pp. 416–422, Mar. 2006.

[33] ——, "A robust timing and frequency synchronization for OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 822–839, July 2003.

[34] Y. Ogawa, K. Nishio, T. Nishimura, and T. Ohgane, "Channel and frequency offset estimation for a MIMO-OFDM system," in *IEEE 60th Vehicular Technology Conference, 2004 (VTC2004-Fall)*, vol. 2, Los Angeles, CA, Sept. 2004, pp. 1523–1527.

[35] A. Saemi, V. Meghdadi, J.-P. Cances, M. R. Zahabi, and J.-M. Dumas, "ML time-frequency synchronization for MIMO-OFDM systems in unknown frequency selective fading channels," in *Proc. 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC06)*, Helsinki, Finland, Sept. 2006, pp. 1–5.

[36] M. Ghogho, A. Swami, and G. B. Giannakis, "Optimized null-subcarrier selection for CFO estimation in OFDM over frequency-selective fading channels," in *Proc. IEEE Global Telecommunications Conference 2001 (GLOBECOM '01)*, vol. 1, San Antonio, TX, Nov. 2001, pp. 202–206.

[37] H. Minn and S. Xing, "An optimal training signal structure for frequency-offset estimation," *IEEE Transactions on Communcations*, vol. 53, no. 2, pp. 343–355, Feb. 2005.

[38] Y. Mostofi and D. C. Cox, "A robust timing synchronization design in OFDM systems–part I: Low-mobility cases," *IEEE Transactions on Wireless Communications*, vol. 6, no. 12, pp. 4329–4339, Dec. 2007.

[39] H. Zhou and Y.-F. Huang, "A maximum likelihood fine timing estimation for wireless OFDM systems," *IEEE Transactions on Broadcasting*, vol. 55, no. 1, pp. 31–41, Mar. 2009.

[40] W. J. Bangs, "Array processing with generalized beamformers," Ph.D. dissertation, Yale Univ., New Haven, CT, 1971.

[41] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Part I.* New York, NY: John Wiley and Sons, Inc., 1968.

[42] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Communications Surveys and Tutorials*, vol. 9, no. 2, pp. 18–48, 2nd Quarter 2007.

[43] Y. Li, L. Cimini, Jr., and N. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Transactions on Communications*, vol. 46, no. 7, pp. 902–915, July 1998.

[44] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Transactions on Signal Processing*, vol. 51, pp. 1615–1624, June 2003.

[45] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1065–1076, May 2002.

[46] Z. Liu, X. Ma, and G. B. Giannakis, "Space-time coding and Kalman filtering for time-selective fading channels," *IEEE Transactions on Communications*, vol. 50, pp. 183–186, Feb. 2002.

[47] A. Duel-Hallen, S. Hu, and H. Hallen, "Long-range prediction of fading signals," *IEEE Signal Processing Magazine*, vol. 17, pp. 62–75, May 2000.

[48] B. Narendran, J. Sienicki, S. Yajnik, and P. Agrawal, "Evaluation of an adaptive power and error control algorithm for wireless systems," in *Proc. IEEE Internation Conference on Communications (ICC'97)*, vol. 1, 1997, pp. 349–355.

[49] J. B. Andersen, J. Jensen, S. H. Jensen, and F. Fredriksen, "Prediction of future fading based on past measurements," in *Proc. IEEE VTC 1999 Fall*, Amsterdam, Sept. 1999, pp. 151–155.

[50] L. Dong, G. Xu, and H. Ling, "Prediction of fast fading mobile radio channels in wideband communication systems," in *Proc. IEEE Globecom '01*, San Antonio, TX, Nov. 2001, pp. 3287–3291.

[51] T. Ekman, "Prediction of mobile radio channels: Modeling and design," Ph.D. dissertation, Signals and Systems, Uppsala Univ., Sweden, 2002.

[52] J. Hwang and J. H. Winters, "Sinusoidal modeling and prediction of fast fading processes," in *Proc. IEEE Globecom '98*, Sidney, Nov. 1998, pp. 892–897.

[53] A. Arredondo, K. R. Dandekar, and G. Xu, "Vector channel modeling and prediction for the improvement of downlink received power," *IEEE Transactions on Communications*, vol. 50, pp. 1121–1129, July 2002.

[54] A. Swindlehurst and M. Larsen, "Multiple-pass decision-directed channel estimation for highly mobile MIMO communications," in *Proc. IEEE Workshoop on Sensor Array and Multichannel Sig. Proc.*, July 2006, pp. 219–223.

[55] I. Wong and B. Evans, "Exploiting spatio-temporal correlations in MIMO wireless channel prediction," in *Proc. IEEE GlobeCom*, Nov. 2006.

[56] J. Vanderpypen and L. Schumacher, "MIMO channel prediction using ESPRIT based techniques," in *Proc. IEEE Int'l Symposium on Personal, Indoor and Mobile Radio Commun.*, Sept. 2007.

[57] K. Huber and S. Haykin, "Improved Bayesian MIMO channel tracking for wireless communications: incorporating a dynamical model," *IEEE Trans. Wireless Comm.*, vol. 5, pp. 2458–2466, Sept. 2006.

[58] B. Balakumar, S. Shahbazpanahi, and T. Kirubarajan, "Joint MIMO channel tracking and symbol decoding using Kalman filtering," *IEEE Transactions on Signal Processing*, vol. 55, pp. 5873–5879, Dec. 2007.

[59] P. D. Teal and R. G. Vaughan, "Simulation and performance bounds for real-time prediction of the mobile multipath channel," in *Proc. IEEE SSP'01*, Singapore, Aug. 2001, pp. 548–551.

[60] T. Svantesson and A. L. Swindlehurst, "A performance bound for prediction of MIMO channels," *IEEE Transactions on Signal Processing*, vol. 54, pp. 520–529, Feb. 2006.

[61] O. Simeone and U. Spagnolini, "Lower bound on training-based channel estimation error for frequency-selective block-fading Rayleigh MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, pp. 3265–3277, Nov. 2004.

[62] G. D. Forney Jr. and M. V. Eyuboglu, "Combined equalization and coding using precoding," *IEEE Communications Magazine*, vol. 43, pp. 25–34, Dec. 1991.

[63] X. Zhang and B. Ottersten, "Power allocation and bit loading for spatial multiplexing in MIMO systems," in *Proc. IEEE Internation Conference on Acoustics, Sound, and Signal Processing (ICASSP '03)*, Honolulu, Hawaii, Apr. 2003, pp. V 53–56.

[64] Z. Zhou, B. Vucetic, M. Dohler, and Y. Li, "MIMO systems with adaptive modulation," *IEEE Transactions on Vehicular Technology*, vol. 54, pp. 1828–1842, Sept. 2005.

[65] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Transactions on Information Theory*, vol. 47, pp. 2632–2639, Sept. 2001.

[66] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanathi, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 648–702, June 2003.

[67] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 1165–1175, July 2004.

[68] S. H. Simon and A. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 406–417, Apr. 2003.

[69] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding," *IEEE Transactions on Signal Processing*, vol. 50, pp. 2599–2613, Oct. 2002.

[70] ——, "Adaptive modulation for multiantenna transmissions with channel mean feedback," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 1626–1636, Sept. 2004.

[71] L. M. Garth, P. J. Smith, and M. Shafi, "Exact symbol error probabilities for SVD transmission of BPSK data over fading channels," in *Proc. IEEE International Conference on Communication (ICC 2005)*, Seoul, Korea, May 2005, pp. 2271–2276.

[72] E. K. S. Au *et al.*, "BER analysis of MIMO-SVD systems with channel estimation error and feedback delay," in *Proc. IEEE International Conference on Communication (ICC 2007)*, Glasgow, Scotland, June 2007, pp. 4375–4380.

[73] H. Zhang, D. L. Goeckel, and G. Ananthaswamy, "Adaptive modulation in MIMO eigenbeamforming with outdated channel state information," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM 2006)*, San Fransisco, CA, USA, Nov. 2006, pp. 1–5.

[74] J. Du *et al.*, "Estimation of performance loss due to delay in channel feedback in MIMO systems," in *Proc. Vehicular Technology Conference (VTC2004-Fall)*, Los Angeles, CA, USA, Sept. 2004, pp. 1619–1622.

[75] Z. Zhou and B. Vucetic, "The effect of CSI imperfection on the performance of SVD based adaptive modulation in MIMO systems," in *Proc. IEEE International Symposium on Information Theory (ISIT 2004)*, Chicago, USA, July 2004, p. 317.

[76] ——, "Design of adaptive modulation using imperfect CSI in MIMO systems," *IEEE Electron. Lett.*, vol. 40, no. 17, pp. 1073–1075, Aug. 2004.

[77] J. Liu, X. Liu, and X. Ma, "First-order perturbation analysis of singular vectors in the singular value decomposition," *IEEE Transactions on Signal Processing*, vol. 56, pp. 3044–3049, July 2008.

[78] S. M. Kay, *Fundamentals of Statistical Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1993.

[79] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, 1st ed. Upper Saddle River, NJ: Prentice Hall, 2000.

[80] A. N. D'Andrea, U. Mangali, and R. Reggiannini, "The modified Cramer-Rao bound and its application to synchronization problems," *IEEE Transactions on Communications*, vol. 42, no. 2/3/4, pp. 1391–1399, Feb./Mar./Apr. 1994.

[81] J. Barabell, "Improve the resolution performance of eigenstructure-based direction finding algorithms," in *IEEE International Conference on Acoustics, Sound, and Signal Processing (ICASSP)*, Boston, MA, Apr. 1983, pp. 336–339.

[82] R. Roy and T. Kailath, "ESPRIT—estimation of signal parameters by rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.

[83] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York, NY: Wiley-Interscience, 1991.

[84] S. Vishwanath and A. Goldsmith, "Adaptive turbo-coded modulation for flat-fading channels," *IEEE Transactions on Communications*, vol. 51, no. 6, pp. 964–972, June 2003.

[85] W. Jakes, *Microwave Mobile Communications.* New York, NY: Wiley-Interscience, 1974.

[86] A. F. Molisch, M. Steinbauer, M. Toeltsch, E. Bonek, and R. S. Thomä, "Capacity of MIMO systems based on measured wireless channels," *IEEE Journal on Selected Areas in Communications*, vol. 20, pp. 561–569, Apr. 2002.

[87] A. Richter, D. Hampicke, G. Sommerkorn, and R. S. Thomä, "MIMO measurement and joint M-D parameter estimation of mobile radio channels," in *Proc. IEEE VTC 2001 Spring*, May 2001, pp. 214–218.

[88] J. W. Wallace, M. A. Jensen, A. L. Swindlehurst, and B. D. Jeffs, "Experimental characterization of the MIMO wireless channel: Data acquisition and analysis," *IEEE Transactions on Wireless Communications*, vol. 2, pp. 335–343, Mar. 2003.

[89] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY: McGraw-Hill, 2001.

[90] P. Stoica and R. Moses, *Introduction to Spectral Analysis.* Upper Saddle River, NJ: Prentice-Hall, 1997.

[91] M. Larsen, A. L. Swindlehurst, and T. Svantesson, "A performance bound for interpolation of MIMO-OFDM channels," in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2006, pp. 1801–1805.

[92] ——, "A performance bound for MIMO-OFDM channel estimation and prediction," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM08)*, Darmstadt, Germany, July 2008, pp. 141–145.

[93] M. D. Larsen, A. L. Swindlehurst, and T. Svantesson, "Performance bounds for MIMO-OFDM channel estimation," *IEEE Transactions on Signal Processing*, vol. 57, no. 5, pp. 1901–1916, May 2009.

[94] M. Larsen and A. L. Swindlehurst, "A MIMO channel perturbation analysis for robust bit loading," in *Proc. IEEE Internation Conference on Acoustics, Sound, and Signal Processing (ICASSP '09)*, Taipei, Taiwan, Apr. 2009, pp. 2825–2828.

[95] B. Friedlander and A. J. Weiss, "On the second-order statistics of the eigenvectors of sample covariance matrices," *IEEE Transactions on Signal Processing*, vol. 46, pp. 3136–3139, Nov. 1998.

[96] Z. Xu, "Perturbation analysis for subspace decomposition with applications in subspace-based algorithms," *IEEE Transactions on Signal Processing*, vol. 50, pp. 2820–2830, Nov. 2002.

[97] R. J. Vaccaro, "A second-order perturbation expansion for the SVD," *SIAM J. Matrix Anal. Appl.*, vol. 15, pp. 661–671, Apr. 1994.

[98] M. Kristensson and B. Ottersten, "A statistical approach to subspace based blind identification," *IEEE Transactions on Signal Processing*, vol. 46, pp. 1612–1623, June 1998.

[99] G. W. Stewart and J. Sun, *Matrix Perturbation Theory.* San Diego, CA: Academic Press, Inc., 1990.

[100] T. Svantesson and A. L. Swindlehurst, "A performance bound for prediction of MIMO channels," *IEEE Transactions on Signal Processing*, vol. 54, pp. 520–529, Feb. 2006.

[101] C. B. Peel and A. L. Swindlehurst, "Performance of space-time modulation for a generalized time-varying rician channel model," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 1003–1012, May 2004.

[102] W. C. Jakes, *Microwave Mobile Communications.* New York, NY: Wiley-IEEE Press, 1994.

[103] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed. Reading, MA: Addison-Wesley, 1994.

[104] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY: McGraw-Hill, 2001.

Appendix A

Derivation of a CRB for Biased Estimators

Let $\mathbf{R}$ and $\mathbf{V}$ be vector random variables associated with the channel measurements and calibration errors, respectively. We desire to develop a bound for the error variance of estimators with the property

$$E_{\mathbf{r},\mathbf{v}|\mathbf{\Theta}}[\hat{\mathbf{h}}(\mathbf{R})] = E_{\mathbf{v}}[\mathbf{h}(\mathbf{\Theta},\mathbf{V}) + \mathbf{b}(\mathbf{\Theta},\mathbf{V})], \tag{A.1}$$

where $\hat{\mathbf{h}}(\mathbf{R})$ is the channel estimate, which depends only on the channel measurements, $\mathbf{h}(\mathbf{\Theta},\mathbf{V})$ is the true channel, and $\mathbf{b}(\mathbf{\Theta},\mathbf{V})$ is the estimation bias. Note that the authors have not seen this particular CRB formulation previously in the literature. However, the setup and derivation follow the methods shown in [41,78] for related bounds. We present the derivation here for completeness. We will work first with the left-hand side (LHS). Taking the derivative with respect to the vector parameter $\mathbf{\Theta}$ results in

$$\begin{aligned}\frac{\partial}{\partial\mathbf{\Theta}}E_{\mathbf{r},\mathbf{v}|\mathbf{\Theta}}[\hat{\mathbf{h}}(\mathbf{R})] &= \frac{\partial}{\partial\mathbf{\Theta}}\int \hat{\mathbf{h}}(\mathbf{R})p(\mathbf{R},\mathbf{V}|\mathbf{\Theta})\,\partial\mathbf{V}\partial\mathbf{R} \\ &= \int \hat{\mathbf{h}}(\mathbf{R})\left(\frac{\partial p(\mathbf{R},\mathbf{V}|\mathbf{\Theta})}{\partial\mathbf{\Theta}}\right)^{H}\partial\mathbf{V}\partial\mathbf{R}\end{aligned} \tag{A.2}$$

assuming that the integral in the second equality converges. Similarly, we take the derivative of the right-hand side (RHS) of (A.1):

$$\begin{aligned}\frac{\partial}{\partial\mathbf{\Theta}}E_{\mathbf{v}|\mathbf{\Theta}}[\mathbf{h}(\mathbf{\Theta},\mathbf{V}) + \mathbf{b}(\mathbf{\Theta},\mathbf{V})] &= \frac{\partial}{\partial\mathbf{\Theta}}E_{\mathbf{r},\mathbf{v}|\mathbf{\Theta}}[\mathbf{h}(\mathbf{\Theta},\mathbf{V}) + \mathbf{b}(\mathbf{\Theta},\mathbf{V})] \\ &= \int p(\mathbf{R},\mathbf{V}|\mathbf{\Theta})\frac{\partial\mathbf{g}(\mathbf{\Theta},\mathbf{V})}{\partial\mathbf{\Theta}} \\ &\quad + \mathbf{g}(\mathbf{\Theta},\mathbf{V})\left(\frac{\partial p(\mathbf{R},\mathbf{V}|\mathbf{\Theta})}{\partial\mathbf{\Theta}}\right)^{H}\partial\mathbf{V}\partial\mathbf{R} \\ &= E_{\mathbf{v}|\mathbf{\Theta}}\left[\frac{\partial\mathbf{g}(\mathbf{\Theta},\mathbf{V})}{\partial\mathbf{\Theta}}\right] \\ &\quad + \int \mathbf{g}(\mathbf{\Theta},\mathbf{V})\left(\frac{\partial p(\mathbf{R},\mathbf{V}|\mathbf{\Theta})}{\partial\mathbf{\Theta}}\right)^{H}\partial\mathbf{V}\partial\mathbf{R},\end{aligned} \tag{A.3}$$

where $\mathbf{g}(\mathbf{\Theta}, \mathbf{V}) = \mathbf{h}(\mathbf{\Theta}, \mathbf{V}) + \mathbf{b}(\mathbf{\Theta}, \mathbf{V})$. Combining (A.2) with (A.3), multiplying on the left by the arbitrary row vector $\mathbf{a}^H$ and on the right by the arbitrary vector $\mathbf{d}$, and noting that

$$\frac{\partial p(\mathbf{R}, \mathbf{V}|\mathbf{\Theta})}{\partial \mathbf{\Theta}} = \frac{\partial \ln p(\mathbf{R}, \mathbf{V}|\mathbf{\Theta})}{\partial \mathbf{\Theta}} p(\mathbf{R}, \mathbf{V}|\mathbf{\Theta}), \tag{A.4}$$

we obtain

$$\int u(\mathbf{R}, \mathbf{V}, \mathbf{\Theta}) v(\mathbf{R}, \mathbf{V}, \mathbf{\Theta}) p(\mathbf{R}, \mathbf{V}|\mathbf{\Theta}) \, \partial \mathbf{V} \partial \mathbf{R} = \mathbf{a}^H E \left[\frac{\partial (\mathbf{h}(\mathbf{\Theta}, \mathbf{V}) + \mathbf{b}(\mathbf{\Theta}, \mathbf{V}))}{\partial \mathbf{\Theta}} \right] \mathbf{d}, \tag{A.5}$$

where

$$u(\mathbf{R}, \mathbf{V}, \mathbf{\Theta}) = \mathbf{a}^H [\hat{\mathbf{h}}(\mathbf{R}) - \mathbf{h}(\mathbf{\Theta}, \mathbf{V}) - \mathbf{b}(\mathbf{\Theta}, \mathbf{V})] \tag{A.6}$$

and

$$v(\mathbf{R}, \mathbf{V}, \mathbf{\Theta}) = \left(\frac{\partial \ln p(\mathbf{R}, \mathbf{V}|\mathbf{\Theta})}{\partial \mathbf{\Theta}} \right)^H \mathbf{d}. \tag{A.7}$$

Applying the Cauchy-Schwarz inequality, we find

$$E[u(\mathbf{R}, \mathbf{V}, \mathbf{\Theta})^2] E[v(\mathbf{R}, \mathbf{V}, \mathbf{\Theta})^2] \geq \left(\mathbf{a}^H E \left[\frac{\partial (\mathbf{h}(\mathbf{\Theta}, \mathbf{V}) + \mathbf{b}(\mathbf{\Theta}, \mathbf{V}))}{\partial \mathbf{\Theta}} \right] \mathbf{d} \right)^2. \tag{A.8}$$

Since $\mathbf{d}$ was arbitrary, we can select it to be anything we want that does not depend on specific realizations of $\mathbf{R}$ and $\mathbf{V}$. We choose

$$\mathbf{d} = \mathbf{J}(\mathbf{\Theta})^{-1} E \left[\frac{\partial (\mathbf{h}(\mathbf{\Theta}, \mathbf{V}) + \mathbf{b}(\mathbf{\Theta}, \mathbf{V}))}{\partial \mathbf{\Theta}} \right]^H \mathbf{a}, \tag{A.9}$$

where

$$\mathbf{J}(\mathbf{\Theta}) = E \left[\frac{\partial \ln p(\mathbf{R}, \mathbf{V}|\mathbf{\Theta})}{\partial \mathbf{\Theta}} \left(\frac{\partial \ln p(\mathbf{R}, \mathbf{V}|\mathbf{\Theta})}{\partial \mathbf{\Theta}} \right)^H \right]. \tag{A.10}$$

Plugging this result into (A.8) and simplifying, we obtain

$$E[u(\mathbf{R}, \mathbf{V}, \mathbf{\Theta})^2] \geq \mathbf{a}^H \mathbf{H}' \mathbf{B} \mathbf{H}'^H \mathbf{a}, \tag{A.11}$$

where $\mathbf{B}$ and $\mathbf{H}'$ are defined in (4.50) and (4.48), respectively. Applying (A.1) to the LHS of (A.11), we have

$$E[u(\mathbf{R}, \mathbf{V}, \mathbf{\Theta})^2] = \mathbf{a}^H \left(\mathbf{C}_{\hat{\mathbf{h}}} - E[\mathbf{b}\mathbf{b}^T] \right) \mathbf{a}, \tag{A.12}$$

where

$$\mathbf{C}_{\hat{\mathbf{h}}} = E \left[(\hat{\mathbf{h}}(\mathbf{R}) - \mathbf{h}(\mathbf{\Theta}, \mathbf{V}))(\hat{\mathbf{h}}(\mathbf{R}) - \mathbf{h}(\mathbf{\Theta}, \mathbf{V}))^H \right]. \tag{A.13}$$

Equations (A.11) and (A.12) lead directly to the bound in (4.49).

Appendix B

Derivation of CRB Matrices

B.1 Derivatives for the Jacobian and Fisher Matrices

This section presents expressions for the derivatives used to calculate the Jacobian matrices in (4.25) and (4.48) and the Fisher matrix of (4.26), i.e., $\frac{\partial \mathbf{C}}{\partial \theta_i}$ and $\frac{\partial \mathbf{h}(\omega,t)}{\partial \theta_i}$.

B.1.1 Derivatives with Respect to σ

The covariance $\mathbf{C} = \sigma\mathbf{I}$ depends only on σ, so all of the derivatives are zero except $\frac{\partial \mathbf{C}}{\partial \theta_1} = \frac{\partial \mathbf{C}}{\partial \sigma} = \mathbf{I}$. The channel vector $\mathbf{h}(\omega, t)$ does not depend on σ, so $\frac{\partial \mathbf{h}(\omega,t)}{\partial \sigma} = \mathbf{0}$.

B.1.2 Derivatives with Respect to Re[$\boldsymbol{\alpha}$] (DOD/DOA)

$$\begin{aligned}\frac{\partial \mathbf{h}(\omega,t)}{\partial \mathrm{Re}[\alpha_i]} &= \frac{\partial (\mathbf{A}_t \otimes \mathbf{A}_r \mathbf{X})\mathbf{w}}{\partial \mathrm{Re}[\alpha_i]} \\ &= \left(\mathbf{A}_t \otimes \mathbf{A}_r \frac{\partial \mathbf{X}}{\partial \mathrm{Re}[\alpha_i]}\right)\mathbf{w} \\ &= \left(\mathbf{A}_t \otimes \mathbf{A}_r \mathbf{O}_{ii}\right)\mathbf{w}, \end{aligned} \tag{B.1}$$

where $\mathbf{w} = \mathrm{vec}(\mathbf{W}(\omega, t))$ and $\mathbf{O}_{ij}$ is a matrix of all zeros except for a one at the ij-th position.

B.1.3 Derivatives with Respect to Im[$\boldsymbol{\alpha}$] (DOD/DOA)

$$\frac{\partial \mathbf{h}(\omega,t)}{\partial \mathrm{Im}[\alpha_i]} = \left(\mathbf{A}_t \otimes \mathbf{A}_r \frac{\partial \mathbf{X}}{\partial \mathrm{Im}[\alpha_i]}\right)\mathbf{w} = j\left(\mathbf{A}_t \otimes \mathbf{A}_r \mathbf{O}_{ii}\right)\mathbf{w} \tag{B.2}$$

B.1.4 Derivatives with Respect to $\boldsymbol{\omega}_d$

$$\frac{\partial \mathbf{h}(\omega,t)}{\partial \omega_{d,i}} = \left(\mathbf{A}_t \otimes \mathbf{A}_r \mathbf{X}\right)\frac{\partial \mathbf{w}}{\partial \omega_{d,i}} = \left(\mathbf{A}_t \otimes \mathbf{A}_r \mathbf{X}\right)\mathbf{d}_{\omega,i}, \tag{B.3}$$

where

$$\mathbf{d}_{\omega,i} = \frac{\partial \mathbf{w}}{\partial \omega_{d,i}} = \mathrm{vec}\left(-jte^{j\gamma_i(\omega,t)}\mathbf{O}_{ii}\right). \tag{B.4}$$

The derivatives for models other than the DOD/DOA follow directly from this derivation.

B.1.5 Derivatives with Respect to $\boldsymbol{\tau}$

$$\frac{\partial \mathbf{h}(\omega,t)}{\partial \tau_i} = (\mathbf{A}_t \otimes \mathbf{A}_r\mathbf{X})\,\mathbf{d}_{\tau,i}, \tag{B.5}$$

where

$$\mathbf{d}_{\tau,i} = \frac{\partial \mathbf{w}}{\partial \tau_i} = \mathrm{vec}\left(-j(\omega-\omega_c)e^{j\gamma_i(\omega,t)}\mathbf{O}_{ii}\right). \tag{B.6}$$

Note that to improve the conditioning of the Fisher matrix, the derivative may be taken with respect to τ(bandwidth) instead of with respect to τ alone.

B.1.6 Derivatives with Respect to $\boldsymbol{\Omega}_t$ (DOD/DOA)

$$\frac{\partial \mathbf{h}(\omega,t)}{\partial \Omega_{t,i}} = \left(\frac{\partial \mathbf{A}_t}{\partial \Omega_{t,i}} \otimes \mathbf{A}_r\mathbf{X}\right)\mathbf{w} = (\mathbf{D}_{t,i} \otimes \mathbf{A}_r\mathbf{X})\,\mathbf{w}, \tag{B.7}$$

where

$$\mathbf{D}_{t,i} = \frac{\partial \mathbf{A}_t}{\partial \Omega_{t,i}} = \left[\begin{array}{ccccccc} \mathbf{0} & \cdots & \mathbf{0} & \frac{\partial [\mathbf{A}_t]_{:,i}}{\partial \Omega_{t,i}} & \mathbf{0} & \cdots & \mathbf{0} \end{array}\right]. \tag{B.8}$$

Note that this expression is valid for any array type.

B.1.7 Derivatives with Respect to $\boldsymbol{\Omega}_\mathbf{r}$ (DOD/DOA)

$$\frac{\partial \mathbf{h}(\omega,t)}{\partial \Omega_{r,i}} = \left(\mathbf{A}_t \otimes \frac{\partial \mathbf{A}_r}{\partial \Omega_{r,i}}\mathbf{X}\right)\mathbf{w} = (\mathbf{A}_t \otimes \mathbf{D}_{r,i}\mathbf{X})\,\mathbf{w} \tag{B.9}$$

with $\mathbf{D}_{r,i}$ defined in a similar manner to (B.8).

B.1.8 Derivatives with Respect to $\mathbf{A}_t$ (VSS)

For the VSS model, the elements of $\mathbf{A}_t$ and $\mathbf{A}_r$ are themselves parameters. Hence, we must find derivatives with respect to the elements of these matrices.

$$\frac{\partial \mathbf{h}(\omega,t)}{\partial \mathrm{Re}[[\mathbf{A}_t]_{ij}]} = \frac{\partial\left((\mathbf{A}_t \otimes \mathbf{A}_r)\mathbf{w}\right)}{\partial \mathrm{Re}[[\mathbf{A}_t]_{ij}]} = (\mathbf{O}_{ij} \otimes \mathbf{A}_r)\,\mathbf{w}. \tag{B.10}$$

Likewise,

$$\frac{\partial \mathbf{h}(\omega,t)}{\partial \mathrm{Im}[[\mathbf{A}_t]_{ij}]} = j\,(\mathbf{O}_{ij} \otimes \mathbf{A}_r)\,\mathbf{w}. \tag{B.11}$$

B.1.9 Derivatives with Respect to $\mathbf{A}_r$ (VSS)

Following the same procedure as for $\mathbf{A}_t$, we obtain

$$\frac{\partial \mathbf{h}(\omega,t)}{\partial \mathrm{Re}[[\mathbf{A}_r]_{ij}]} = (\mathbf{A}_t \otimes \mathbf{O}_{ij})\,\mathbf{w} \tag{B.12}$$

and

$$\frac{\partial \mathbf{h}(\omega,t)}{\partial \mathrm{Im}[[\mathbf{A}_r]_{ij}]} = j\frac{\partial \mathbf{h}(\omega,t)}{\partial \mathrm{Re}[[\mathbf{A}_r]_{ij}]}. \tag{B.13}$$

B.1.10 Derivatives with Respect to A (MSS)

For the MSS model, we must take derivatives with respect to the LM_rM_t elements of the matrix $\mathbf{A}$ from (4.20).

$$\begin{aligned}\frac{\partial \mathbf{h}(\omega,t)}{\partial \mathrm{Re}[[\mathbf{A}]_{ij}]} &= \frac{\partial\left((\mathbf{I}_{M_t}\otimes\mathbf{A})\mathrm{vec}(\mathbf{b}\otimes\mathbf{I}_{M_t})\right)}{\partial \mathrm{Re}[[\mathbf{A}]_{ij}]} \\ &= (\mathbf{I}_{M_t}\otimes\mathbf{O}_{ij})\mathrm{vec}(\mathbf{b}\otimes\mathbf{I}_{M_t}).\end{aligned} \tag{B.14}$$

As in previous cases,

$$\frac{\partial \mathbf{h}(\omega,t)}{\partial \mathrm{Im}[[\mathbf{A}]_{ij}]} = j\frac{\partial \mathbf{h}(\omega,t)}{\partial \mathrm{Re}[[\mathbf{A}]_{ij}]}. \tag{B.15}$$

B.2 Derivation of CRB Submatrix

The derivatives with respect to the vector parameters $\boldsymbol{\alpha}$, $\boldsymbol{\omega}_d$, $\boldsymbol{\tau}$, and so forth, lead to a block structure in the Fisher information matrix. Simple closed-form expressions are obtained for these matrix subblocks by substituting the derivative expressions above into (4.26). We will now look at a representative subblock given by derivatives with respect to τ and $\boldsymbol{\Omega}_t$. First note that

$$2\mathrm{Re}\left[\frac{\partial \mathbf{h}^H}{\partial \tau_i}\mathbf{C}^{-1}\frac{\partial \mathbf{h}}{\partial \Omega_{t,j}}\right] = \frac{2}{\sigma}\sum_{n=1}^{N_M}\mathrm{Re}\left[\frac{\partial \mathbf{h}_n^H}{\partial \tau_i}\frac{\partial \mathbf{h}_n}{\partial \Omega_{t,j}}\right], \tag{B.16}$$

where $\mathbf{h}_n = \mathbf{h}(\omega_n, t_n)$ and the summation come as a result of the stacked nature of $\mathbf{h}$ and the diagonal structure of $\mathbf{C}$, which is a reflection of the independence of the measurement noise at different sample locations:

$$\frac{\partial \mathbf{h}_n^H}{\partial \tau_i}\frac{\partial \mathbf{h}_n}{\partial \Omega_{t,j}} = \mathbf{d}_{\tau,i}^H(\mathbf{A}_t\otimes\mathbf{A}_r\mathbf{X})^H(\mathbf{D}_{t,j}\otimes\mathbf{A}_r\mathbf{X})\mathbf{w} \tag{B.17}$$

$$= \mathbf{d}_{\tau,i}^H(\mathbf{A}_t^H\mathbf{D}_{t,j}\otimes\mathbf{X}^H\mathbf{A}_r^H\mathbf{A}_r\mathbf{X})\mathbf{w}\ . \tag{B.18}$$

Note that $\mathbf{d}_{\tau,i}$ and $\mathbf{w}$ are evaluated at (ω_n, t_n). We omit this dependence from the notation momentarily for simplicity. Expanding out the Kronecker product and summing over the matrix subblocks, we obtain

$$\frac{\partial \mathbf{h}_n^H}{\partial \tau_i}\frac{\partial \mathbf{h}_n}{\partial \Omega_{t,j}} = \sum_{p=1}^{L}\sum_{q=1}^{L}\breve{\mathbf{d}}_{\tau,ip}^H[\mathbf{A}_t^H\mathbf{D}_{t,j}]_{pq}\mathbf{X}^H\mathbf{A}_r^H\mathbf{A}_r\mathbf{X}\breve{\mathbf{w}}_q, \tag{B.19}$$

where

$$\breve{\mathbf{d}}_{\tau,ip} = [\mathbf{d}_{\tau,i}]_{((p-1)L+1):pL} = [-j\omega e^{j\gamma_i(\omega,t)}\mathbf{O}_{ii}]_{:,p} \tag{B.20}$$

and

$$\breve{\mathbf{w}}_q = [\mathbf{w}]_{((q-1)L+1):qL} = [\mathbf{W}]_{:,q}. \tag{B.21}$$

Note that the terms of (B.19) are only nonzero when $q = j$ due to $\mathbf{D}_{t,i}$ and $p = i$ since $\breve{\mathbf{d}}_{\tau,ip} = 0$ when $p \neq i$. Only one term remains and (B.19) becomes

$$\frac{\partial \mathbf{h}_n^H}{\partial \tau_i}\frac{\partial \mathbf{h}_n}{\partial \Omega_{t,j}} = [\mathbf{D}_\tau^H]_{ii}[\mathbf{A}_t^H\mathbf{D}_{t,j}]_{ij}[\mathbf{X}^H\mathbf{A}_r^H\mathbf{A}_r\mathbf{X}]_{ij}[\mathbf{W}]_{jj}, \tag{B.22}$$

where $\mathbf{D}_\tau = -j(\omega_n - \omega_c)\mathbf{W}$. Now that we have the individual derivative terms, the entire block is formed by arranging the individual terms to obtain

$$2\mathrm{Re}\left[\frac{\partial \mathbf{h}^H}{\partial \tau}\mathbf{C}^{-1}\frac{\partial \mathbf{h}}{\partial \mathbf{\Omega_t}}\right] = \frac{2}{\sigma}\mathrm{Re}\left[\sum_{n=1}^{N_M}\mathbf{D}_{\tau,n}^H(\mathbf{A}_t^H\mathbf{D}_t \odot \mathbf{X}^H\mathbf{A}_r^H\mathbf{A}_r\mathbf{X})\mathbf{W}_n\right], \tag{B.23}$$

where $\mathbf{W}_n = \mathbf{W}(\omega_n, t_n)$, $\mathbf{D}_{\tau,n}$ is defined similarly, and $\odot$ is the element-wise or Hadamard matrix product. The matrix $\mathbf{D}_t$ is given by

$$\mathbf{D}_t = \left[\begin{array}{cccc} \frac{\partial[\mathbf{A}_t]_{:,1}}{\partial\Omega_{t,1}} & \frac{\partial[\mathbf{A}_t]_{:,2}}{\partial\Omega_{t,2}} & \cdots & \frac{\partial[\mathbf{A}_t]_{:,L}}{\partial\Omega_{t,L}} \end{array}\right]. \tag{B.24}$$

All of the other subblocks of the Fisher matrix may be derived in a like manner.

The Jacobian matrices will also have a block structure. Simplified forms for these blocks may be found by following a procedure similar to the one above for the Fisher blocks.

B.3 Expressions for FIM Subblocks

B.3.1 DOD/DOA CRB

The Fisher information matrix for the DOD/DOA model is given in (4.41). The subblocks can be determined by following the example in Appendix B.2 and are given by

$$\mathbf{P}_1 = \frac{2}{\sigma}\sum_{n=1}^{N_M}\mathbf{W}_n^H(\mathbf{A}_t^H\mathbf{A}_t) \odot (\mathbf{A}_r^H\mathbf{A}_r)\mathbf{W}_n,$$

$$\mathbf{P}_2 = \frac{2}{\sigma}\sum_{n=1}^{N_M}\mathbf{W}_n^H(\mathbf{A}_t^H\mathbf{A}_t) \odot (\mathbf{A}_r^H\mathbf{A}_r\mathbf{X})\mathbf{D}_{\tau,n},$$

$$\mathbf{P}_3 = \frac{2}{\sigma}\sum_{n=1}^{N_M}\mathbf{W}_n^H(\mathbf{A}_t^H\mathbf{A}_t) \odot (\mathbf{A}_r^H\mathbf{A}_r\mathbf{X})\mathbf{D}_{\omega,n},$$

$$\mathbf{P}_4 = \frac{2}{\sigma}\sum_{n=1}^{N_M}\mathbf{W}_n^H(\mathbf{A}_t^H\mathbf{D}_t) \odot (\mathbf{A}_r^H\mathbf{A}_r\mathbf{X})\mathbf{W}_n,$$

$$\mathbf{P}_5 = \frac{2}{\sigma}\sum_{n=1}^{N_M}\mathbf{W}_n^H(\mathbf{A}_t^H\mathbf{A}_t) \odot (\mathbf{A}_r^H\mathbf{D}_r\mathbf{X})\mathbf{W}_n,$$

$$\mathbf{P}_6 = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathbf{D}_{\tau,n}^H(\mathbf{A}_t^H\mathbf{A}_t) \odot (\mathbf{X}^H\mathbf{A}_r^H\mathbf{A}_r\mathbf{X})\mathbf{D}_{\tau,n},$$

$$\mathbf{P}_7 = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathbf{D}_{\tau,n}^H(\mathbf{A}_t^H\mathbf{A}_t) \odot (\mathbf{X}^H\mathbf{A}_r^H\mathbf{A}_r\mathbf{X})\mathbf{D}_{\omega,n},$$

$$\mathbf{P}_8 = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathbf{D}_{\tau,n}^H(\mathbf{A}_t^H\mathbf{D}_t) \odot (\mathbf{X}^H\mathbf{A}_r^H\mathbf{A}_r\mathbf{X})\mathbf{W}_n,$$

$$\mathbf{P}_9 = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathbf{D}_{\tau,n}^H(\mathbf{A}_t^H\mathbf{A}_t) \odot (\mathbf{X}^H\mathbf{A}_r^H\mathbf{D}_r\mathbf{X})\mathbf{W}_n,$$

$$\mathbf{P}_{10} = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathbf{D}_{\omega,n}^H(\mathbf{A}_t^H\mathbf{A}_t) \odot (\mathbf{X}^H\mathbf{A}_r^H\mathbf{A}_r\mathbf{X})\mathbf{D}_{\omega,n},$$

$$\mathbf{P}_{11} = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathbf{D}_{\omega,n}^H(\mathbf{A}_t^H\mathbf{D}_t) \odot (\mathbf{X}^H\mathbf{A}_r^H\mathbf{A}_r\mathbf{X})\mathbf{W}_n,$$

$$\mathbf{P}_{12} = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathbf{D}_{\omega,n}^H(\mathbf{A}_t^H\mathbf{A}_t) \odot (\mathbf{X}^H\mathbf{A}_r^H\mathbf{D}_r\mathbf{X})\mathbf{W}_n,$$

$$\mathbf{P}_{13} = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathbf{W}_n^H(\mathbf{D}_t^H\mathbf{D}_t) \odot (\mathbf{X}^H\mathbf{A}_r^H\mathbf{A}_r\mathbf{X})\mathbf{W}_n,$$

$$\mathbf{P}_{14} = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathbf{W}_n^H(\mathbf{D}_t^H\mathbf{A}_t) \odot (\mathbf{X}^H\mathbf{A}_r^H\mathbf{D}_r\mathbf{X})\mathbf{W}_n,$$

and

$$\mathbf{P}_{15} = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathbf{W}_n^H(\mathbf{A}_t^H\mathbf{A}_t) \odot (\mathbf{X}^H\mathbf{D}_r^H\mathbf{D}_r\mathbf{X})\mathbf{W}_n,$$

where $\mathbf{D}_r$ and $\mathbf{D}_{\omega,n}$ are defined in a manner analogous to $\mathbf{D}_t$ and $\mathbf{D}_{\tau,n}$.

B.3.2 DOD/DOA CRB with Calibration Errors

If the Fisher information matrix for the DOD/DOA model with calibration errors is evaluated as required by (4.47), the Fisher subblocks are of the same form as those for the DOD/DOA model, and may be obtained by replacing $\mathbf{A}_t$ with $\mathbf{A}_t + \mathbf{V}_t$ and $\mathbf{A}_r$ with $\mathbf{A}_r + \mathbf{V}_r$ in those expressions. If the Fisher matrix is evaluated as required by (4.49) then the subblocks may be found by replacing $\mathbf{A}_t^H\mathbf{A}_t$ by $\mathbf{A}_t^H\mathbf{A}_t + M_t\sigma_t\mathbf{I}$ and $\mathbf{A}_r^H\mathbf{A}_r$ by $\mathbf{A}_r^H\mathbf{A}_r + M_r\sigma_r\mathbf{I}$ in the expressions given in Appendix B.3.1. Note that the bias terms $\mathbf{b}\mathbf{b}^H$ and $E[\mathbf{b}\mathbf{b}^H]$ in (4.47) and (4.49), respectively, do not have simple closed form expressions. However, the trace of these terms, as required for the normalized error bound, does and this difficulty is avoided.

B.3.3 VSS CRB

Following the same procedure used to find the CRB for the DOD/DOA model, a bound may be found for the VSS model. The Fisher matrix for the VSS model is

$$\mathbf{J}(\mathbf{\Theta}) = \begin{bmatrix} \frac{M_t M_r N_M}{\sigma^2} & \mathbf{0}^T \\ \mathbf{0} & \mathbf{J}_{22} \end{bmatrix}, \tag{B.25}$$

where

$$\mathbf{J}_{22} = \begin{bmatrix} \mathrm{Re}[\mathbf{P}_1] & \mathrm{Re}[\mathbf{P}_2] & \mathrm{Re}[\mathbf{P}_3] & -\mathrm{Im}[\mathbf{P}_3] & \mathrm{Re}[\mathbf{P}_4] & -\mathrm{Im}[\mathbf{P}_4] \\ \mathrm{Re}[\mathbf{P}_2^T] & \mathrm{Re}[\mathbf{P}_5] & \mathrm{Re}[\mathbf{P}_6] & -\mathrm{Im}[\mathbf{P}_6] & \mathrm{Re}[\mathbf{P}_7] & -\mathrm{Im}[\mathbf{P}_7] \\ \mathrm{Re}[\mathbf{P}_3^T] & \mathrm{Re}[\mathbf{P}_6^T] & \mathrm{Re}[\mathbf{P}_8] & -\mathrm{Im}[\mathbf{P}_8] & \mathrm{Re}[\mathbf{P}_9] & -\mathrm{Im}[\mathbf{P}_9] \\ -\mathrm{Im}[\mathbf{P}_3^T] & -\mathrm{Im}[\mathbf{P}_6^T] & \mathrm{Im}[\mathbf{P}_8] & \mathrm{Re}[\mathbf{P}_8] & \mathrm{Im}[\mathbf{P}_9] & \mathrm{Re}[\mathbf{P}_9] \\ \mathrm{Re}[\mathbf{P}_4^T] & \mathrm{Re}[\mathbf{P}_7^T] & \mathrm{Re}[\mathbf{P}_9^T] & \mathrm{Im}[\mathbf{P}_9^T] & \mathrm{Re}[\mathbf{P}_{10}] & -\mathrm{Im}[\mathbf{P}_{10}] \\ -\mathrm{Im}[\mathbf{P}_4^T] & -\mathrm{Im}[\mathbf{P}_7^T] & -\mathrm{Im}[\mathbf{P}_9^T] & \mathrm{Re}[\mathbf{P}_9^T] & \mathrm{Im}[\mathbf{P}_{10}] & \mathrm{Re}[\mathbf{P}_{10}] \end{bmatrix} \tag{B.26}$$

with

$$\mathbf{P}_1 = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathbf{D}_{\omega,n}(\mathbf{A}_t^H\mathbf{A}_t \odot \mathbf{A}_r^H\mathbf{A}_r)\mathbf{D}_{\omega,n},$$

$$\mathbf{P}_2 = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathbf{D}_{\omega,n}(\mathbf{A}_t^H\mathbf{A}_t \odot \mathbf{A}_r^H\mathbf{A}_r)\mathbf{D}_{\tau,n},$$

$$\mathbf{P}_3 = \frac{2}{\sigma}\sum_{n=1}^{N_M} (\mathbf{D}_{\omega,n}\mathbf{A}_r^H\mathbf{A}_r\mathbf{W}_n) \circledast \mathbf{A}_t^H,$$

$$\mathbf{P}_4 = \frac{2}{\sigma}\sum_{n=1}^{N_M} (\mathbf{D}_{\omega,n}\mathbf{A}_t^H\mathbf{A}_t\mathbf{W}_n) \circledast \mathbf{A}_r^H,$$

$$\mathbf{P}_5 = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathbf{D}_{\tau,n}(\mathbf{A}_t^H\mathbf{A}_t \odot \mathbf{A}_r^H\mathbf{A}_r)\mathbf{D}_{\tau,n},$$

$$\mathbf{P}_6 = \frac{2}{\sigma}\sum_{n=1}^{N_M} (\mathbf{D}_{\tau,n}\mathbf{A}_r^H\mathbf{A}_r\mathbf{W}_n) \circledast \mathbf{A}_t^H,$$

$$\mathbf{P}_7 = \frac{2}{\sigma}\sum_{n=1}^{N_M} (\mathbf{D}_{\tau,n}\mathbf{A}_t^H\mathbf{A}_t\mathbf{W}_n) \circledast \mathbf{A}_r^H,$$

$$\mathbf{P}_8 = \frac{2}{\sigma}\sum_{n=1}^{N_M} (\mathbf{W}_n^H\mathbf{A}_r^H\mathbf{A}_r\mathbf{W}_n \otimes \mathbf{I}_{M_t}),$$

$$\mathbf{P}_9 = \frac{2}{\sigma}\sum_{n=1}^{N_M} (\mathbf{W} \otimes \mathbf{I}_{M_t})^H\mathbf{G}(\mathbf{W} \otimes \mathbf{I}_{M_r}),$$

and

$$\mathbf{P}_{10} = \frac{2}{\sigma}\sum_{n=1}^{N_M}(\mathbf{W}_n^H\mathbf{A}_t^H\mathbf{A}_t\mathbf{W}_n \otimes \mathbf{I}_{M_r}),$$

where

$$\mathbf{G} = \begin{bmatrix} \mathbf{a}_{t,1}\mathbf{a}_{r,1}^H & \cdots & \mathbf{a}_{t,L}\mathbf{a}_{r,1}^H \\ \vdots & \ddots & \vdots \\ \mathbf{a}_{t,1}\mathbf{a}_{r,L}^H & \cdots & \mathbf{a}_{t,L}\mathbf{a}_{r,L}^H \end{bmatrix}, \tag{B.27}$$

$\mathbf{a}_{t,i}$ and $\mathbf{a}_{r,i}$ are the i-th columns of $\mathbf{A}_t$ and $\mathbf{A}_r$, respectively, and $\circledast$ represents a matrix product defined for two matrices $\mathbf{A}$ and $\mathbf{B}$ as follows. If

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_M \end{bmatrix} \tag{B.28}$$

and

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_N \end{bmatrix}, \tag{B.29}$$

then

$$\mathbf{A} \circledast \mathbf{B} = \begin{bmatrix} \mathbf{a}_1 \odot \mathbf{b}_1 & \mathbf{a}_1 \odot \mathbf{b}_2 & \cdots & \mathbf{a}_1 \odot \mathbf{b}_N | & \cdots \\ & & & | \mathbf{a}_M \odot \mathbf{b}_1 & \mathbf{a}_M \odot \mathbf{b}_2 & \cdots & \mathbf{a}_M \odot \mathbf{b}_N \end{bmatrix}. \tag{B.30}$$

B.3.4 MSS CRB

The Fisher matrix for the MSS model is

$$\mathbf{J}(\mathbf{\Theta}) = \begin{bmatrix} \frac{M_t M_r N_M}{\sigma^2} & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T \\ \mathbf{0} & \mathrm{Re}[\mathbf{Q}_1] & -\mathrm{Im}[\mathbf{P}_1] & \mathrm{Re}[\mathbf{P}_2] & \mathrm{Re}[\mathbf{P}_3] \\ \mathbf{0} & \mathrm{Im}[\mathbf{P}_1] & \mathrm{Re}[\mathbf{P}_1] & \mathrm{Im}[\mathbf{P}_2] & \mathrm{Im}[\mathbf{P}_3] \\ \mathbf{0} & \mathrm{Re}[\mathbf{P}_2^T] & \mathrm{Im}[\mathbf{P}_2^T] & \mathrm{Re}[\mathbf{P}_4] & \mathrm{Re}[\mathbf{P}_5] \\ \mathbf{0} & \mathrm{Re}[\mathbf{P}_3^T] & \mathrm{Im}[\mathbf{P}_3^T] & \mathrm{Re}[\mathbf{P}_5^T] & \mathrm{Re}[\mathbf{P}_6] \end{bmatrix}, \tag{B.31}$$

and the individual matrix subblocks are given by

$$\mathbf{P}_1 = \frac{2}{\sigma}\sum_{n=1}^{N_M}\mathbf{b}_n^*\mathbf{b}_n^T \otimes \mathbf{I}_{M_t M_r},$$

$$\mathbf{P}_2 = \frac{2}{\sigma}\sum_{n=1}^{N_M}(\mathbf{B}_n^H \otimes \mathbf{I}_{M_t M_r})(\mathbf{1}_{L\times L} \otimes \mathbf{I}_{M_t M_r}) \begin{bmatrix} \mathrm{vec}(\mathbf{A}_1) & & \\ & \vdots & \\ & & \mathrm{vec}(\mathbf{A}_L) \end{bmatrix} \mathcal{D}_{\tau,n},$$

$$\mathbf{P}_3 = \frac{2}{\sigma}\sum_{n=1}^{N_M}(\mathbf{B}_n^H \otimes \mathbf{I}_{M_t M_r})(\mathbf{1}_{L\times L} \otimes \mathbf{I}_{M_t M_r}) \begin{bmatrix} \mathrm{vec}(\mathbf{A}_1) & & \\ & \vdots & \\ & & \mathrm{vec}(\mathbf{A}_L) \end{bmatrix} \mathcal{D}_{\omega,n},$$

$$\mathbf{P}_4 = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathcal{D}_{\tau,n}^H \begin{bmatrix} \mathrm{Tr}[\mathbf{A}_1^H\mathbf{A}_1] & \cdots & \mathrm{Tr}[\mathbf{A}_1^H\mathbf{A}_L] \\ \vdots & \ddots & \vdots \\ \mathrm{Tr}[\mathbf{A}_L^H\mathbf{A}_1] & \cdots & \mathrm{Tr}[\mathbf{A}_L^H\mathbf{A}_L] \end{bmatrix} \mathcal{D}_{\tau,n},$$

$$\mathbf{P}_5 = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathcal{D}_{\tau,n}^H \begin{bmatrix} \mathrm{Tr}[\mathbf{A}_1^H\mathbf{A}_1] & \cdots & \mathrm{Tr}[\mathbf{A}_1^H\mathbf{A}_L] \\ \vdots & \ddots & \vdots \\ \mathrm{Tr}[\mathbf{A}_L^H\mathbf{A}_1] & \cdots & \mathrm{Tr}[\mathbf{A}_L^H\mathbf{A}_L] \end{bmatrix} \mathcal{D}_{\omega,n},$$

and

$$\mathbf{P}_6 = \frac{2}{\sigma}\sum_{n=1}^{N_M} \mathcal{D}_{\omega,n}^H \begin{bmatrix} \mathrm{Tr}[\mathbf{A}_1^H\mathbf{A}_1] & \cdots & \mathrm{Tr}[\mathbf{A}_1^H\mathbf{A}_L] \\ \vdots & \ddots & \vdots \\ \mathrm{Tr}[\mathbf{A}_L^H\mathbf{A}_1] & \cdots & \mathrm{Tr}[\mathbf{A}_L^H\mathbf{A}_L] \end{bmatrix} \mathcal{D}_{\omega,n},$$

where $\mathbf{B}_n = \mathrm{diag}(\mathbf{b}_n)$, $\mathcal{D}_{\tau,n} = \partial\mathbf{B}_n/\partial\tau$, and $\mathcal{D}_{\omega,n} = \partial\mathbf{B}_n/\partial\omega$.

Appendix C

Additional Beamforming/Multiplexing Thresholds

C.1 Thresholds Accounting for Estimation Error

More precisely defined thresholds than those of (5.92) and (5.94) can be obtained if we separate $\sigma^2_{H,2}$ into its two component terms, as in (5.36). This is useful since $\sigma^2_{H,2}$ is itself a function of σ^2_n. In particular, if we assume that $\sigma^2_{H,n} = K\sigma^2_n$ for some constant K that depends on the particular channel estimation scheme, then using a derivation analogous to those previously given, we may derive the following beamforming/multiplexing thresholds for $\sigma^2_{H,o}$ and σ^2_n, which are given by

$$\sigma^{2*}_{H,o}(\sigma^2_n) = \frac{-c_1 + \sqrt{c_1^2 - 4c_2c_0}}{2c_2} \tag{C.1}$$

and

$$\sigma^{2*}_n = \frac{2P_T(e^\alpha - 1)\sigma_2^2}{(Q^{-1}(\mathcal{E}))^2(P_T^2K^2 + 2P_TK + e^\alpha) - 2P_T(e^\alpha - 1)K}\ , \tag{C.2}$$

where

$$c_2 = P_T^2(Q^{-1}(\mathcal{E}))^2, \tag{C.3}$$

$$c_1 = 2P_T^2(Q^{-1}(\mathcal{E}))^2\sigma^2_{H,n} + 2P_T((Q^{-1}(\mathcal{E}))^2 - e^\alpha + 1)\sigma^2_n, \tag{C.4}$$

and

$$\begin{aligned} c_0 = {} & (Q^{-1}(\mathcal{E}))^2(P_T^2\sigma^4_{H,n} + e^\alpha\sigma^4_n) - 2P_T(e^\alpha - 1)\sigma_2^2\sigma^2_n \\ & + 2P_T((Q^{-1}(\mathcal{E}))^2 - e^\alpha + 1)\sigma^2_n\sigma^2_{H,n}\ . \end{aligned} \tag{C.5}$$

Note that the threshold in (5.87) applies as is. If $\sigma^2_{H,n} = 0$ (or $K = 0$) then (C.1) and (C.2) reduce to (5.92) and (5.94), respectively.

C.2 Thresholds for the Gauss-Markov Channel Model

The derived thresholds may also be modified to accommodate the use of the Gauss-Markov channel mobility model in (5.37). For simplicity, we will only consider the CCSI case as discussed in Chapter 5.2.4. Extensions to the MCSI case depend on assumptions about how the Gauss-Markov model relates to $\Delta\mathbf{H}_1$ and $\Delta\mathbf{H}_2$. While the consideration of these issues was not addressed in this work, once such determinations are made, derivation of the corresponding thresholds is straightforward. For example, if $\Delta\mathbf{H}_1 = 0$, then the threshold

of (5.92) becomes a threshold on $1 - \alpha_r$ instead of $\sigma^2_{H,2}$. The only other modification needed is in (5.90) where $e^\alpha - 1$ should be replaced by $(e^\alpha - 1)(1 - \sigma_2^2)$. The threshold of (5.94) is applicable without modification.