

## ABSTRACT

WU, SHAOHAN. Joint Antenna Impedance and Channel Estimation in Multiple-input, Multiple-output Receivers. (Under the direction of Dr. Brian L. Hughes.)

Over the past two decades, several authors have demonstrated that impedance matching between the receive antenna and front-end can significantly impact channel capacity in wireless, multipath channels. In order to implement capacity-optimal matching, the receiver must know the antenna impedance. However, this impedance can often vary significantly with time in an unpredictable way in response to near-field loading of the antenna. To address this problem, we propose an approach to joint antenna impedance and channel estimation suitable for fading, wireless channels. The goal of this work is to use the extensive resources available in most wireless systems for channel estimation to estimate the antenna impedance as well as the channel. In this approach, we assume the receiver switches its load impedance in a predetermined way during reception of a known training sequence. Based on the observations of this training sequence at the receiver, we derive joint estimators for both the antenna impedance and channel, as well as Cramer-Rao lower bounds (CRB) on the error performance of any unbiased estimator.

First, we consider two different estimation approaches for single-antenna channels. The first is based on a hybrid Bayesian and classical estimation framework that leads to a joint maximum *a posteriori* and maximum-likelihood estimate of the channel and impedance, respectively. The estimators are simple and perform well at high signal-to-noise ratios; however, numerical results suggest that the hybrid estimators are generally inconsistent, and perform poorly at low signal-to-noise ratios. Our second approach focuses on classical estimation of the impedance, followed by minimum-mean-squared error estimation of the channel conditioned on knowledge of the antenna impedance. The latter

approach leads to a simple, efficient estimator based on a principal-components analysis (PCA) of the received data covariance matrix.

We then consider multiple-input, multiple-output (MIMO) wireless channels, with multiple antennas at the transmitter and/or receiver, as well as spatially and temporally correlated fading. We extend the PCA-based estimators to MIMO channels, and we also explore alternative iterative estimators based on the Newton-Raphson method, as well as simpler estimators based on the method of moments (MM). We investigate the performance of these estimators through a series of experiments, under conditions similar to those of long-term evolution (LTE) cellular channels. Our results suggest that the performance of our estimator is highly sensitive to the available diversity in the channel. In the presence of rich diversity, however, both estimators perform close to the CRB under a broad range of conditions of practical interest. Through numerical examples, we further show that these estimators can be used together with adaptive matching at the receiver to significantly enhance the signal-to-noise ratio and capacity of MIMO fading channels.

© Copyright 2019 by Shaohan Wu

All Rights Reserved

Joint Antenna Impedance and Channel Estimation in Multiple-input, Multiple-output  
Receivers

by  
Shaohan Wu

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

Electrical Engineering

Raleigh, North Carolina

2019

APPROVED BY:

---

Dr. Brian A. FLoyd

---

Dr. Cranos M. Williams

---

Dr. Luo Xiao

---

Dr. Brian L. Hughes  
Chair of Advisory Committee

# DEDICATION

To my parents.

## BIOGRAPHY

Shaohan Wu began working toward his Ph.D. in Electrical Engineering at North Carolina State University under the direction of Prof. Brian L. Hughes in 2014. During this doctoral program, he worked at Bose Corporation as a DSP intern from May 2016 to August 2016. Starting January 2019, he will become a staff engineering at MediaTek USA.

Shaohan's research interests include Statistical Signal Processing, Information Theory, and Wireless Communications.

## ACKNOWLEDGEMENTS

First and foremost, I thank my advisor Dr. Brian L. Hughes for his continuous guidance, support, and encouragement during this Ph.D. pursuit. His prowess as a mathematician and ability to communicate complex ideas in simple language tremendously helped me complete this dissertation. I am fortunate to observe him at work and follow in his footsteps.

My gratitude goes to Drs. Brian A. Floyd, Cranos M. Williams, and Xiao Luo, for kindly accepting the role as graduate advisory committee member and sharing their respective expertise. I also benefited from technical discussions and collaborations with many colleagues and (former) graduate students here at NC State, including Lopamudra Kundu, Wuyuan Li, Divyakumar Badheka, Charley Wilson, and Binbin Yang. Furthermore, financial support from the National Science Foundation and the state of North Carolina is deeply appreciated. I am grateful to many administrative staff across ECE department and in graduate school, who facilitated me meeting various milestones that eventually lead to this dissertation.

I owe gratitude to many friends and the local Buddhist community for memorable moments, especially services to society. It provided purposes in life outside my chosen profession. Last but very importantly, I am indebted to my parents for their love, understanding, and sacrifices since the very beginning of this life. Without them, any success of mine is unthinkable. This dissertation is dedicated to them.

# TABLE OF CONTENTS

<b>List of Figures</b> . . . . .	<b>vii</b>
<b>Chapter 1 Introduction</b> . . . . .	<b>1</b>
1.1 Relevant Prior Research . . . . .	4
1.1.1 Circuits Architectures and Algorithms for Adaptive Matching . . . . .	4
1.1.2 Adaptive Matching at the Transmitter . . . . .	5
1.1.3 Adaptive Matching or Impedance Estimation at the Receiver . . . . .	5
1.2 Technical Background . . . . .	7
1.2.1 Channel Capacity . . . . .	7
1.2.2 Impedance Matching . . . . .	10
1.2.3 Channel Estimation . . . . .	15
1.3 Outline of this Dissertation . . . . .	20
<b>Chapter 2 A Hybrid Approach to Joint Channel and Antenna Impedance Estimation</b> . . . . .	<b>23</b>
2.1 Introduction . . . . .	23
2.2 System Model . . . . .	24
2.3 Joint MAP and ML Estimators . . . . .	28
2.4 Estimators for Multiple Packets . . . . .	31
2.4.1 Estimators for General $\mathbf{C}_H$ . . . . .	32
2.4.2 Special Cases of $\sigma_n^{-2}\mathbf{C}_H$ . . . . .	35
2.5 Properties of the Joint MAP/ML Estimators . . . . .	36
2.5.1 Single-Packet Estimators . . . . .	37
2.5.2 Consistency . . . . .	39
2.5.3 Hybrid Cramér-Rao Bound . . . . .	40
2.6 Numerical Results . . . . .	41
2.7 Conclusions . . . . .	47
<b>Chapter 3 A Principal-Components Approach to Antenna Impedance Estimation</b> . . . . .	<b>49</b>
3.1 Introduction . . . . .	49
3.2 System Model . . . . .	51
3.3 Maximum-Likelihood Estimators . . . . .	55
3.4 Estimators for Multiple Packets . . . . .	64
3.4.1 ML Estimators for Temporally i.i.d. Fading . . . . .	81
3.4.2 ML Estimators for General $\mathbf{C}_h$ . . . . .	82
3.5 Numerical Results . . . . .	84
3.6 Conclusion . . . . .	95

<b>Chapter 4 Antenna Impedance Estimation at MIMO Receivers . . . . .</b>	<b>97</b>
4.1 Introduction . . . . .	97
4.2 System Model . . . . .	98
4.3 Maximum-Likelihood Estimators . . . . .	103
4.4 Estimators for Multiple Packets . . . . .	112
4.5 Numerical Results . . . . .	123
4.6 Conclusion . . . . .	131
<b>Chapter 5 Likelihood Ratio Tests for Detecting Antenna Impedance Variations . . . . .</b>	<b>132</b>
5.1 Introduction . . . . .	132
5.2 System Model . . . . .	133
5.3 Antenna Impedance Variation Detection . . . . .	136
5.4 Numerical Results . . . . .	139
5.5 Conclusion . . . . .	142
<b>Chapter 6 Conclusion . . . . .</b>	<b>144</b>
<b>References . . . . .</b>	<b>146</b>
<b>APPENDICES . . . . .</b>	<b>153</b>
Appendix A Proof of Theorem 3 in Chapter 2 . . . . .	154
Appendix B Details on Chapter 3 . . . . .	158
B.1 Another ML Estimator for $F$ . . . . .	158
B.2 Derivation of Complex Gradient and Complex Hessian . . . . .	159
Appendix C Details on Chapter 4 . . . . .	161
C.1 Details on Miller-Chang Bound . . . . .	161
C.2 Another Classical ML Estimator . . . . .	163

## LIST OF FIGURES

Figure 1.1	Circuit model of a single-antenna receiver . . . . .	11
Figure 2.1	Circuit model of a single-antenna receiver . . . . .	24
Figure 2.2	Relative MSE of $\hat{H}_{MAP}$ versus SNR. . . . .	42
Figure 2.3	Relative MAE of $\hat{F}_{ML}$ and $\hat{F}_U$ versus SNR. . . . .	43
Figure 2.4	Histograms of $\text{Re}[\hat{F}_{ML}]$ and $\text{Im}[\hat{F}_{ML}]$ for $L = 5, 10, 20$ . . . . .	44
Figure 2.5	MSE of $\hat{F}_{ML}$ and $\hat{F}_C$ versus SNR for selected $L > 2$ . . . . .	45
Figure 2.6	Relative Bias of $\hat{F}_{ML}$ versus SNR and $L$ . . . . .	46
Figure 2.7	MSE of $\hat{H}_{MAP}$ versus SNR for a correlated channel. . . . .	47
Figure 2.8	MAE of $\hat{F}_{ML}$ versus SNR for a correlated channel. . . . .	47
Figure 3.1	Circuit model of a single-antenna receiver . . . . .	51
Figure 3.2	Relative MSE of ML Estimators in i.i.d. Fading with $N = 4$ . . . . .	85
Figure 3.3	Relative RMSE and Absolute Bias of $\hat{F}_{MM}$ in Fast Fading, $N = 4$ . . . . .	87
Figure 3.4	Effects of Transmit Diversity under Medium Fading, $N = 2, 4$ . . . . .	89
Figure 3.5	Benefits of $\hat{F}_{ML}$ over $\hat{F}_{MM}$ in Slow Fading, $N = 1, 4$ . . . . .	90
Figure 3.6	Benefits of $\hat{F}_{MM}$ on MMSE Channel Estimation, $L = 10, N = 4$ . . . . .	92
Figure 3.7	Ergodic Capacity in (3.115), $L = 10, N = 4$ . . . . .	94
Figure 4.1	Circuit model of a multiple-antenna receiver . . . . .	99
Figure 4.2	Relative MSE of $\hat{\mathbf{F}}_{ML}$ versus SNR in i.i.d. Fading, $L = 5$ . . . . .	125
Figure 4.3	Properties of $\hat{\mathbf{F}}_{MM}$ for a 4 by 2 MIMO, $L = 10$ . . . . .	126
Figure 4.4	Excess Power of $\hat{\mathbf{F}}_{MM}$ for 4 by 2 MIMO. . . . .	127
Figure 4.5	Ergodic Capacity over SNR for 4 by 2 MIMO. . . . .	129
Figure 4.6	Ergodic Capacity over $d/\lambda$ for 4 by 2 MIMO, SNR = 10dB. . . . .	130
Figure 5.1	Circuit model of a multiple-antenna receiver . . . . .	133
Figure 5.2	RoC Curves for 4 by 2 MIMO under i.i.d. Fading. . . . .	140
Figure 5.3	RoC Curves for 4 by 2 MIMO under Correlated Rayleigh Fading. . . . .	143

# Chapter 1

## Introduction

Over the past two decades, several authors have demonstrated that antenna impedance matching at the receiver can significantly impact capacity and diversity in wireless communication channels [26–29, 51, 79]. For multiple-input, multiple-output (MIMO) channels, multi-port matching techniques that optimize capacity for different front-end configurations have been investigated in [26–29, 51, 79]. These works show that capacity is sensitive to receiver impedance matching, and optimal matching can dramatically increase the capacity of MIMO channels relative to conventional single-port matching.

To implement the capacity-optimal matching in [26–29, 51, 79], the receiver must know the antenna input impedance  $\mathbf{Z}_A$ . In practice, this is complicated by the fact that  $\mathbf{Z}_A$  varies with loading conditions in the antenna near-field. For example, the position of a hand on a cellular handset can significantly affect the input-impedance of handset antennas [7, 57, 74]. For this reason, several authors have proposed the use of adaptive matching networks that estimate and adapt to variations in antenna impedance. In [57, 74], the authors propose adaptive matching techniques based on empirical capacity estimates at the receiver. Capacity is optimized by periodically adjusting the receiver impedance, and

calculating the impact on empirical capacity estimates. Capacity is then optimized by searching for the receiver impedance that maximizes the observed capacity. This approach has several advantages: it is general, requires only information that is easily available at the receiver, and numerical results suggest it converges to the optimal capacity in a matter of seconds. On the other hand, because this approach requires a search of the receiver impedance space, it is also computationally-intensive, and convergence is slow compared to other common communication tasks, such as channel estimation, which often converges in milliseconds.

In this dissertation, we explore a new, more direct, approach to capacity-optimal adaptive matching, based on direct estimation of the receiver antenna impedance  $\mathbf{Z}_A$ . This approach seeks to first estimate  $\mathbf{Z}_A$  at the receiver, and then use the analytical results of [26–29, 51, 79] to calculate the capacity-optimal matching for this estimate. In contrast to the empirical capacity metric considered in [57, 74],  $\mathbf{Z}_A$  does not depend on the transmitted data, and thus should be easier to estimate. More importantly, since this approach *calculates* the optimal matching network, rather than searching over all possible networks, it should use far less data and computation than in [57, 74], and so has the potential to optimize capacity more quickly.

Most current wireless receivers have no mechanism to estimate antenna impedance. However, they do have extensive resources for channel estimation, in the form of training sequences and pilot symbols. These resources are often underutilized, in the sense that they are designed for worst-case conditions which rarely occur (e.g., high-speed trains in dense urban multi-path). In this work, we consider diverting some of these resources to estimate the receiver antenna impedance in addition to the channel path gains. We consider an estimation approach in which the receiver perturbs its impedance during reception of a known training sequence. Using these observations, the receiver

then performs joint channel and antenna impedance estimation based on the received data. Finally, the impedance estimates are used by the receiver to adaptively adjust the antenna matching network in order to maximize the resulting ergodic capacity, using the results of [26–29, 51, 79].

The problem of measuring impedance has been extensively investigated in the circuits literature; see [7, 8, 15, 33, 34, 65, 72, 85] and the references therein. Most of these works focus on measuring the impedance of the transmit antenna using special circuits that sense reflected power due to mismatch [15, 34, 65]. Among these papers, the most relevant to our work are [7, 8, 36], which consider dynamic measurement of the antenna impedance in handsets. In these works,  $\mathbf{Z}_A$  is measured by first perturbing the receiver impedance, and then calculating the resulting impact on average received power. In this work, we also perturb the receiver’s load impedance. However, rather than using special circuits, we estimate  $\mathbf{Z}_A$  by applying estimation theory to observations based on known training sequences. This approach has several important advantages: (1) it uses resources which are already present in most wireless systems; (2) no additional measurement circuits are needed, except the ability to perturb the receiver’s load; and (3) estimation theory provides tools to evaluate how performance depends on the training sequences and receiver perturbations, which may enable optimization over these parameters.

The rest of this introduction is organized as follows. In Sec. 1.1, we briefly review the relevant prior literature on adaptive matching and impedance estimation. In Sec. 1.2, we provide an overview of some important wireless performance metrics, such as channel capacity, and discuss how these metrics are affected by antenna impedance mismatch and channel estimation errors. This overview motivates the performance metrics used in later chapters of this dissertation.

## 1.1 Relevant Prior Research

Prior research on adaptive matching can be grouped into three broad categories: (1) circuit architectures and algorithms for adaptive matching; (2) adaptive matching at the transmitter, and (3) adaptive matching and/or antenna impedance estimation at the receiver. We now give a brief overview of the literature in each area.

### 1.1.1 Circuits Architectures and Algorithms for Adaptive Matching

In the circuits literature, several tunable matching network architectures have been proposed in order to realize adaptive or reconfigurable receivers. While this dissertation does not consider hardware design of matching networks, these works could provide a way to physically implement the adaptive matching component of this work. Most of the proposed architectures are based on a cascade of L-networks or  $\pi$ -networks [15, 33, 34, 65, 85], which usually consist of off-chip high-Q fixed inductors and tunable capacitors.

For certain tunable matching networks, several authors have also studied tuning algorithms to adjust the tunable network components to minimize the signal reflected by the matching network, as measured by the voltage standing wave ratio. Several authors have considered approaches based on the genetic algorithm and simulated annealing; for a discussion of these and related methods see [34]. In particular, Gu *et al* [34] have proposed an analytical algorithm that can adaptively adjust a tunable  $\pi$ -network to minimize the signal reflection for any given transmit antenna impedance.

### 1.1.2 Adaptive Matching at the Transmitter

In the circuits literature, several authors have also considered the application of adaptive matching at the transmitter [33,34,72]. In these works, the transmit antenna serves as the load for a power amplifier. These systems typically incorporate some kind of special circuit capable of detecting load mismatch, combined with a closed-loop control system that seeks to adjust a tunable matching network to maximize the radiated power. For further information, see [33,34,72] and references therein. Several types of special circuits have been considered to detect mismatch, such as directional couplers [15] and received power meters [34,65].

As in this dissertation, the authors above seek to adaptively adjust the antenna matching network in order to optimize system performance. However, the problems considered in [33,34,72] differ in two important ways from our work: First, all of the studies above focus on adaptive matching at the transmitter, where the transmitted signal is known and noise may be assumed to be negligible, whereas we consider the corresponding problem at the receiver, where the received signal is *a priori* unknown and corrupted by observation noise. Second, the studies above all employ special circuits to directly measure antenna mismatch, whereas we explore the possibility of estimating mismatch indirectly using observations of known training sequences at the receiver.

### 1.1.3 Adaptive Matching or Impedance Estimation at the Receiver

By contrast, relatively few papers have considered receivers equipped with tunable matching networks [7,36,57,74,75]. As noted earlier, in [57,74], the authors propose adaptive matching techniques based on empirical capacity estimates at the receiver. The match-

ing network seeks to optimize capacity by periodically adjusting the receiver impedance, and calculating the impact on an estimate of empirical capacity. Since the algorithms used in these papers involve either a random search or exhaustive sweep of the receiver impedance space, they tend to be computationally intensive and slow to converge to the optimal capacity, relative to other communication tasks.

More recently, two other studies have considered direct estimation of the receiver antenna impedance. In [7], Ali *et al* propose an approach to measure the antenna impedance that compares the received signal power at different frequencies and different loads. Perhaps most similar to this dissertation is the work of Hassan and Wittneben [36], who have considered joint channel and impedance estimation for MIMO receivers [36]. In this work, the authors vary the receiver load impedance and rely on known training sequences to jointly estimate the channel and the antenna impedance.

Similar to this dissertation, [7] and [36] both share the goal of estimating the antenna impedance directly, and both involve changing the receiver load to do so. However, these works differ significantly from this dissertation in modeling, technical approach and performance metrics. Ali *et al* assume a deterministic receiver circuit model that directly observes power, whereas we assume a noisy receiver which observes only the demodulated and sampled output of the receiver front-end. Their approach involves solving deterministic nonlinear equations via simulations and measurements, whereas our approach is grounded in estimation theory. The results of [36] also differ from ours in several important ways: First, the authors use a different model of noise in the receiver front end, which includes both antenna and load resistor noise, whereas we consider a scenario in which amplifier noise dominates. Second, [36] considers only uncorrelated, fast-fading conditions, whereas we consider the more general and realistic case of spatially and temporally-correlated fading. Third, [36] considers estimation of the channel fading

path gains and antenna impedance, whereas we consider a different parameterization of the channel, that estimates quantities that are more relevant to those needed by the communication algorithms, and which leads to better behaved estimators. Fourth, [36] considers least-squares estimation, whereas we consider maximum-likelihood estimation as well as fundamental lower bounds on performance, such as the Cramér-Rao bound (CRB). Finally, through numerical examples, we also investigate the impact the channel and impedance estimation error on the ergodic capacity and on the receiver signal-to-noise ratio when used in conjunction with adaptive matching.

## 1.2 Technical Background

In this dissertation, we study joint antenna impedance and channel estimation techniques for MIMO receivers and their impact on the channel capacity. In this section, we briefly review some important technical background that will be needed to correctly model the relationship between the channel capacity, impedance matching and channel estimation error in the chapters that follow.

### 1.2.1 Channel Capacity

Consider a point-to-point MIMO system in which a transmitter with  $N$  antennas sends data to a receiver with  $M$  antennas. In a narrowband channel, the signal observed at the receiver can usually be modeled as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} , \tag{1.1}$$

where  $\mathbf{x} \in \mathbb{C}^N$  is a vector of transmitted symbols,  $\mathbf{n} \in \mathbb{C}^M$  represents observation noise, and  $\mathbf{H} \in \mathbb{C}^{M \times N}$  is a matrix of channel fading path gains. When rich multipath is present and the transmit antennas are sufficiently separated, the columns of  $\mathbf{H}$ , denoted by  $\mathbf{h}_i$ , are often modeled as independent, zero-mean, circularly-symmetric Gaussian random vectors with a known covariance  $\mathbf{\Sigma}_{\mathbf{h}} = E[\mathbf{h}_i \mathbf{h}_i^H]$ . In the sequel, we denote this distribution by  $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{\Sigma}_{\mathbf{h}})$ . Similarly, the observation noise  $\mathbf{n}$  is also modeled as a zero-mean white Gaussian random vector,  $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}_M, \sigma_n^2 \mathbf{I}_M)$ .

For the MIMO receiver model (1.1), one of the most important metrics of wireless system performance is the (ergodic) channel capacity - the largest rate at which information can be reliably transmitted over this channel subject to a fundamental bound on the transmit power,  $E[\mathbf{x}^H \mathbf{x}] \leq P$ . According to information theory, the capacity can be calculated by maximizing the mutual information of the channel over all probability distributions of  $\mathbf{x}$  that satisfy the transmit power constraint. To define the capacity of (1.1), it is convenient to first consider the special case where  $\mathbf{H}$  is a known, deterministic matrix. In this case, the mutual information can be written as [73]

$$I(\mathbf{x}; \mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) = h(\mathbf{y}) - h(\mathbf{n}), \quad (1.2)$$

where  $h(\mathbf{y})$  is the differential entropy of the observation  $\mathbf{y}$  [73, Sec. II], and the second equality follows from the assumption that  $\mathbf{n}$  is independent of  $\mathbf{x}$ . Thus, maximizing  $I(\mathbf{x}; \mathbf{y})$  is equivalent to maximizing the differential entropy  $h(\mathbf{y})$ .

By Lemma 2 of Telatar [73], the maximum value of  $h(\mathbf{y})$  is always attained by a zero-mean Gaussian input distribution, say  $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{Q})$ , where  $\mathbf{Q} = E[\mathbf{x} \mathbf{x}^H]$  and the transmit power constraint implies  $\text{Tr}[\mathbf{Q}] \leq P$ . This implies that  $\mathbf{y}$  is also a zero-mean Gaussian vector with covariance  $E[\mathbf{y} \mathbf{y}^H] = \mathbf{H} \mathbf{Q} \mathbf{H}^H + \mathbf{I}_M$ , and hence the differential

entropy is

$$h(\mathbf{y}) = \log_2 \det (\pi e E[\mathbf{y}\mathbf{y}^H]) = \log_2 \det (\mathbf{H}\mathbf{Q}\mathbf{H}^H + \sigma_n^2 \mathbf{I}_M) + M \cdot \log_2(\pi e) . \quad (1.3)$$

It is straightforward to show that  $h(\mathbf{n}) = M \log_2(\pi e \sigma_n^2)$ , and hence the channel capacity can be expressed as

$$C = \max_{\mathbf{Q}: \text{Tr}[\mathbf{Q}] \leq P} \log_2 \det \left( \mathbf{I}_M + \frac{1}{\sigma_n^2} \mathbf{H}\mathbf{Q}\mathbf{H}^H \right) . \quad (1.4)$$

When  $\mathbf{H}$  is known to the transmitter, the optimal covariance  $\mathbf{Q}$  consists of a matrix that allocates power among the singular vectors of  $\mathbf{H}^H \mathbf{H}$  according to a water-filling power distribution [73, Sec. III-B].

In fast-fading scenarios, it is often more appropriate to model  $\mathbf{H}$  as a random matrix, that is known at the receiver but not at the transmitter [73, Sec. IV]. This is referred to as *channel side information at the receiver* (CSIR). If  $\mathbf{H}$  is random matrix, the capacity of the MIMO channel becomes [27]

$$C = \max_{\mathbf{Q}: \text{Tr}[\mathbf{Q}] \leq P} E \left[ \log_2 \det \left( \mathbf{I}_M + \frac{1}{\sigma_n^2} \mathbf{H}\mathbf{Q}\mathbf{H}^H \right) \right] , \quad (1.5)$$

where  $E[\cdot]$  denotes the expectation with respect to the probability distribution of  $\mathbf{H}$ . When the elements of  $\mathbf{H}$  are independent and identically distributed (i.i.d.), it was shown in [73, Th. 1] that the optimal covariance is given by

$$\mathbf{Q} = \frac{P}{N} \mathbf{I}_N , \quad (1.6)$$

where the total power  $P$  is equally distributed among the  $N$  transmit antennas. In [27],

this result was extended to any  $\mathbf{H}$  with i.i.d. columns, as considered here. In this case, the capacity reduces to [73, Th. 1]

$$C = E \left[ \log_2 \det \left( \mathbf{I}_M + \frac{P}{N\sigma_n^2} \mathbf{H}\mathbf{H}^H \right) \right]. \quad (1.7)$$

Note the capacity depends on the number of transmit and receive antennas, the transmit power constraint  $P$ , the observation noise variance  $\sigma_n^2$ , and also implicitly on the column covariance,  $\Sigma_{\mathbf{h}}$ . This last dependence can be expressed more explicitly by observing, as in [27], that  $\mathbf{H}_w \triangleq \Sigma_{\mathbf{h}}^{-1/2} \mathbf{H}$  is a random matrix with i.i.d.  $\mathcal{CN}(0, 1)$  elements. It follows that the ergodic capacity can be expressed in the alternate form [27, eq. 15]

$$C = E \left[ \log_2 \det \left( \mathbf{I}_M + \frac{P}{N\sigma_n^2} \Sigma_{\mathbf{h}} \mathbf{H}_w \mathbf{H}_w^H \right) \right], \quad (1.8)$$

from which it is clear that capacity depends only on  $N, M$  and the matrix  $\frac{P}{N\sigma_n^2} \Sigma_{\mathbf{h}}$ . In general,  $\Sigma_{\mathbf{h}}$  will depend on both the fading characteristics of the channel and the impedance matching between the receiver antenna and load. This dependence is explored in greater detail in the next section.

## 1.2.2 Impedance Matching

We now briefly review the dependence of the column covariance  $\Sigma_{\mathbf{h}}$  on the fading characteristics of the channel and impedance matching between the antenna and receiver front end.

First consider a receiver with a single antenna. The front-end of this receiver can be modeled as a load impedance  $Z_L$  connected to the receive antenna, as illustrated in Fig. 1.1. Here the antenna is represented by a Thevenin-equivalent impedance  $Z_A$  and

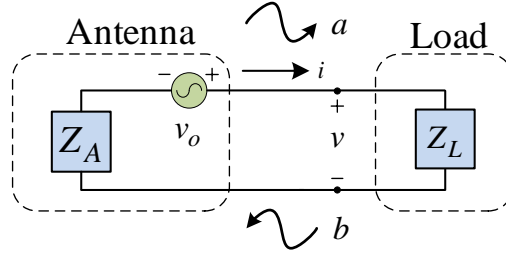


Figure 1.1: Circuit model of a single-antenna receiver

open-circuit voltage  $v_o$  [64, pg. 186]. From elementary circuit theory, the power delivered to the load in this system is [64, eq. 4.65],

$$P_L = \frac{1}{2} \operatorname{Re}\{v^* i\} = \frac{|v_o|^2 R_L}{2|Z_L + Z_A|^2}, \quad (1.9)$$

where  $R_L = \operatorname{Re}\{Z_L\}$ , and  $v$  and  $i$  are voltage across, and current into, the load  $Z_L$ , which are given by

$$v = \frac{v_o Z_L}{Z_L + Z_A}, \quad i = \frac{v_o}{Z_L + Z_A}. \quad (1.10)$$

In microwave network analysis, it is often more convenient to express circuit relationships in terms of generalized scattering parameters [64, eq. 4.60], defined by

$$a \triangleq \frac{v + i Z_R}{2\sqrt{R_R}}, \quad b = \frac{v - i Z_R^*}{2\sqrt{R_R}}, \quad (1.11)$$

where  $a$  and  $b$  are the *incident and reflected power waves*, respectively. Here  $Z_R = R_R + jX_R$  is a complex reference impedance, which can be chosen to suit the problem at hand. We can calculate the power delivered to the load directly from the power waves (1.11)

via the formula

$$\begin{aligned}
P_L &\triangleq \frac{1}{2} (|a|^2 - |b|^2) = \frac{1}{8R_R} (|v + iZ_R|^2 - |v - iZ_R^*|^2) \\
&= \frac{1}{8R_R} (v^*iZ_R + vi^*Z_R^* + v^*iZ_R^* + vi^*Z_R) \\
&= \frac{1}{2} \operatorname{Re}\{v^*i\} .
\end{aligned} \tag{1.12}$$

When there is no transmission line, the reference impedance is often chosen as the conjugate of load impedance [66],  $Z_R = Z_L^*$ . This leads to simplified expressions for the power waves [64, eq. 4.66],

$$a = v_o \frac{\sqrt{R_L}}{Z_L + Z_A}, \quad b = 0, \tag{1.13}$$

where the reflected power wave is always zero. As observed in [64, pg. 187],  $b = 0$  does not necessarily mean the load is matched to the generator; in fact, the power delivered to the load is maximized only when the antenna and load are conjugately matched,  $Z_L = Z_A^*$ . Nevertheless, power waves provide a convenient way to express power in terms of the squared-magnitudes of the incident and reflected waves.

Next, we consider a system with  $M$  antennas at the receiver. In this case, the antenna and load are described by  $M \times M$  complex impedances matrices,  $\mathbf{Z}_A$  and  $\mathbf{Z}_L$ , and the open circuit voltage is a complex  $M$ -vector  $\mathbf{v}_o$ . We assume that the open-circuit voltage can be modeled as

$$\mathbf{v}_o = \mathbf{G}\mathbf{x}, \tag{1.14}$$

where  $\mathbf{x}$  is the vector of transmitted symbols and  $\mathbf{G} \in \mathbb{C}^{M \times N}$  is a complex matrix of fading path gains that depends only on the fading characteristics of the channel. It follows that

the voltage observed across the load can be modeled as [29, 51, 79],

$$\mathbf{y} = \mathbf{R}_L^{1/2} (\mathbf{Z}_A + \mathbf{Z}_L)^{-1} \mathbf{G}\mathbf{x} + \mathbf{n} , \quad (1.15)$$

where  $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}_M, \sigma_n^2 \mathbf{I}_M)$  represents amplifier noise. It is easy to see that this observation model is consistent with the channel model (1.1) introduced in the last section, provided we define

$$\mathbf{H} = \mathbf{R}_L^{1/2} (\mathbf{Z}_A + \mathbf{Z}_L)^{-1} \mathbf{G} , \quad (1.16)$$

If the columns of  $\mathbf{G}$  are independent Gaussian random vectors, say  $\mathbf{g}_i \sim \mathcal{CN}(\mathbf{0}, \Sigma_{\mathbf{g}})$ , then the columns of  $\mathbf{H}$  will also be i.i.d. zero-mean vectors with covariance

$$\Sigma_{\mathbf{h}} = \mathbf{R}_L^{1/2} (\mathbf{Z}_A + \mathbf{Z}_L)^{-1} \Sigma_{\mathbf{g}} (\mathbf{Z}_A + \mathbf{Z}_L)^{-H} \mathbf{R}_L^{1/2} . \quad (1.17)$$

Note the original observation model is inferred from Ivrlač and Nossek [43, eq. 16],

$$\mathbf{v}'_L = \mathbf{Z}_L (\mathbf{Z}_A + \mathbf{Z}_L)^{-1} \mathbf{G}\mathbf{x} + \mathbf{Z}_L \mathbf{R}_L^{-1/2} \mathbf{n} , \quad (1.18)$$

where the  $\mathbf{Z}_L \mathbf{R}_L^{-1/2}$  before  $\mathbf{n}$  is the matrix generalization of [82, eq. 4], such that the variance of entries therein represents power. Pre-whitening the noise of this signal results in (1.15).

This signal model (1.15) properly connects the total received signal power, i.e., mean

of Frobenius norm of the signal [37, eq. 10], to that of circuit theory [43, eq. 22],

$$\begin{aligned}
P_r &\triangleq E [||\mathbf{H}\mathbf{x}||_F^2] \\
&= E \text{Tr} \left[ \mathbf{x}^H \mathbf{G}^H (\mathbf{Z}_A + \mathbf{Z}_L)^{-H} \mathbf{R}_L^{1/2} \mathbf{R}_L^{1/2} (\mathbf{Z}_A + \mathbf{Z}_L)^{-1} \mathbf{G} \mathbf{x} \right] \\
&= \text{Tr} \left( (\mathbf{Z}_A + \mathbf{Z}_L)^{-H} \mathbf{R}_L (\mathbf{Z}_A + \mathbf{Z}_L)^{-1} E [\mathbf{G} E[\mathbf{x}\mathbf{x}^H] \mathbf{G}^H] \right) \\
&= P \text{Tr} \left[ (\mathbf{Z}_A + \mathbf{Z}_L)^{-H} \mathbf{R}_L (\mathbf{Z}_A + \mathbf{Z}_L)^{-1} \boldsymbol{\Sigma}_g \right] = P \text{Tr} [\boldsymbol{\Sigma}_h] , \quad (1.19)
\end{aligned}$$

where an identity of Frobenius norm is applied in the second step [41, eq. 5.6.0.2],

$$||\mathbf{A}||_F^2 = \text{Tr}(\mathbf{A}^H \mathbf{A}) , \quad (1.20)$$

the second to last equality uses (1.6).

It has been shown that, for a given open-circuit voltage  $\mathbf{v}_o$ , multi-port conjugate matching is the optimal matching condition in terms of maximum power transfer [25, eq. 10] [64],

$$\mathbf{Z}_{L,opt} = \mathbf{Z}_A^* . \quad (1.21)$$

However, if optimizing ergodic capacity (1.8) is the goal, then the optimal matching condition may differ from (1.21), as ergodic capacity depends on the structure of the signal correlation matrix  $\boldsymbol{\Sigma}_h$  (1.17), not only its trace. The optimal load impedance  $\mathbf{Z}_{L,opt}$  then should depend on both  $\mathbf{Z}_A$  and  $\boldsymbol{\Sigma}_g$ .

When the incident signal arrives isotropically, the signal correlation is proportional to the antenna radiation resistance, which we approximate as its resistance [48, eq. 3.8], i.e.,  $\boldsymbol{\Sigma}_g \propto \mathbf{R}_A$ . With this assumption, the channel correlation matrix under conjugate matching (1.21) leads to an optimal signal correlation  $\boldsymbol{\Sigma}_{h,opt}$  that is proportional to  $\mathbf{I}_M$ . This optimal matching in (1.21) maximizes received power, decouples the signal, and

simultaneously maximizes ergodic capacity.

Thus far, we have assumed perfect knowledge of  $\mathbf{H}$  at the receiver. However, practical wireless communication systems estimate the fading channel with errors, despite extensive resources allocated to channel estimation. These channel estimation errors decrease ergodic capacity [37]. Thus, in the next section, we overview two important channel estimators and their mean-square errors.

### 1.2.3 Channel Estimation

Training-based MIMO channel estimation has been well studied [16]. Here we briefly review the linear least-squares (LS) and the linear minimum mean-square error (MMSE) estimators. Consider a frequency-flat block-fading MIMO system with  $N$  transmit and  $M$  receive antennas. From (1.1), we consider  $T \geq N$  training symbols under i.i.d. complex Gaussian noise,

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}, \quad (1.22)$$

where  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_T] \in \mathbb{C}^{N \times T}$  is subject to the same power constraint as in (1.1), i.e.,  $\text{Tr}(\mathbf{x}_i \mathbf{x}_i^H) \leq P$ , and  $\text{vec } \mathbf{N} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{MT})$ . The task of training-based channel estimation is to recover the channel information matrix  $\mathbf{H}$  from observations in  $\mathbf{Y}$ , using knowledge of the training sequence  $\mathbf{X}$ .

The LS estimator for  $\mathbf{H}$  is [16, eq. 4]

$$\hat{\mathbf{H}}_{LS} = \mathbf{Y}\mathbf{X}^H (\mathbf{X}\mathbf{X}^H)^{-1}. \quad (1.23)$$

The average mean-square error (MSE) of  $\hat{\mathbf{H}}_{LS}$  can be calculated as [16, eq. 7],

$$\begin{aligned}
J_{LS} &\triangleq \frac{1}{MN} E \left[ \|\mathbf{H} - \hat{\mathbf{H}}_{LS}\|_F^2 \right] \\
&= \frac{1}{MN} \text{Tr} \left[ (\mathbf{X}\mathbf{X}^H)^{-1} \mathbf{X} E[\mathbf{N}^H \mathbf{N}] \mathbf{X}^H (\mathbf{X}\mathbf{X}^H)^{-1} \right] \\
&= \frac{\sigma_n^2}{N} \text{Tr} \left[ (\mathbf{X}\mathbf{X}^H)^{-1} \right], \tag{1.24}
\end{aligned}$$

where  $E[\mathbf{N}^H \mathbf{N}] = M\sigma_n^2 \mathbf{I}_T$ . It has been shown that under power constraint  $\text{Tr}(\mathbf{X}\mathbf{X}^H) \leq T \cdot P$ , the MSE is minimized when orthogonal training sequences are used [16, eq. 8], i.e.,

$$\mathbf{X}\mathbf{X}^H = \frac{P \cdot T}{N} \mathbf{I}_N. \tag{1.25}$$

Among all possible orthogonal training sequences, a normalized submatrix of the discrete Fourier transform (DFT) matrix has entries with identical magnitude [16, eq. 10], e.g.,

$$\mathbf{X} = \sqrt{\frac{P}{N}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & W_T & \cdots & W_T^{T-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & W_T^{N-1} & \cdots & W_T^{(N-1)(T-1)} \end{bmatrix}, \tag{1.26}$$

where  $W_T = e^{j2\pi/T}$ . In other words, the  $i$ -th row  $j$ -th column entry of  $\mathbf{X}$  is  $X_{ij} = W_T^{(i-1)(j-1)}$  for  $1 \leq i \leq N$  and  $1 \leq j \leq T$ . The minimized MSE is

$$J_{LS,min} = \frac{\sigma_n^2}{N} \frac{N}{P \cdot T} \text{Tr}[\mathbf{I}_N] = \frac{N}{T} \frac{\sigma_n^2}{P}. \tag{1.27}$$

This can be interpreted as the average MSE (for each entry of  $\mathbf{H}$ ) is inversely proportional to SNR. Note the absolute (i.e., not relative) MSE in  $J_{LS,min}$  does not depend on the

receiver characteristics. Also, intuitively, it is scaled by the ratio of  $N/T$ , i.e., for fixed  $N$  more training symbols reduces MSE linearly. The optimal MSE for LS estimator in  $J_{LS,min}$  is often not the lowest. The linear MMSE estimator exhibits lower MSE than the LS, at the cost of losing local unbiasedness, which is a common approach in statistics, e.g., see [46] and references therein.

The linear MMSE can be defined as [16, eq. 22],

$$\hat{\mathbf{H}}_{MMSE} = \mathbf{Y}\mathbf{A}_o, \quad (1.28)$$

where  $\mathbf{Y}$  is the observation (1.22) and  $\mathbf{A}_o$  is chosen such that the MSE is minimized,

$$\mathbf{A}_o = \arg \min_{\mathbf{A} \in \mathbb{C}^{T \times N}} E [ \|\mathbf{H} - \mathbf{Y}\mathbf{A}\|_F^2 ] . \quad (1.29)$$

We explicitly write out the expectation of the MSE as a function of  $\mathbf{A}$ , using (1.20),

$$\begin{aligned} J(\mathbf{A}) &\triangleq E [ \|\mathbf{H} - \mathbf{Y}\mathbf{A}\|_F^2 ] = E \text{Tr} [ (\mathbf{H}^H - \mathbf{A}^H \mathbf{Y}^H) (\mathbf{H} - \mathbf{Y}\mathbf{A}) ] \\ &= \text{Tr} [ \mathbf{R}_{\mathbf{H}} - \mathbf{R}_{\mathbf{H}} \mathbf{X} \mathbf{A} - \mathbf{A}^H \mathbf{X}^H \mathbf{R}_{\mathbf{H}} + \mathbf{A}^H (\mathbf{X}^H \mathbf{R}_{\mathbf{H}} \mathbf{X} + M \sigma_n^2 \mathbf{I}_T) \mathbf{A} ] , \end{aligned} \quad (1.30)$$

where the channel correlation matrix is defined as

$$\mathbf{R}_{\mathbf{H}} = E [ \mathbf{H}^H \mathbf{H} ] = \left( \sum_{i=1}^M \sigma_{H,i}^2 \right) \mathbf{I}_N \triangleq M \cdot \sigma_H^2 \mathbf{I}_N . \quad (1.31)$$

Note the second equality follows from the independence of columns of  $\mathbf{H}$ , but the diagonal entries of  $\mathbf{\Sigma}_{\mathbf{h}}$  might not be identical; see (1.17). To find  $\mathbf{A}_o$ , set the gradients to zero [30,

Tab. V],

$$\frac{\partial J(\mathbf{A})}{\partial \mathbf{A}} = -(\mathbf{R}_H \mathbf{X})^T + (\mathbf{X}^H \mathbf{R}_H \mathbf{X} + M\sigma_n^2 \mathbf{I}_T)^T \mathbf{A}^* = \mathbf{0}_{N \times T}, \quad (1.32)$$

which leads to [16, eq. 25]

$$\mathbf{A}_o = (\mathbf{X}^H \mathbf{R}_H \mathbf{X} + M\sigma_n^2 \mathbf{I}_T)^{-1} \mathbf{X}^H \mathbf{R}_H. \quad (1.33)$$

Then the average MSE of this MMSE estimator is

$$\begin{aligned} J_{MMSE} &\triangleq \frac{1}{MN} J(\mathbf{A}_o) \\ &= \frac{1}{MN} \text{Tr} \left[ \mathbf{R}_H - \mathbf{R}_H \mathbf{X} (\mathbf{X}^H \mathbf{R}_H \mathbf{X} + M\sigma_n^2 \mathbf{I}_T)^{-1} \mathbf{X}^H \mathbf{R}_H \right] \\ &= \frac{1}{MN} \text{Tr} \left[ \left( \mathbf{R}_H^{-1} + \frac{1}{M\sigma_n^2} \mathbf{X} \mathbf{X}^H \right)^{-1} \right] = \frac{N\sigma_n^2 \cdot \sigma_H^2}{N\sigma_n^2 + PT\sigma_H^2}, \end{aligned} \quad (1.34)$$

where the third step follows from Woodbury matrix identity, and with (1.25) and (1.31), we find the MMSE estimator has zero-mean and i.i.d. estimation errors [16, eq. 27],

$$\left( \mathbf{R}_H^{-1} + \frac{1}{M\sigma_n^2} \mathbf{X} \mathbf{X}^H \right)^{-1} = \left( \frac{1}{M\sigma_H^2} \mathbf{I}_N + \frac{1}{M} \frac{PT}{N\sigma_n^2} \mathbf{I}_N \right)^{-1} = \frac{\sigma_n^2 \cdot MN\sigma_H^2}{N\sigma_n^2 + PT\sigma_H^2} \mathbf{I}_N. \quad (1.35)$$

This error definitely depends on the receiver characteristics. It is straightforward to show the MMSE estimator has reduced MSE compared to the LS, i.e,

$$\frac{J_{LS,min}}{J_{MMSE}} = 1 + \frac{N}{T} \frac{\sigma_n^2}{P\sigma_H^2} > 1. \quad (1.36)$$

Furthermore, it has been implied that the MMSE estimate  $\hat{\mathbf{H}}_{MMSE}$  in (1.28) follows

a zero-mean, circularly-symmetric, complex Gaussian distribution [37, eq. 20],

$$\text{vec } \hat{\mathbf{H}}_{MMSE} \sim \mathcal{CN} \left( \mathbf{0}_{MN}, \frac{\sigma_{\hat{H}}^2}{\sigma_H^2} \mathbf{I}_N \otimes \boldsymbol{\Sigma}_{\mathbf{h}} \right), \quad (1.37)$$

where the estimation error is uncorrelated to the estimates [37],

$$\sigma_{\hat{H}}^2 \triangleq \sigma_H^2 - J_{MMSE} = \frac{PT\sigma_H^2}{N\sigma_n^2 + PT\sigma_H^2} \sigma_H^2. \quad (1.38)$$

Now we combine the three pillars of technicality reviewed above to form the unified metric for this dissertation, i.e., channel capacity with impedance mismatch and channel estimation errors.

Consider a data transmission phase, where channel information has been estimated using MMSE. The MMSE estimate of the channel  $\hat{\mathbf{H}}$  is treated as correct, such that

$$\mathbf{y}_d = \hat{\mathbf{H}}\mathbf{x}_d + \mathbf{H}_e\mathbf{x}_d + \mathbf{n}_d \triangleq \hat{\mathbf{H}}\mathbf{x}_d + \mathbf{n}_e, \quad (1.39)$$

where  $\mathbf{H}_e = \mathbf{H} - \hat{\mathbf{H}}$  is the channel estimation error, the noise is ZMCSCG, i.e.,  $\mathbf{n}_d \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_M)$ . The equivalent noise (and error) has power

$$\begin{aligned} \sigma_e^2 &= \frac{1}{M} E \text{Tr} [\mathbf{n}_e \mathbf{n}_e^H] = \frac{1}{M} E [\mathbf{n}_d^H \mathbf{n}_d] + \frac{1}{M} E \text{Tr} [\mathbf{H}_e \mathbf{x}_d \mathbf{x}_d^H \mathbf{H}_e^H] \\ &= \sigma_n^2 + \frac{1}{M} \text{Tr} (E [\mathbf{x}_d \mathbf{x}_d^H] E [\mathbf{H}_e^H \mathbf{H}_e]) = \sigma_n^2 + \frac{1}{M} \text{Tr} \left[ \frac{P}{N} \mathbf{I}_N \cdot M J_{MMSE} \mathbf{I}_N \right] \\ &= \sigma_n^2 + P \cdot J_{MMSE}. \end{aligned} \quad (1.40)$$

Hassibi and Hochwald derived a lower bound on ergodic capacity when channel estimation errors exists [37, Th. 1]. In particular, if the MMSE estimate in (1.28) is treated as correct during data transmission for our setting, then the ergodic capacity is lower

bounded by [37, eq. 21]

$$C \geq C_l = E \left[ \log_2 \det \left( \mathbf{I}_M + \rho_{\text{eff}} \cdot \frac{1}{N} \boldsymbol{\Sigma}_{\mathbf{h}} \mathbf{H}_w \mathbf{H}_w^H \right) \right], \quad (1.41)$$

where the distribution of  $\hat{\mathbf{H}}$  is given in (1.37), and the effective SNR is defined as,

$$\rho_{\text{eff}} \triangleq \frac{P\sigma_{\hat{\mathbf{H}}}^2}{\sigma_n^2 + P \cdot J_{\text{MMSE}}} = \frac{P\sigma_H^2}{\sigma_n^2} \frac{PT\sigma_H^2}{PT\sigma_H^2 + N(P\sigma_H^2 + \sigma_n^2)}. \quad (1.42)$$

Comparing this capacity lower bound to that in (1.8) with perfect channel information at the receiver, we find, if defining  $\rho \triangleq P\sigma_H^2/\sigma_n^2$  as the SNR,

$$\rho_{\text{eff}} = \eta \cdot \rho, \quad \eta \triangleq \frac{1}{1 + (1 + 1/\rho)N/T}. \quad (1.43)$$

where  $\eta$  is defined as the efficiency.

Observe that a larger training symbol to transmit antenna ratio, i.e.,  $T/N$ , leads to a higher efficiency and higher ergodic capacity. Recall that the signal correlated  $\boldsymbol{\Sigma}_{\mathbf{h}}(\mathbf{Z}_{\mathbf{A}})$  is a function of antenna impedance and becomes a scaled identity matrix under optimal matching (1.21).

### 1.3 Outline of this Dissertation

This dissertation considers joint channel and antenna impedance estimation in fading, multipath channels, as well as the application of these estimates to capacity-optimal adaptive matching. The rest of this dissertation is organized as follows.

In Chapter 2, we first develop a hybrid estimation framework for single-input, single-output (SISO) channels. In this framework, the channel gain is modeled as a Gaussian

random parameter, while the antenna impedance is deterministic. We observe these parameters by synchronously switching the receiver load during the observation of known training sequences. Based on these observations, we derive the joint maximum *a posteriori* and maximum-likelihood (MAP/ML) estimators for the channel and impedance. We investigate the bias, consistency, and efficiency of these estimators analytically and through numerical experiments. Our results suggest that the hybrid approach generally leads to inconsistent estimators, which perform poorly at low signal-to-noise ratios.

In Chapter 3, we consider multiple-input, single-output (MISO) channels, where multiple antennas are deployed at the transmitter but not the receiver. Based on the results of the previous chapter, we consider a different estimation methodology, in which channel estimation is decoupled from impedance estimation. In particular, we first derive the maximum-likelihood (ML) estimator for the antenna impedance over multiple training packets, as well as the Cramer-Rao lower bound (CRB) on the performance of any unbiased estimator. We then consider minimum-mean square estimation of the channel, conditioned on knowledge of the antenna impedance. Numerical results suggest that the performance of these estimators are consistent and perform better than the hybrid approach at low signal-to-noise ratios. However, they are sensitive to the spatial and temporal diversity available in the fading channel. When diversity is limited, these estimators tend to be inefficient and the mean-squared error can be infinite. In the presence of rich diversity, however, both estimators perform close to the CRB under a broad range of conditions of practical interest.

In Chapter 4, we consider a general multiple-input, multiple-out (MIMO) channel, where multiple antennas are used at both the transmitter and receiver, and where channel fading path gains may be both spatially and temporally correlated at the receiver. Extending the approach of Chapter 3, we derive the ML estimator for the antenna impedance

matrix, as well as a simpler estimators for correlated fading based on the iterative Newton-Raphson and the method-of-moments (MM). In addition, we also derive the Cramér-Rao bounds (CRB) for these estimation problems, and compare the performance of these estimators to this fundamental lower bound. Due to the greater diversity available in MIMO channel, our numerical results suggest that these estimators perform close to the CRBs for most estimation conditions of practical interest.

Lastly, we conclude this dissertation by summarizing our contributions and possible future research directions.

# Chapter 2

## A Hybrid Approach to Joint Channel and Antenna Impedance Estimation

### 2.1 Introduction

We now develop a hybrid framework for joint channel and antenna impedance estimation in single-input, single-output (SISO) channels. In this approach, the channel path gain is modeled as a complex Gaussian random variable, while the antenna impedance is a deterministic parameter. We assume the receiver switches its load during reception of a known training sequence. Based on the received signal over multiple training packets, we derive the joint maximum *a posteriori* and maximum-likelihood (MAP/ML) estimators for the channel and impedance, respectively. Through numerical examples, we investigate the bias, consistency, and efficiency of these estimators, as well as the impact of channel correlation on estimation accuracy.

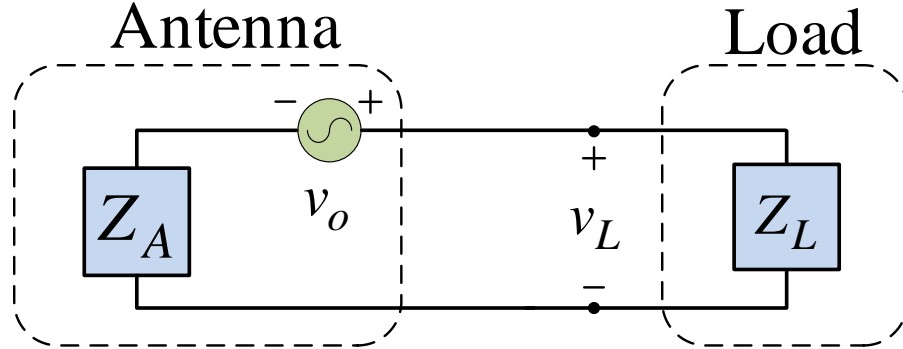


Figure 2.1: Circuit model of a single-antenna receiver

The rest of this chapter is organized as follows. We present the system model in Sec. 2.2, and derive the joint MAP/ML estimators for the channel and antenna impedance in Sec. 2.3. These estimators are then extended to multiple packets in Sec. 2.4. We study the properties of these estimators, such as bias and efficiency, in Sec. 2.5; illustrate their performance through numerical examples in Sec. 2.6; and summarize our conclusions in Sec. 2.7.

## 2.2 System Model

Consider a narrowband communications link with one transmit antenna and one receive antenna. We adopt the receiver model illustrated in Fig. 2.1, which has been widely used to model scenarios in which amplifier noise dominates at the receiver [29, 51, 79]. More complex models that consider multiple noise sources and impedance matching are considered in [26, 27].

In Fig. 2.1, the antenna is modeled by the Thevenin equivalent

$$v = Z_A i + v_o, \quad (2.1)$$

where  $v, i \in \mathbb{C}$  are the voltage across, and current into, the antenna terminals. Here

$$Z_A = R_A + jX_A, \quad (2.2)$$

is the antenna impedance, and  $R_A$  and  $X_A$  are the resistance and reactance, respectively. In (2.1),  $v_o \in \mathbb{C}$  is the open-circuit voltage induced by the incident signal field, which can be modeled in a flat-fading environment as [27]

$$v_o = Gx, \quad (2.3)$$

where  $x \in \mathbb{C}$  is the transmitted symbol and  $G \in \mathbb{C}$  is the fading path gain.

We assume the estimation algorithms observe a noisy version of the load voltage in Fig. 2.1. With a slight abuse of notation, we write this as [29, 51, 79]

$$v_L = \frac{Z_L G x}{Z_A + Z_L} + n, \quad (2.4)$$

where the observation noise  $n$  is a zero-mean, circularly-symmetric, complex Gaussian random variable with variance  $\sigma_n^2$ , which we hereafter denote by  $n \sim \mathcal{CN}(0, \sigma_n^2)$ .

Suppose the channel gain  $G$  and antenna impedance  $Z_A$  are unknown to the receiver. In many communication systems, the channel is often estimated in part from training sequences embedded in the transmitted data. In this chapter, we would like to jointly estimate both unknown parameters by combining training sequences with known variations in the receiver load impedance  $Z_L$ .

We assume the transmitter sends a predetermined training sequence,  $x_1, \dots, x_T$ , and the receiver shifts synchronously through a sequence of known impedances  $Z_{L,1}, \dots, Z_{L,T}$ . If  $G$  and  $Z_A$  are modeled as fixed over the duration of the training sequence, the received

observations are given by

$$v_{L,t} = \frac{Z_{L,t}Gx_t}{Z_A + Z_{L,t}} + n_t, \quad t = 1, \dots, T, \quad (2.5)$$

where  $n_t \sim \mathcal{CN}(0, \sigma_n^2)$  are independent and identically distributed (i.i.d.). In this chapter, we consider load impedances that take on only two possible values,

$$Z_{L,t} = \begin{cases} Z_1, & 1 \leq t \leq K, \\ Z_2, & K < t \leq T. \end{cases} \quad (2.6)$$

where  $Z_1$  and  $Z_2$  are known. Here we assume  $Z_L = Z_1$  is the load impedance used to receive the transmitted data, which is matched to our best estimate of  $Z_A$ , and  $Z_L = Z_2$  is an impedance variation introduced in order to make  $Z_A$  observable.

The estimation problem above can be simplified with a slightly different parameterization of the unknowns. If  $Z_1$  is the load impedance used to transmit the data, then the channel estimate needed by the communication algorithms would normally be defined as [80, eq. 13],

$$H \triangleq \frac{Z_1G}{Z_A + Z_1}. \quad (2.7)$$

With this definition, the sequence of observations can be described compactly as

$$v_{L,t} = \begin{cases} Hx_t + n_t, & 1 \leq t \leq K, \\ FHx_t + n_t, & K < t \leq T, \end{cases} \quad (2.8)$$

where

$$F \triangleq \frac{1 + Z_A/Z_1}{1 + Z_A/Z_2}. \quad (2.9)$$

Note  $F$  is a one-to-one mapping of  $Z_A$  provided  $Z_1 \neq Z_2$ , so knowing  $F$  is equivalent to knowing  $Z_A$ . Intuitively,  $F$  can be viewed as a ratio of the channel transmission coefficients associated with the loads  $Z_1$  and  $Z_2$ ; more precisely,  $F = T_2/T_1$  where  $T_i = 2Z_i/(Z_A + Z_i), i = 1, 2$ . Note that redefining the estimation problem in terms of  $H$  and  $F$  renders the observation bilinear in the unknowns. In the channel estimation literature, since  $G$  usually results from a random superposition of multipaths,  $H$  is usually modeled as a Gaussian random variable. On the other hand,  $F$  would appear to be more appropriately modeled as a fixed unknown constant. In practical applications, observing  $F$  may actually be more convenient than observing  $Z_A$  directly.

With these definitions, we can now precisely state the estimation problem considered in this chapter. We would like to estimate the complex-valued parameters

$$\boldsymbol{\theta} \triangleq [H, F]^T \tag{2.10}$$

based on the observations (2.8). Here,  $H$  is a zero-mean Gaussian random variable,  $H \sim \mathcal{CN}(0, \sigma_H^2)$  and  $F$  is deterministic. Note the prior probability density function (pdf) of  $H$ ,

$$p(H) = \frac{1}{\pi\sigma_H^2} \exp\left(-\frac{|H|^2}{\sigma_H^2}\right), \tag{2.11}$$

does not depend on  $F$ . For simplicity, we initially assume  $\sigma_H^2$  is known to the receiver. Although this does not appear consistent with the assumption that  $Z_A$  is unknown, it can be justified by observing  $\sigma_H^2$  can be accurately estimated from the data received using the receiver impedance  $Z_1$ . To illustrate, we will relax this assumption later in the next chapter.

## 2.3 Joint MAP and ML Estimators

In classical estimation theory, unknown parameters are modeled as deterministic; in Bayesian estimation, they are modeled as random variables. The problem introduced in Sec. 2.2 is a *hybrid* estimation problem [77, pg. 329], because  $\boldsymbol{\theta}$  contains both random and deterministic parameters. Several authors have investigated hybrid estimation problems [12, 60, 67–69]. Estimators and Cramér-Rao-type bounds for hybrid estimation were formulated in [69] in the context of passive source localization. In this section, we apply similar tools to the problem formulated in Sec. 2.2.

Before discussing estimators, it is convenient to introduce sufficient statistics that summarize the information in the observations  $\mathbf{v}_L \triangleq [v_{L,1}, v_{L,2}, \dots, v_{L,T}]^T$  relevant to estimation of  $\boldsymbol{\theta}$ . In classical estimation, a statistic  $\mathbf{V} = f(\mathbf{v}_L)$  is sufficient to estimate  $\boldsymbol{\theta}$  based on  $\mathbf{v}_L$  if  $p(\mathbf{v}_L|\mathbf{V}, \boldsymbol{\theta}) = p(\mathbf{v}_L|\mathbf{V})$ . In Bayesian settings, a statistic is sufficient if  $p(\mathbf{v}_L, \boldsymbol{\theta}|\mathbf{V}) = p(\mathbf{v}_L|\mathbf{V})p(\boldsymbol{\theta}|\mathbf{V})$ . It is well known that classical sufficiency implies Bayesian sufficiency; it also clearly implies sufficiency for the hybrid estimation problem, which is defined in an analogous way.

**Lemma 1 (Sufficient Statistic) :** *For  $\boldsymbol{\theta}$  fixed and unknown, the statistic  $\mathbf{V} = [V_1, V_2]^T$ , where*

$$V_1 \triangleq \frac{1}{S_1} \sum_{t=1}^K v_{L,t} x_t^*, \quad V_2 \triangleq \frac{1}{S_2} \sum_{t=K+1}^T v_{L,t} x_t^*, \quad (2.12)$$

and

$$S_1 \triangleq \sum_{t=1}^K |x_t|^2, \quad S_2 \triangleq \sum_{t=K+1}^T |x_t|^2, \quad (2.13)$$

is sufficient (in the classical sense) to estimate  $\boldsymbol{\theta}$  based on the observations (2.8).  $\diamond$

**Proof :** From the Neyman-Fisher Factorization Theorem [45, pg. 117], it suffices to show the conditional pdf of the observations can be factored in the form  $p(\mathbf{v}_L|\boldsymbol{\theta}) = g(\mathbf{V}, \boldsymbol{\theta})h(\mathbf{v}_L)$ . From (2.8), the pdf of the observations can be written as

$$\begin{aligned}
(\pi\sigma_n^2)^T \cdot p(\mathbf{v}_L|\boldsymbol{\theta}) &= \exp\left(-\frac{1}{\sigma_n^2} \sum_{t=1}^K |v_{L,t} - Hx_t|^2 - \frac{1}{\sigma_n^2} \sum_{t=K+1}^T |v_{L,t} - FHx_t|^2\right) \\
&= \exp\left[-\frac{1}{\sigma_n^2} \sum_{t=1}^K (|Hx_t|^2 - v_{L,t}^* x_t H - v_{L,t} x_t^* H^*)\right] \times \\
&\quad \exp\left[-\frac{1}{\sigma_n^2} \sum_{t=K+1}^T (|FHx_t|^2 - v_{L,t}^* x_t FH - v_{L,t} x_t^* F^* H^*)\right] \exp\left(-\frac{\sum_{t=1}^T |v_{L,t}|^2}{\sigma_n^2}\right) \\
&= \exp\left[-\frac{S_1}{\sigma_n^2} (|H|^2 - V_1^* H - V_1 H^*) - \frac{S_2}{\sigma_n^2} (|FH|^2 - V_2^* FH - V_2 F^* H^*)\right] \exp\left(-\frac{|\mathbf{v}_L|^2}{\sigma_n^2}\right).
\end{aligned}$$

To complete the proof, equate the first exponential with  $(\pi\sigma_n^2)^T g(\mathbf{V}, \boldsymbol{\theta})$  and the second with  $h(\mathbf{v}_L)$ .

For simplicity, we will hereafter consider the observations to be the sufficient statistics  $V_1$  and  $V_2$ . From (2.8), the conditional pdf of these observations is

$$p(\mathbf{v}|H; F) = \frac{1}{\pi^2 \det(\mathbf{C}_\mathbf{V})} \exp\left[-(\mathbf{v} - \boldsymbol{\mu}_\mathbf{V})^H \mathbf{C}_\mathbf{V}^{-1} (\mathbf{v} - \boldsymbol{\mu}_\mathbf{V})\right], \quad (2.14)$$

where

$$\mathbf{v} = [v_1, v_2]^T, \quad \boldsymbol{\mu}_\mathbf{V} = [H, FH]^T, \quad \mathbf{C}_\mathbf{V} = \begin{bmatrix} \frac{\sigma_n^2}{S_1} & 0 \\ 0 & \frac{\sigma_n^2}{S_2} \end{bmatrix}. \quad (2.15)$$

We now use these observations to estimate  $H$  and  $F$  under the hybrid assumptions  $H$  is a random variable with pdf (2.11), and  $F$  is an unknown constant. For these assumptions, it is natural to consider an estimator that maximizes the hybrid log-likelihood

function [77, pg. 329], [69],

$$\hat{\boldsymbol{\theta}}(\mathbf{V}) \triangleq \arg \max_{\boldsymbol{\theta} \in \mathbb{C}^2} \ln p(\mathbf{V}, H; F) . \quad (2.16)$$

This results in a joint maximum *a posteriori* probability estimate of  $H$  and maximum-likelihood estimate of  $F$ , which are presented in the following theorem.

**Theorem 1 (Joint MAP and ML Estimators)** : *Given the sufficient statistics  $V_1$  and  $V_2$  in (2.12), the joint maximum a posteriori probability (MAP) and maximum-likelihood (ML) estimators for  $H$  and  $F$ , respectively, are*

$$\hat{\boldsymbol{\theta}} \triangleq \left[ \hat{H}_{MAP}, \hat{F}_{ML} \right]^T = \left[ cV_1, \frac{V_2}{cV_1} \right]^T , \quad \text{where } c \triangleq \frac{S_1\sigma_H^2}{S_1\sigma_H^2 + \sigma_n^2} , \quad (2.17)$$

and  $\sigma_H^2$  and  $\sigma_n^2$  are the variances of the channel path gain and noise, respectively.  $\diamond$

**Proof** : From (2.16), (2.14) and (2.11), we can write

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \arg \max_{\boldsymbol{\theta}} [\ln p(\mathbf{V}|H; F) + \ln p(H)] \\ &= \arg \max_{\boldsymbol{\theta}} \left[ -(\mathbf{V} - \boldsymbol{\mu}_{\mathbf{V}})^H \mathbf{C}_{\mathbf{V}}^{-1} (\mathbf{V} - \boldsymbol{\mu}_{\mathbf{V}}) - \frac{|H|^2}{\sigma_H^2} \right] \\ &= \arg \max_{\boldsymbol{\theta}} \left[ -\frac{S_1}{\sigma_n^2} |V_1 - H|^2 - \frac{S_2}{\sigma_n^2} |V_2 - FH|^2 - \frac{1}{\sigma_H^2} |H|^2 \right] \\ &\triangleq \arg \max_{\boldsymbol{\theta}} \mathcal{L}(H, F) , \end{aligned} \quad (2.18)$$

where  $S_1$  and  $S_2$  are defined in (2.13). If we denote the real and imaginary parts of the parameter vector by  $\boldsymbol{\theta} = \boldsymbol{\theta}_r + j\boldsymbol{\theta}_i$ , the complex gradient is given by [45, Sec. 15.6]

$$\frac{\partial \mathcal{L}(H, F)}{\partial \boldsymbol{\theta}^*} \triangleq \frac{1}{2} \left[ \frac{\partial \mathcal{L}(H, F)}{\partial \boldsymbol{\theta}_r} + j \frac{\partial \mathcal{L}(H, F)}{\partial \boldsymbol{\theta}_i} \right] . \quad (2.19)$$

With this definition, in order for  $H, F$  to maximize  $\mathcal{L}(H, F)$  it is necessary that

$$\begin{aligned}\frac{\partial \mathcal{L}(H, F)}{\partial H^*} &= \frac{S_1}{\sigma_n^2}(V_1 - H) + \frac{S_2}{\sigma_n^2}(V_2 - FH)F^* - \frac{H}{\sigma_H^2} = 0, \\ \frac{\partial \mathcal{L}(H, F)}{\partial F^*} &= \frac{S_2}{\sigma_n^2}(V_2 - FH)H^* = 0.\end{aligned}\tag{2.20}$$

The second equation implies  $V_2 - FH = 0$  or  $H = 0$ . The only solution satisfying the first condition is  $H_1 = cV_1, F_1 = V_2/(cV_1)$ . The only solution that satisfies the second condition is  $H_2 = 0$  and  $F_2 = -V_1^*/(\alpha V_2^*)$ . Substituting each back into the log-likelihood, we have

$$\begin{aligned}\mathcal{L}(H_1, F_1) &= -\frac{S_1(1-c)^2}{\sigma_n^2}|V_1|^2 - \frac{c^2|V_1|^2}{\sigma_H^2} = -\frac{S_1}{\sigma_n^2}|V_1|^2(1-c) \\ \mathcal{L}(H_2, F_2) &= -\frac{S_1}{\sigma_n^2}|V_1|^2 - \frac{S_2}{\sigma_n^2}|V_2|^2,\end{aligned}$$

where the second equality follows from  $S_1\sigma_H^2/\sigma_n^2 = c/(1-c)$ . Since  $\mathcal{L}(H_1, F_1) \geq \mathcal{L}(H_2, F_2)$  for all  $V_1, V_2, 0 < c < 1$ , we conclude that  $H_1, F_1$  is the global maximum of  $\mathcal{L}(H, F)$ .

The hybrid ML estimator for  $Z_A$  can be found by substituting  $\hat{F}_{ML}$  in (2.9) based on the invariance principle of MLE [45, pg. 176]. However, it is not difficult to show that this ML estimator is biased. In the next section, we investigate optimal estimators when multiple packets are available for joint processing.

## 2.4 Estimators for Multiple Packets

We now consider estimators that combine observations from multiple training packets. Consider a sequence of  $L$  packets, where  $H_i$  denotes the channel during the  $i$ -th packet transmission. We assume the channel path gains are unknown, jointly-distributed Gaus-

sian RVs, such that  $\mathbf{H} = (H_1, \dots, H_L)^T \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_\mathbf{H})$ . For simplicity, we assume  $\mathbf{C}_\mathbf{H}$  is known at the receiver, although this assumption will be relaxed later. We assume  $F$  changes more slowly with time, and can be regarded as fixed over the  $L$  packets. If each packet is formatted as in (2.8), then the observations during the  $i$ -th packet can be written as

$$v_{L,t,i} = \begin{cases} H_i x_t + n_{1,t,i} , & 1 \leq t \leq K , \\ FH_i x_t + n_{2,t,i} , & K < t \leq T , \end{cases} \quad (2.21)$$

where  $n_{1,t,i} \sim \mathcal{CN}(0, \sigma_n^2/S_1)$  and  $n_{2,t,i} \sim \mathcal{CN}(0, \sigma_n^2/S_2)$  are mutually independent for all  $i$  and  $t$ . The aim of this section is to use these observations to estimate the parameters  $\boldsymbol{\theta} \triangleq [\mathbf{H}^T, F]^T$ .

### 2.4.1 Estimators for General $\mathbf{C}_\mathbf{H}$

Let  $V_{1i}, V_{2i}$  denote statistics for the  $i$ -th packet defined analogously to (2.12). Proceeding as in Lemma 1, we can show  $\mathbf{V}_1 = (V_{11}, \dots, V_{1L})^T$  and  $\mathbf{V}_2 = (V_{21}, \dots, V_{2L})^T$  are sufficient statistics to estimate  $\mathbf{H}$  and  $F$  based on (2.21). We can therefore consider the observations to be

$$\begin{aligned} \mathbf{V}_1 &= \mathbf{H} + \mathbf{N}_1 , \\ \mathbf{V}_2 &= F\mathbf{H} + \mathbf{N}_2 , \end{aligned} \quad (2.22)$$

where  $\mathbf{N}_1 \sim \mathcal{CN}(\mathbf{0}, (\sigma_n^2/S_1)\mathbf{I})$  and  $\mathbf{N}_2 \sim \mathcal{CN}(\mathbf{0}, (\sigma_n^2/S_2)\mathbf{I})$  are independent noise vectors.

As in the last section, we estimate  $\mathbf{H}$  and  $F$  under the hybrid assumptions that  $F$  is

an unknown constant and  $\mathbf{H}$  is a random vector with pdf

$$p(\mathbf{H}) = \frac{1}{\det(\pi \mathbf{C}_{\mathbf{H}})} \exp(-\mathbf{H}^H \mathbf{C}_{\mathbf{H}}^{-1} \mathbf{H}) . \quad (2.23)$$

Again, we consider estimators that maximize the hybrid log-likelihood function

$$\hat{\boldsymbol{\theta}}(\mathbf{V}_1, \mathbf{V}_2) \triangleq \arg \max_{\boldsymbol{\theta} \in \mathbb{C}^{L+1}} \ln p(\mathbf{V}_1, \mathbf{V}_2, \mathbf{H}; F) , \quad (2.24)$$

which are presented in the following theorem.

**Theorem 2 (Multi-packet Joint MAP/ML Estimators) :** *Given the sufficient statistics  $\mathbf{V}_1$  and  $\mathbf{V}_2$  in (2.22), the joint MAP/ML estimators for  $\mathbf{H}$  and  $F$  are, respectively,*

$$\hat{\mathbf{H}}_{MAP} \triangleq \left[ \left( \frac{S_1}{\sigma_n^2} + \frac{|\hat{F}_{ML}|^2 S_2}{\sigma_n^2} \right) \mathbf{I} + \mathbf{C}_{\mathbf{H}}^{-1} \right]^{-1} \left( \frac{S_1}{\sigma_n^2} \mathbf{V}_1 + \frac{S_2}{\sigma_n^2} \hat{F}_{ML}^* \mathbf{V}_2 \right) \quad (2.25)$$

and  $\hat{F}_{ML}$ , where  $\hat{F}_{ML}$  is a zero of the rational function

$$g(F) \triangleq \left( \frac{S_1}{\sigma_n^2} \mathbf{V}_1 + \frac{S_2}{\sigma_n^2} F^* \mathbf{V}_2 \right)^H \mathbf{A}^H(F) \mathbf{A}(F) \left( \frac{S_1}{\sigma_n^2} (\mathbf{V}_2 - F \mathbf{V}_1) + \mathbf{C}_{\mathbf{H}}^{-1} \mathbf{V}_2 \right), \quad (2.26)$$

and

$$\mathbf{A}(F) \triangleq \left[ \left( \frac{S_1}{\sigma_n^2} + \frac{|F|^2 S_2}{\sigma_n^2} \right) \mathbf{I} + \mathbf{C}_{\mathbf{H}}^{-1} \right]^{-1} . \quad (2.27)$$

**Proof :** From (2.22) and (2.23), we can write

$$\begin{aligned} \mathcal{L}(\mathbf{H}, F) &\triangleq \ln p(\mathbf{V}_1, \mathbf{V}_2, \mathbf{H}; F) \\ &= -\frac{S_1}{\sigma_n^2} |\mathbf{V}_1 - \mathbf{H}|^2 - \frac{S_2}{\sigma_n^2} |\mathbf{V}_2 - F \mathbf{H}|^2 - \mathbf{H}^H \mathbf{C}_{\mathbf{H}}^{-1} \mathbf{H} + K \end{aligned} \quad (2.28)$$

where  $K$  is a constant. In order for  $\mathbf{H}, F$  to maximize  $\mathcal{L}(\mathbf{H}, F)$  it is necessary that

$$\begin{aligned}\frac{\partial \mathcal{L}(\mathbf{H}, F)}{\partial \mathbf{H}^*} &= \frac{S_1}{\sigma_n^2}(\mathbf{V}_1 - \mathbf{H}) + \frac{S_2}{\sigma_n^2}(\mathbf{V}_2 - F\mathbf{H})F^* - \mathbf{C}_{\mathbf{H}}^{-1}\mathbf{H} = \mathbf{0} , \\ \frac{\partial \mathcal{L}(\mathbf{H}, F)}{\partial F^*} &= \frac{S_2}{\sigma_n^2}\mathbf{H}^H(\mathbf{V}_2 - F\mathbf{H}) = 0 .\end{aligned}\tag{2.29}$$

Solving the first equation for  $\mathbf{H}$ , we obtain (2.25). To find  $F$ , we substitute (2.25) into (2.29), which yields (2.26), thereby completing the proof.

Since  $g(F)$  is a ratio of polynomials, it may have multiple zeros, so the theorem does not necessarily specify unique estimators. Given multiple zeros  $F_1, \dots, F_m$ , however, we can easily identify the MAP/ML solution: For each  $F_i$ , we can, in principle, calculate a corresponding estimate of  $\mathbf{H}_i$  from (2.25). The MAP/ML solution will be the pair  $\mathbf{H}_i, F_i$  that maximizes  $\mathcal{L}(\mathbf{H}_i, F_i)$ . We conjecture the ML zero is always the one that maximizes  $\text{Re}[\mathbf{V}_2^H \mathbf{V}_1 F_i]$ .

In principle, Theorem 2 requires us to extract and compare all of the zeros of  $g(F)$ . However, the form of equations (2.25) and (2.29) suggests a simpler, alternative approach to approximate the MAP/ML estimators. Note the second equation in (2.29) can be rewritten as

$$F = \mathbf{H}^H \mathbf{V}_2 / \mathbf{H}^H \mathbf{H} .\tag{2.30}$$

Thus, we could attempt to solve the equations iteratively by setting  $F_0 = 0$  and solving for  $\mathbf{H}_0$  using (2.25). This is equivalent to estimating  $\mathbf{H}$  using only  $\mathbf{V}_1$ . Using  $\mathbf{H}_0$ , we can then form improved estimate of  $F$ , say  $F_1$ , from (2.30). We can then iterate between these two equations to improve the estimates.

### 2.4.2 Special Cases of $\sigma_n^{-2}\mathbf{C}_\mathbf{H}$

We now consider three extreme cases of  $\sigma_n^{-2}\mathbf{C}_\mathbf{H}$  in which Theorem 2 yields explicit, closed-form estimators for  $\mathbf{H}$  and  $F$ . First suppose  $H_1, \dots, H_L$  is an i.i.d. sequence, so  $\mathbf{C}_\mathbf{H} = \sigma_H^2 \mathbf{I}$ . In this case,  $g(F) = 0$  reduces to

$$(\mathbf{V}_1 + \alpha F^* \mathbf{V}_2)^H (\mathbf{V}_2 - cF \mathbf{V}_1) = 0, \alpha = \frac{S_2}{S_1}, c = \frac{S_1 \sigma_H^2}{S_1 \sigma_H^2 + \sigma_n^2}. \quad (2.31)$$

Expanding the product and defining  $P_{ij} \triangleq (1/L) \mathbf{V}_i^H \mathbf{V}_j$ , we obtain

$$P_{12} + (\alpha P_{22} - cP_{11})F - \alpha c P_{21} F^2 = 0. \quad (2.32)$$

If we multiply both sides by  $-P_{21} = -P_{12}^*$  and define  $E \triangleq P_{21}F$ , this equation becomes

$$\alpha c E^2 - (\alpha P_{22} - cP_{11})E - |P_{12}|^2 = 0. \quad (2.33)$$

This is a quadratic equation with real coefficients, which has two real roots:

$$E = \frac{\alpha P_{22} - cP_{11} \pm \sqrt{(\alpha P_{22} - P_{11})^2 + 4\alpha c |P_{12}|^2}}{2\alpha c}. \quad (2.34)$$

It follows the ML estimate of  $F$  is given by one of the following two roots,

$$F = \frac{\alpha P_{22} - cP_{11} \pm \sqrt{(\alpha P_{22} - cP_{11})^2 + 4\alpha c |P_{12}|^2}}{2\alpha P_{21}}. \quad (2.35)$$

We conjecture the positive root above is always the ML estimate; however, as noted earlier, we can determine which root is the actual solution by calculating the MAP  $\mathbf{H}$  for each possible  $F$ , and then choosing the pair  $\mathbf{H}, F$  that gives the larger value of  $\mathcal{L}(\mathbf{H}, F)$ .

The second extreme case we consider is an arbitrary non-singular  $\mathbf{C}_{\mathbf{H}}$  in the low noise limit,  $\sigma_n^2 \rightarrow 0$ . In (2.26), applying the approximation  $\sigma_n^2 \mathbf{C}_{\mathbf{H}}^{-1} \approx \mathbf{0}$  to the equation  $g(F) = 0$  leads to

$$(\mathbf{V}_1 + \alpha F^* \mathbf{V}_2)^H (\mathbf{V}_2 - F \mathbf{V}_1) = 0. \quad (2.36)$$

Comparing this to (2.31), we see this equation is identical to the case of i.i.d.  $H_i$  for  $c = 1$ . It follows immediately that the ML estimate of  $F$  is given by one of the following two roots,

$$F = \frac{\alpha P_{22} - P_{11} \pm \sqrt{(\alpha P_{22} - P_{11})^2 + 4\alpha |P_{12}|^2}}{2\alpha P_{21}}. \quad (2.37)$$

This result suggests that the ML estimator for arbitrary non-singular  $\mathbf{C}_{\mathbf{H}}$  is asymptotically the same as the i.i.d. ML estimator in (2.35) in the low-noise limit,  $\sigma_n^2 \rightarrow 0$ .

Finally, the last extreme case we consider is  $H_1 = H_2 = \dots = H_L$ , so  $[\mathbf{C}_{\mathbf{H}}]_{ij} = \sigma_H^2$  for all  $i, j$ . Strictly speaking, Theorem 2 does not apply here since it implicitly assumes  $\mathbf{C}_{\mathbf{H}}$  is invertible. However, the MAP/ML estimators follow directly from Theorem 1, by regarding all  $L$  packets as one large packet of size  $LT$ . In this case, Theorem 1 yields the estimators

$$\hat{H}_{MAP} = c \sum_{i=1}^L V_{1,i}, \quad \hat{F}_{ML} = \frac{1}{c} \frac{\sum_{i=1}^L V_{2,i}}{\sum_{i=1}^L V_{1,i}}, \quad c \triangleq \frac{S_1 \sigma_H^2}{S_1 \sigma_H^2 + \sigma_n^2}. \quad (2.38)$$

## 2.5 Properties of the Joint MAP/ML Estimators

The performance of estimators is often measured by low-order central moments, such as bias and mean-squared error. In this section, we explore the behavior of these moments for the joint MAP/ML estimators in Theorems 1 and 2, as well as the consistency of

these estimators.

### 2.5.1 Single-Packet Estimators

We first consider the single training packet estimators in Theorem 1. Note  $\hat{F}_{ML}$  in (2.17) is a ratio of complex Gaussian random variables, and so has the complex Gaussian ratio distribution [6, 14, 42, 58, 59, 80]. Ratios of Gaussian variables often have heavy tails, in the sense that no finite moments exist [6, 42, 49, 80]. For example, if  $U$  and  $V$  are independent, real, standard Gaussian random variables, the ratio  $U/V$  is a standard Cauchy random variable, which has no finite moments of order greater than or equal to one; only fractional moments exist.

Complex Gaussian ratios are somewhat better behaved: Hyder *et al* [42] showed that a ratio of zero-mean complex Gaussian RVs generally has a finite mean but an infinite variance. Astély *et al* [6] derived a similar result for a ratio of independent complex Gaussian RVs, and proved its mean-square is unbounded. In [80, Th. 2], we show the moments  $E[(V_2/V_1)^k|H]$  do not exist for any  $H$  for  $k > 1$ . Proceeding in a similar way, it is easy to show that  $\hat{F}_{ML}$  in (2.17) has infinite variance. To study the bias, we would like to calculate the mean of a ratio of complex Gaussian RVs with non-zero means and arbitrary correlation. The following result, which is proved in the Appendix, provides closed-form expression for the mean of a general complex Gaussian ratio.

**Theorem 3 :** *Suppose  $\mathbf{X} = [X_1, X_2]^T$  is a complex Gaussian random vector with mean and covariance*

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho^*\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

If  $\mu_2 \neq 0$  and  $\sigma_2 > 0$ , then

$$E \left[ \frac{X_1}{X_2} \right] = \frac{\mu_1}{\mu_2} + \left( \frac{\rho\sigma_1}{\sigma_2} - \frac{\mu_1}{\mu_2} \right) e^{-\gamma_2} , \quad (2.39)$$

where  $\gamma_2 \triangleq |\mu_2|^2/\sigma_2^2$ . In addition, the first absolute moment,  $E|X_1/X_2|$  is finite.  $\diamond$

Using this result, we can show the ML estimator in (2.17) is unbiased.

**Theorem 4 :** *The ML estimator in (2.17) satisfies  $E[\hat{F}_{ML}] = F$  for all  $F \in \mathbb{C}$ .*  $\diamond$

**Proof :** Applying Theorem 3 to the conditional pdf (2.14), we obtain for any  $F$

$$\begin{aligned} E[\hat{F}_{ML}] &= E \left[ \frac{V_2}{cV_1} \right] = E_H \left[ E \left\{ \frac{V_2}{cV_1} \middle| H \right\} \right] \\ &= \frac{F}{c} E_H \left[ \left( 1 - e^{-\frac{S_1|H|^2}{\sigma_n^2}} \right) \right] = F , \end{aligned} \quad (2.40)$$

where the last step follows from

$$\begin{aligned} E_H \left[ \exp \left( -\frac{S_1|H|^2}{\sigma_n^2} \right) \right] &= \int_{\mathbb{C}} \frac{1}{\pi\sigma_H^2} \exp \left( -\frac{S_1|H|^2}{\sigma_n^2} \right) \exp \left( -\frac{|H|^2}{\sigma_H^2} \right) dH \\ &= \frac{\sigma^2}{\sigma_H^2} \int_{\mathbb{C}} \frac{1}{\pi\sigma^2} \exp \left( -\frac{|H|^2}{\sigma^2} \right) dH = \frac{\sigma^2}{\sigma_H^2} = 1 - c , \end{aligned}$$

where  $\sigma^2 \triangleq (S_1/\sigma_n^2 + 1/\sigma_H^2)^{-1}$ .

While the single-packet estimator  $\hat{F}_{ML}$  is unbiased, it follows from the results above that its mean-squared error is infinite, and therefore cannot be used to measure performance. We therefore measure the performance of this estimator for  $L = 1$  by the mean absolute error (MAE), i.e.,  $E[|\hat{F}_{ML} - F|]$ .

Thus far, we have considered only the single-packet ( $L = 1$ ) estimator in Theorem 1. However, these results also imply unbiased (suboptimal) estimators for  $L > 1$ . First,

note that all of the results of this section also apply to the estimator (2.38) for  $L > 1$  under the conditions  $H_1 = H_2 = \dots = H_L$ , since this estimator is derived from Theorem 1. We further note the results of this section imply a simple, unbiased estimator for the general multi-packet problem

$$\hat{F}_U \triangleq \frac{1}{cL} \sum_{i=1}^L \frac{V_{2,i}}{V_{1,i}}, \quad (2.41)$$

whose performance will be explored via simulations in the next section.

## 2.5.2 Consistency

Except for the special case  $H_1 = H_2 = \dots = H_L$ , it appears difficult to evaluate the mean of  $\hat{F}_{ML}$  for  $L > 1$ , although numerical examples suggest it may be biased (more on this in the next section). To gain insight into why the estimator might be biased, it is helpful to consider the behavior of  $\hat{F}_{ML}$  for i.i.d.  $\mathbf{H}$  when  $L$  becomes large. As  $L \rightarrow \infty$ , the coefficients in (2.33) converge in probability,

$$\lim_{L \rightarrow \infty} P_{11} = \sigma_H^2 + \sigma_n^2/S_1, \quad \lim_{L \rightarrow \infty} P_{22} = |F|^2 \sigma_H^2 + \sigma_n^2/S_2, \quad \lim_{L \rightarrow \infty} P_{12} = \sigma_H^2 F,$$

so the roots (2.35) converge to

$$F = \frac{\alpha|F|^2 \sigma_H^2 - cd\sigma_H^2 \pm \sqrt{(\alpha|F|^2 \sigma_H^2 - cd\sigma_H^2)^2 + 4\alpha c\sigma_H^4 |F|^2}}{2\alpha c\sigma_H^2 F^*}$$

where  $d \triangleq 1 - (\sigma_n^2/S_1 \sigma_H^2)^2$ , so  $cd\sigma_H^2 = c(\sigma_H^2 + \sigma_n^2/S_1) - \alpha\sigma_n^2/S_2$ . Since neither root is  $F$ , it follows  $\hat{F}_{ML}$  is not consistent in  $L$ . However, if the second term in the square root is multiplied by  $d$ , the positive root becomes

$$\frac{\alpha|F|^2 \sigma_H^2 - cd\sigma_H^2 + \sqrt{(\alpha|F|^2 \sigma_H^2 - cd\sigma_H^2)^2 + 4\alpha dc\sigma_H^4 |F|^2}}{2\alpha c\sigma_H^2 F^*} = \frac{2\alpha|F|^2 \sigma_H^2}{2\alpha c\sigma_H^2 F^*} = \frac{F}{c}. \quad (2.42)$$

This suggests a modified estimator

$$\hat{F}_C \triangleq \frac{\alpha P_{22} - cP_{11} + \sqrt{(\alpha P_{22} - cP_{11})^2 + 4\alpha cd|P_{12}|^2}}{2\alpha P_{21}}, \quad (2.43)$$

which is consistent in  $L$ . For large signal-to-noise ratios ( $S_1\sigma_H^2 \gg \sigma_n^2$ ), we note  $c$  and  $d$  are approximately 1, and the positive root in (2.35) coincides with this consistent estimator. In the next section, we will compare the performance of  $\hat{F}_{ML}$  and  $\hat{F}_C$  via simulations.

### 2.5.3 Hybrid Cramér-Rao Bound

To evaluate estimator performance, it is useful to consider a fundamental lower bound on the error covariance

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} \triangleq E_{\mathbf{V}_1, \mathbf{V}_2, \mathbf{H}; F} \left[ \left( \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \left( \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right)^H \right],$$

where  $E_{\mathbf{V}_1, \mathbf{V}_2, \mathbf{H}; F}$  denotes expectation with respect to the pdf  $p_{\mathbf{V}_1, \mathbf{V}_2, \mathbf{H}; F}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{h}; F)$  in (2.28). A Cramér-Rao Bound for the hybrid estimation problem was presented in [77, pg. 329], [69]. Let  $\mathcal{L}(\mathbf{H}, F)$  be the log-likelihood in (2.28), and define the pseudo-information as

$$E_{\mathbf{V}_1, \mathbf{V}_2, \mathbf{H}; F} \left[ \left( \frac{\partial \mathcal{L}(\mathbf{H}, F)}{\partial \boldsymbol{\theta}^*} \right) \left( \frac{\partial \mathcal{L}(\mathbf{H}, F)}{\partial \boldsymbol{\theta}^*} \right)^T \right]. \quad (2.44)$$

The hybrid CRB (HCRB) states that, if the pseudo-information vanishes, then the error covariance is lower bounded<sup>1</sup>

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} \geq \boldsymbol{\mathcal{I}}^{-1}, \quad (2.45)$$

---

<sup>1</sup>The HCRB is given for real parameters in [77, pg. 329]. Here we use the approach described in [45, Sec. 15.7] to state this bound in an equivalent form convenient for complex parameters.

for any  $\hat{\boldsymbol{\theta}} = [\hat{\mathbf{H}}^T, \hat{F}]^T$  such that  $\hat{F}$  is unbiased, where  $\boldsymbol{\mathcal{I}}$  is the hybrid information matrix

$$\boldsymbol{\mathcal{I}} \triangleq E_{\mathbf{V}, \mathbf{H}; F} \left[ \left( \frac{\partial \mathcal{L}(\mathbf{H}, F)}{\partial \boldsymbol{\theta}^*} \right) \left( \frac{\partial \mathcal{L}(\mathbf{H}, F)}{\partial \boldsymbol{\theta}^*} \right)^H \right]. \quad (2.46)$$

For the estimation problem (2.22), from (2.29) it is easy to verify that the pseudonoise vanishes and the hybrid information is

$$\boldsymbol{\mathcal{I}} = \begin{bmatrix} \left( \frac{S_1 + |F|^2 S_2}{\sigma_n^2} \right) \mathbf{I} + \mathbf{C}_{\mathbf{H}}^{-1} & \mathbf{0}_{L \times 1} \\ \mathbf{0}_{1 \times L} & \frac{S_2}{\sigma_n^2} \text{Tr}[\mathbf{C}_{\mathbf{H}}] \end{bmatrix}. \quad (2.47)$$

Thus, the HCRB is

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} \geq \begin{bmatrix} \left[ \left( \frac{S_1 + |F|^2 S_2}{\sigma_n^2} \right) \mathbf{I} + \mathbf{C}_{\mathbf{H}}^{-1} \right]^{-1} & \mathbf{0}_{L \times 1} \\ \mathbf{0}_{1 \times L} & \frac{\sigma_n^2}{S_2 \text{Tr}[\mathbf{C}_{\mathbf{H}}]} \end{bmatrix}. \quad (2.48)$$

Note that uncertainty in  $\mathbf{H}$  affects the estimators differently, in the sense that doubling  $\mathbf{C}_{\mathbf{H}}$  increases the lower bound on  $\hat{\mathbf{H}}$  error covariance, but decreases the bound on the  $\hat{F}$  error covariance. This is intuitively reasonable, since a larger  $\text{Tr}[\mathbf{C}_{\mathbf{H}}]$  essentially increases the signal-to-noise of the observation of  $F$ .

## 2.6 Numerical Results

In this section, we explore the performance of the estimators in Secs. 2.3-2.5 through numerical examples. We take the training sequence  $x_t$  in (2.8) to be a unit-magnitude Zadoff-Chu sequence of length  $T = 64$ . The unknown antenna impedance is  $Z_A = 50\Omega$ . The load impedance (2.6) is  $Z_1 = 50\Omega$  for the first  $K = T/2 = 32$  symbols of each training packet, and  $Z_2 = 70\Omega$  for the remaining symbols. From (2.9), it follows  $F = 1.1667$ .

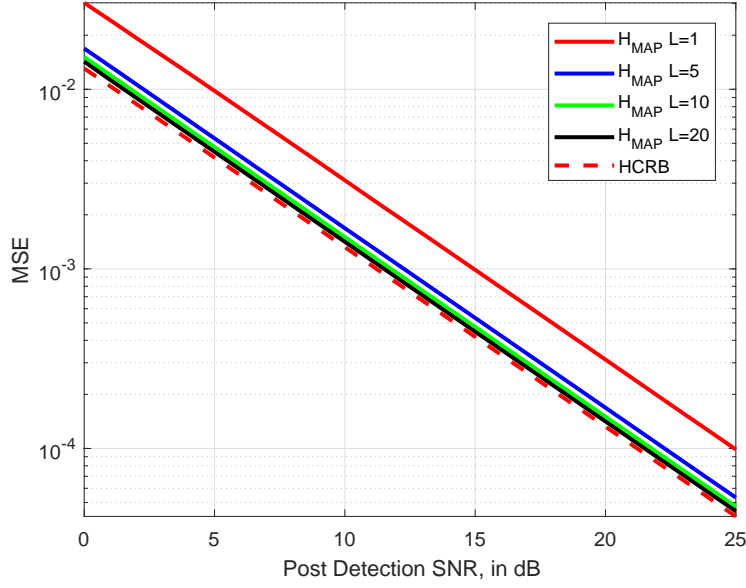


Figure 2.2: Relative MSE of  $\hat{H}_{MAP}$  versus SNR.

Suppose the channel  $\mathbf{H}$  is an i.i.d. sequence, so  $\mathbf{C}_{\mathbf{H}} = \sigma_H^2 \mathbf{I}$ , and define the post-detection signal-to-noise ratio (SNR) of a training symbol as

$$\rho \triangleq \frac{\sigma_H^2}{\sigma_n^2}. \quad (2.49)$$

In Fig. 2.2, we plot the relative mean-squared error (MSE) of the MAP channel estimator versus  $\rho$  for  $L = 1, 5, 10$  and  $20$  training packets. Here the relative MSE is defined as  $E[\|\hat{\mathbf{H}}_{MAP} - \mathbf{H}\|^2]/L\sigma_H^2$ , where  $\hat{\mathbf{H}}_{MAP}$  is given in (2.25) and  $\hat{F}_{ML}$  is the positive root in (2.35). Also shown for comparison is the corresponding HCRB bound on  $\hat{\mathbf{H}}_{MAP}$ , which equals the diagonal elements of the upper-left matrix block in (2.48). Note the single-packet estimator is 3 dB away from the HCRB, since the estimator (2.17) uses only half of the training symbols. When even a few packets are combined, however, the improved estimate of  $F$  enables the receiver to quickly approach the HCRB to within

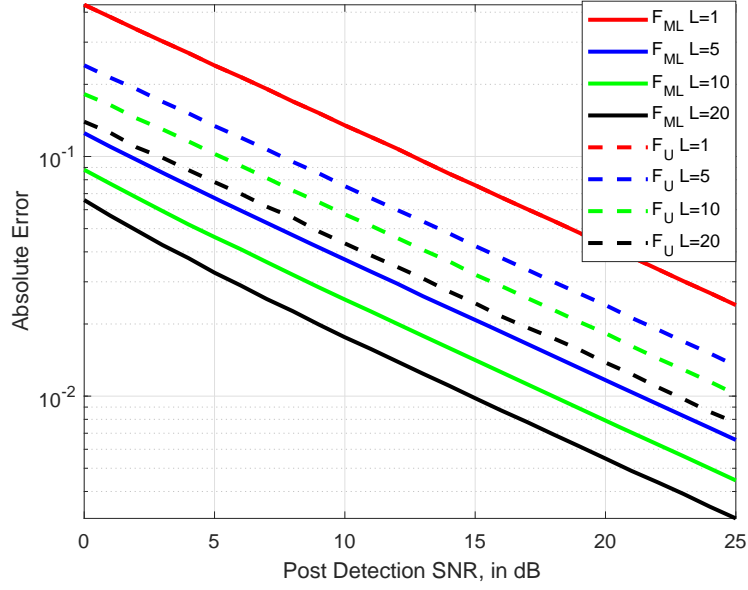


Figure 2.3: Relative MAE of  $\hat{F}_{ML}$  and  $\hat{F}_U$  versus SNR.

a fraction of a dB. In particular, for  $L = 20$ , the efficiency is around 92% for all SNR plotted.

Next consider the performance of  $\hat{F}_{ML}$ . As noted earlier, the  $L = 1$  estimator (2.17) has infinite variance, whereas  $\hat{F}_{ML}$  appears to have finite variance for  $L > 2$ . In order to compare all estimators with the same error metric, we therefore consider the relative mean absolute error (MAE),  $E[|\hat{F}_{ML} - F|]/|F|$ . In Fig. 2.3, we plot the MAE of  $\hat{F}_{ML}$  versus SNR for various numbers of packets  $L$ . Also shown is the simple, unbiased estimator  $\hat{F}_U$ , defined in (2.41). Since  $\hat{F}_{ML}$  and  $\hat{F}_U$  coincide for  $L = 1$ , their MAE does as well. However,  $\hat{F}_{ML}$  significantly outperforms  $\hat{F}_U$  for all  $L > 1$ . In particular, at an MAE of 0.01,  $\hat{F}_{ML}$  saves 5 dB relative to  $\hat{F}_U$  for  $L = 5$ , and 7.5 dB relative to  $\hat{F}_U$  for  $L = 20$ .

As noted earlier,  $\hat{F}_{ML}$  has infinite MSE for  $L = 1$ , but we conjecture MSE is finite for  $L > 2$ . The reason for this conjecture is that  $\hat{F}_{ML}$  has a Gaussian random variable in the denominator for  $L = 1$ , and the inverse Gaussian distribution has infinite variance.

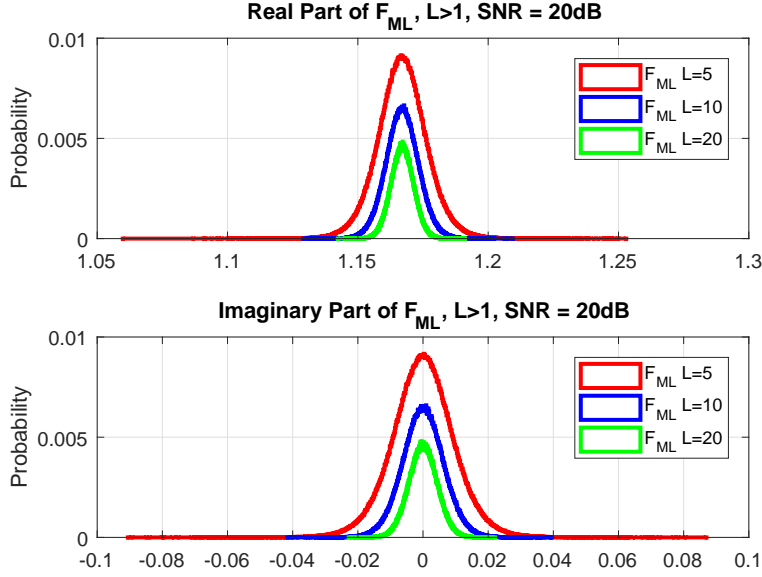


Figure 2.4: Histograms of  $\text{Re}[\hat{F}_{ML}]$  and  $\text{Im}[\hat{F}_{ML}]$  for  $L = 5, 10, 20$ .

For  $L > 1$ , however, the denominator of  $\hat{F}_{ML}$  is chi-square with  $L$  degrees of freedom, and the inverse chi-square has finite variance for  $L > 2$ . Some support for this conjecture can be seen in Fig. 2.4, which plots the empirical distribution  $\hat{F}_{ML}$  for  $L = 5, 10$  and  $20$ , which appear well-described by Gaussian densities in the vicinity of the mode.

If we assume the MSE of  $\hat{F}_{ML}$  is finite for  $L > 2$ , we can compare the performance to the HCRB. In Fig. 2.5, we plot the relative MSE,  $E[|\hat{F}_{ML} - F|^2]/|F|^2$  versus SNR for  $L = 5, 10, 20$ . For comparison, we also plot the HCRB on  $\hat{F}_{ML}$  given in the lower right block of the matrix in (2.48). Note that  $\hat{F}_{ML}$  is less than a dB from the HCRB for high SNRs for  $L = 5$ , and nearly achieves the HCRB for  $L = 10$  and  $20$ . At low SNRs, however,  $\hat{F}_{ML}$  diverges somewhat from the HCRB. One possible explanation for this behavior would be the presence of a bias in  $\hat{F}_{ML}$  at low SNRs. Some support for this hypothesis was given in Sec. 2.5, where we showed that  $\hat{F}_{ML}$  is inconsistent in  $L$ , in the sense that the estimator does not converge to  $F$  as  $L \rightarrow \infty$ . To remedy this situation,

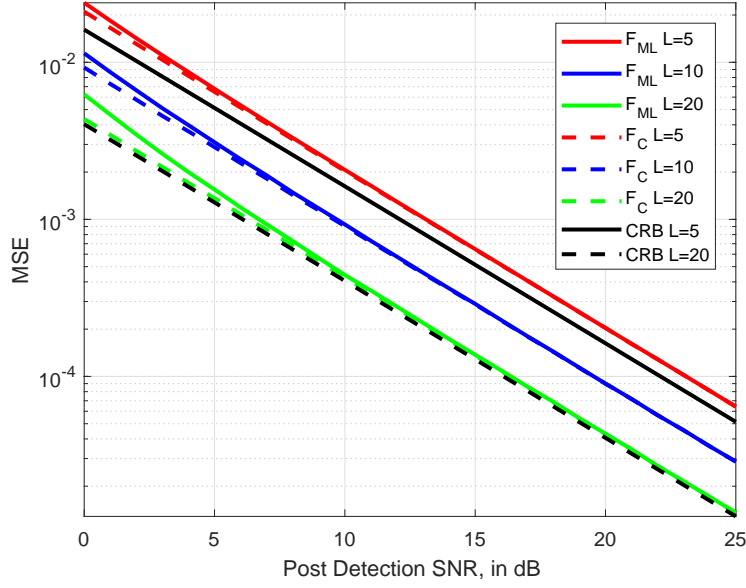


Figure 2.5: MSE of  $\hat{F}_{ML}$  and  $\hat{F}_C$  versus SNR for selected  $L > 2$ .

we defined a modified estimator  $\hat{F}_C$  in (2.43) that is consistent in  $L$ , and which is also plotted in Fig. 2.5. Note  $\hat{F}_C$  performs as well or better than  $\hat{F}_{ML}$  for all SNR and  $L$  in the figure. In particular,  $\hat{F}_C$  outperforms  $\hat{F}_{ML}$  at low SNR.

To better understand the impact of bias, in Fig. 2.6 we plot the absolute relative bias, defined by  $|E[\hat{F}_{ML} - F]|/|F|$ , versus SNR for  $L = 5, 10$  and two values of  $F$ :

$$F_1 = 1.1667, \quad F_2 = 1.2 + 0.4j. \quad (2.50)$$

Note all four curves for  $\hat{F}_{ML}$  seem to coincide, and suggest a bias that decreases with SNR but does not appear to depend on  $L$  or  $F$ . For comparison, we also plot the corresponding curves for the consistent estimator,  $\hat{F}_C$ . Compared with  $\hat{F}_{ML}$ , the bias of  $\hat{F}_C$  at low to medium SNRs appears two orders of magnitude smaller, and does not appear to depend on  $L$ . We do not know if  $\hat{F}_C$  is unbiased, but these results suggest it has a smaller bias

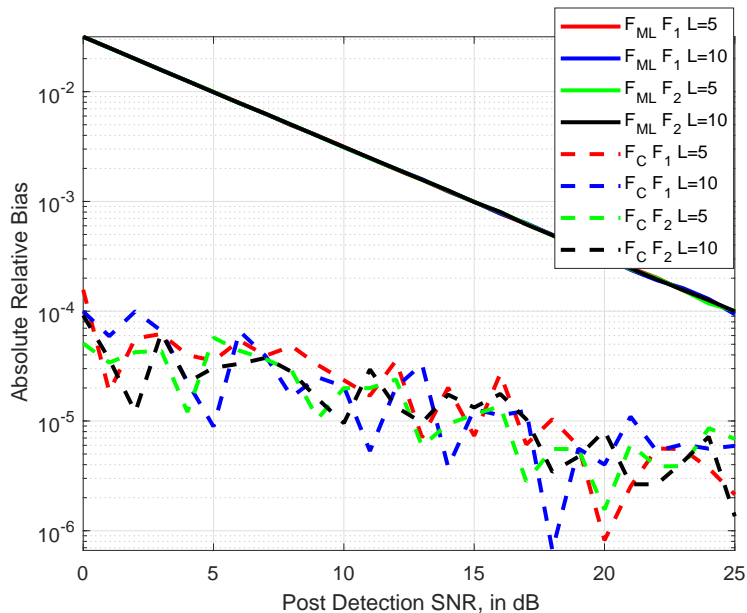


Figure 2.6: Relative Bias of  $\hat{F}_{ML}$  versus SNR and  $L$ .

than  $\hat{F}_{ML}$ .

Thus far, we have considered a fast-fading scenario, in which  $\mathbf{H}$  is i.i.d. As a final example, we now consider a slow fading channel, where  $H_1 = \dots = H_L$ . In Fig. 2.7, we plot the relative MSE of  $\hat{H}_{MAP}$  versus SNR for several values of  $L$ . The performance of  $\hat{H}_{MAP}$  for i.i.d.  $\mathbf{H}$  is also included for comparison. Not surprisingly, the strong correlation in  $\mathbf{H}$  under slow fading conditions leads to a much smaller MSE for  $L > 1$ , since each channel is now observed  $L$  times. However, it is interesting to note that the *opposite* is true of  $\hat{F}_{ML}$ . In Fig. 2.8, we plot the MAE of  $\hat{F}_{ML}$  under the same conditions as Fig. 2.7. Here the i.i.d. channel leads to considerably better estimates of  $F$ , since independent channel observations help us average out variations due to  $\mathbf{H}$ .

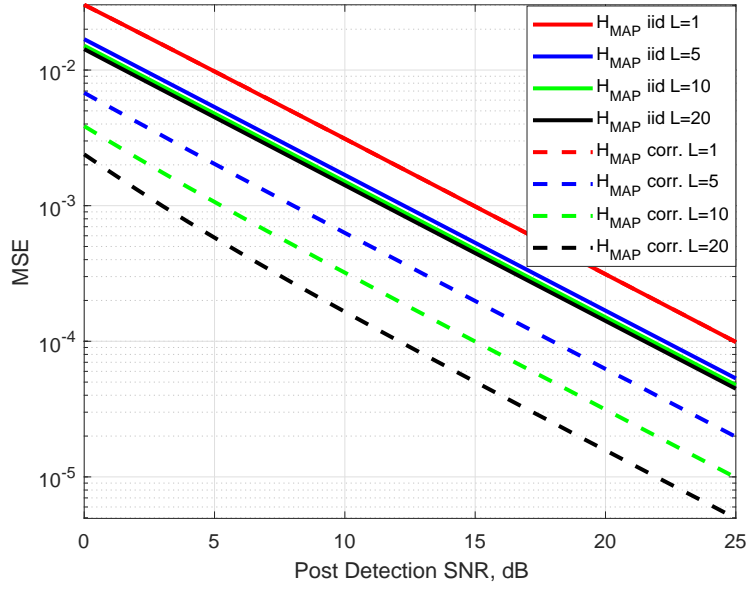


Figure 2.7: MSE of  $\hat{H}_{MAP}$  versus SNR for a correlated channel.

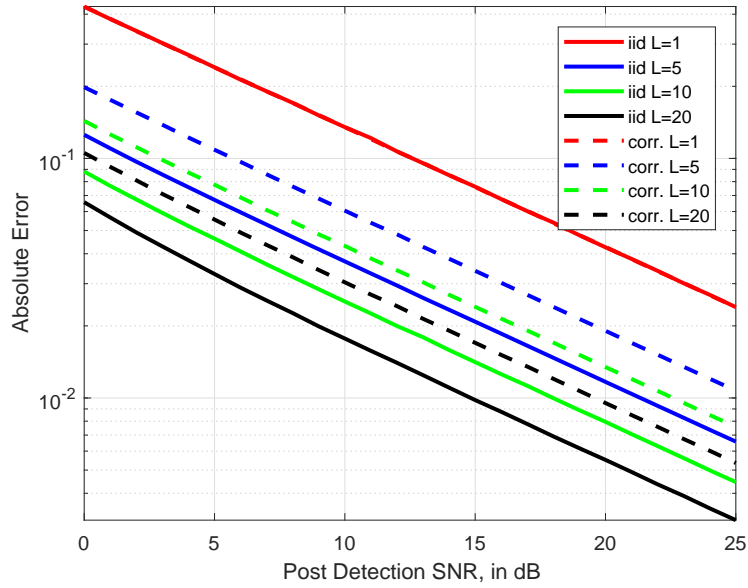


Figure 2.8: MAE of  $\hat{F}_{ML}$  versus SNR for a correlated channel.

## 2.7 Conclusions

In this chapter, we considered joint channel and receiver antenna impedance estimation for fading, multipath channels with a single antenna at both the transmitter and receiver.

We formulated this task as a hybrid estimation problem, where the channel is modeled as an unknown Gaussian random variable, and the impedance is a deterministic parameter. We derived the joint maximum *a posteriori* and maximum-likelihood estimators for the channel and impedance, respectively, as well as the hybrid Cramer-Rao lower bound for any unbiased impedance estimator. Numerical results suggest these estimators perform close to the lower bound when either the SNR or the number of packets is large; however, performance degrades for a small number of packets when the SNR is low. We showed the impedance estimator is generally inconsistent, and numerical results suggest the presence of a non-negligible bias at low SNRs. Motivated by this observation, we proposed an alternate impedance estimator that is provably consistent and appears to be unbiased, which outperforms the joint ML estimator in all of our simulations. In general, we find that temporal correlation of the channel reduces channel estimation error, but increases impedance estimation error. Our results suggest that this approach can utilize existing training data in wireless communications systems to accurately estimate the receiver antenna impedance, in exchange for a very small, controlled increase in channel estimation error.

# Chapter 3

## A Principal-Components Approach to Antenna Impedance Estimation

In the previous chapter, a hybrid estimation framework for joint channel and antenna impedance has been developed. However, this joint ML estimator for antenna impedance is provably inconsistent. In this chapter, we consider the same engineering problem of estimating antenna impedance at a single-antenna receiver but formulate this problem in a classical estimation setting. Under this formulation, the true ML estimator is provably consistent, asymptotically unbiased and consistent [45, Th. 7.3].

### 3.1 Introduction

Antenna impedance matching at mobile receivers has been shown to significantly impact capacity and diversity of wireless channels [26–29, 51, 79]. This matching becomes challenging as antenna impedance changes with time-varying near-field loading, e.g., human users [7, 57, 74]. To mitigate this variation, joint channel and antenna impedance esti-

mators at single-antenna receivers have been derived under classical estimation [80], and hybrid estimation assumptions [81]. However, multiple antennas are often deployed at the transmitter of modern wireless communication systems, e.g., LTE and WiFi [32, 50, 84]. In this chapter, we investigate the effects of transmit diversity on antenna impedance estimation at a single-antenna receiver.

We develop a classical framework of antenna impedance estimation, for single-antenna receivers in Rayleigh fading channels. Based on observation of training sequences via synchronously switched load at the receiver, we derive the true maximum-likelihood (ML) estimators for antenna impedance over multiple packets under different channel correlation assumptions. For the i.i.d. Rayleigh fading channel, this ML estimator uses the eigenvector corresponding to the largest eigenvalue of the sample covariance matrix. When the channel is temporally correlated, the ML estimator is not in closed-form. Thus, necessary conditions are given for the MLE, and we employ numerical methods in finding the MLE. We explore the performance (e.g., bias and efficiency) of these estimators through numerical examples. The impact of channel correlation on impedance estimation accuracy is also investigated numerically.

The rest of the chapter is organized as follows. We present the system model in Sec. 3.2, and derive the true ML estimators for the antenna impedance in Sec. 3.3. We then investigate the more general problem of multiple packets, derive the true ML estimator and a simple method of moments (MM) estimation in Sec. 3.4, and explore the performance of these estimators through numerical examples in Sec. 3.5, and conclude in Sec. 3.6.

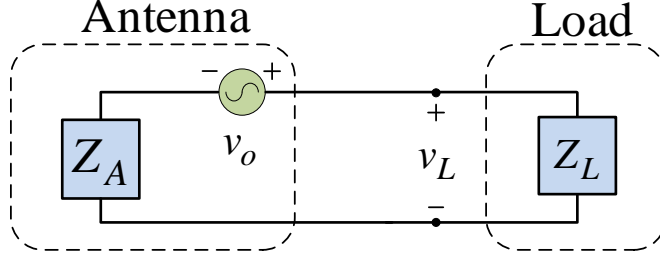


Figure 3.1: Circuit model of a single-antenna receiver

## 3.2 System Model

Consider a narrowband, multiple-input, single-output (MISO) channel with  $N$  transmit antennas and one receive antenna. A circuit model of the receiver is illustrated in Fig. 3.1, which models a scenario in which amplifier noise dominates [29, 51, 79]. In Fig. 3.1, the antenna is modeled by its Thevenin equivalent

$$v = Z_A i + v_o, \quad (3.1)$$

where  $v, i \in \mathbb{C}$  are the voltage across, and current into, the antenna terminals. The antenna impedance is

$$Z_A = R_A + jX_A, \quad (3.2)$$

where  $R_A$  and  $X_A$  are the resistance and reactance, respectively. In (3.1),  $v_o \in \mathbb{C}$  is the open-circuit voltage induced by the incident signal field, which can be modeled in a flat-fading environment as [27] [78, eq. 1],

$$v_o = \mathbf{g}^T \mathbf{x}, \quad (3.3)$$

where  $\mathbf{x} \in \mathbb{C}^N$  is the vector of symbols sent from the  $N$  transmit antennas, and  $\mathbf{g} \in \mathbb{C}^N$  is a vector of channel fading path gains. As in the last chapter, we consider a Rayleigh fading environment in which the transmit antennas are spaced far apart, so the path gains can be modeled as independent, zero-mean Gaussian random variables,  $\mathbf{g} \sim \mathcal{CN}(\mathbf{0}, \sigma_g^2 \mathbf{I}_N)$ .

We assume the estimation algorithms observe a noisy version of the load voltage in Fig. 3.1, given by [29, 51, 79]

$$v_L = \frac{Z_L v_o}{Z_A + Z_L} + n_L, \quad (3.4)$$

where the observation noise  $n_L$  is a zero-mean, Gaussian random variable,  $n_L \sim \mathcal{CN}(0, \sigma_L^2)$ . In estimation theory, the performance of estimators is usually a function of the signal-to-noise ratio (SNR), which is conventionally defined as the squared-magnitude of the signal divided by the noise variance, i.e.  $|Z_L v_o|^2 / (\sigma_L^2 |Z_A + Z_L|^2)$ . As noted in [43], however, since circuit power depends on both voltage and current, this SNR formula does not correctly predict the ratio of the physical signal and noise powers in the receiver front-end. For a given  $v_o$ , the ratio of the physical signal power to noise power at the load is given by [64, eq. 4.65c]

$$\frac{R_L |v_o|^2}{\sigma_n^2 |Z_A + Z_L|^2}, \quad (3.5)$$

where  $\sigma_n^2$  represents the noise power at the output of the amplifier and  $R_L = \text{Re}\{Z_L\}$ . As in [43], we correct this discrepancy by defining  $\sigma_L^2$  in a way that ensures the SNR and physical power ratio (3.5) coincide:

$$\sigma_L^2 \triangleq \frac{|Z_L|^2 \sigma_n^2}{R_L}.$$

With this definition, it is convenient to redefine the observed signal as

$$u \triangleq \frac{\sqrt{R_L}}{Z_L} v_L = \frac{\sqrt{R_L} v_o}{Z_A + Z_L} + n, \quad (3.6)$$

where  $n \sim \mathcal{CN}(0, \sigma_n^2)$  now represents the noise referred to the amplifier output. The model (3.6) now correctly connects estimator performance to the physical signal-to-noise ratio (3.5). This connection is essential in order to accurately predict the impact of impedance mismatch on system-level performance metrics, such as capacity and channel mean-squared estimation error, in the remainder of this chapter.

Suppose the channel gains and antenna impedance are unknown to the receiver. We would like to estimate these parameters using the observations of known training sequences. As in the last chapter, we assume the transmitter sends a known training sequence,  $\mathbf{x}_1, \dots, \mathbf{x}_T \in \mathbb{C}^N$ . During transmission, the receiver shifts synchronously through a sequence of known impedances  $Z_{L,1}, \dots, Z_{L,T}$ . If  $\mathbf{g}$  and  $Z_A$  are fixed for the duration of the training sequence, the received observations are given by

$$u_t = \frac{\sqrt{R_{L,t}} \mathbf{g}^T \mathbf{x}_t}{Z_A + Z_{L,t}} + n_t, \quad t = 1, \dots, T, \quad (3.7)$$

where  $n_t \sim \mathcal{CN}(0, \sigma_n^2)$  are independent and identically distributed.

As in the last chapter, we consider load impedances that take on only two possible values [80],

$$Z_{L,t} = \begin{cases} Z_1, & 1 \leq t \leq K, \\ Z_2, & K < t \leq T. \end{cases} \quad (3.8)$$

where  $Z_1$  and  $Z_2$  are known. Again we assume  $Z_L = Z_1$  is the load impedance used to

receive the transmitted data, and is matched to our best estimate of  $Z_A$ ; additionally  $Z_L = Z_2$  is an impedance variation introduced in order to make  $Z_A$  observable.

With this choice of load impedance, we can express the observations in a simpler, bilinear form. Note that in order to perform optimal detection of the transmitted symbols in (3.6), the receiver needs an estimate of the coefficient of  $\mathbf{x}$ ,

$$\mathbf{h} \triangleq \frac{\sqrt{R_1}}{Z_A + Z_1} \mathbf{g} \sim \mathcal{CN}(\mathbf{0}, \sigma_h^2 \mathbf{I}_N) \ , \quad \sigma_h^2 = \frac{R_1 \cdot \sigma_g^2}{|Z_A + Z_1|^2} \ , \quad (3.9)$$

where  $R_1 = \text{Re}\{Z_1\}$ . It therefore makes sense to estimate  $\mathbf{h}$  directly, rather than the path gains  $\mathbf{g}$ . Expressing the observations (3.7) in terms of  $\mathbf{h}$ , we obtain the bilinear model

$$u_t = \begin{cases} \mathbf{h}^T \mathbf{x}_t + n_t \ , & 1 \leq t \leq K \ , \\ F \mathbf{h}^T \mathbf{x}_t + n_t \ , & K < t \leq T \ , \end{cases} \quad (3.10)$$

where we define

$$F \triangleq \frac{\sqrt{R_1}(Z_1 + Z_A)}{\sqrt{R_2}(Z_2 + Z_A)} \ . \quad (3.11)$$

The goal of this chapter is to derive estimators for  $\mathbf{h}$ ,  $Z_A$  and  $\sigma_h^2$  based on the observations (3.10). From the invariance principle of maximum-likelihood estimation (MLE) [45, pg. 185], knowing the MLE of  $F$  is equivalent to knowing that of  $Z_A$  and vice versa. It therefore suffices to derive estimators for  $\mathbf{h}$ ,  $F$  and  $\sigma_h^2$ . In the last chapter, we proposed a hybrid approach to jointly estimate  $\mathbf{h}$  and  $F$  for the special case  $N = 1$  and  $\sigma_h^2$  known. In this chapter, we consider an alternative approach for  $N \geq 1$  and  $\sigma_h^2$  unknown, in which we solve this problem in two successive steps: First, we consider joint maximum-likelihood estimation of  $F$  and  $\sigma_h^2$ , treating  $\mathbf{h}$  as a nuisance parameter. Second, given estimates of  $F$  and  $\sigma_h^2$ , we then estimate  $\mathbf{h}$  using minimum mean-squared error estimation. Since the

second step involves well-known techniques, we focus exclusively on estimators for  $F$  and  $\sigma_h^2$  in the next few sections; estimators for  $\mathbf{h}$  will be explored through numerical examples in Sec. 3.5.

### 3.3 Maximum-Likelihood Estimators

In this section, we derive joint maximum-likelihood estimators for the channel variance  $\sigma_h^2$  in (3.9) and the antenna impedance, represented here by the parameter  $F$  in (3.11). Before doing so, it is convenient to reduce the observations (3.10) to sufficient statistics.

**Lemma 2 (Sufficient Statistics)** : *Consider the observations  $\mathbf{u} \triangleq (u_1, \dots, u_T)^T$  in (3.10), where  $F$  and  $\sigma_h^2$  are unknown constants. Suppose the matrices*

$$\mathbf{B}_1 \triangleq \sum_{t=1}^K \mathbf{x}_t^* \mathbf{x}_t^T, \quad \mathbf{B}_2 \triangleq \sum_{t=K+1}^T \mathbf{x}_t^* \mathbf{x}_t^T, \quad (3.12)$$

are non-singular, where  $\mathbf{x}_1, \dots, \mathbf{x}_T$  is the training sequence. Then

$$\mathbf{y}_1 \triangleq \mathbf{B}_1^{-1} \sum_{t=1}^K u_t \mathbf{x}_t^*, \quad \mathbf{y}_2 = \mathbf{B}_2^{-1} \sum_{t=K+1}^T u_t \mathbf{x}_t^*, \quad (3.13)$$

are sufficient statistics to estimate  $F$  and  $\sigma_h^2$  based on the observation  $\mathbf{u}$ . Moreover,  $\mathbf{y}_1$  and  $\mathbf{y}_2$  are conditionally independent given  $\mathbf{h}$ , with conditional distributions

$$\mathbf{y}_1 \sim \mathcal{CN}(\mathbf{h}, \sigma_n^2 \mathbf{B}_1^{-1}), \quad \mathbf{y}_2 \sim \mathcal{CN}(F\mathbf{h}, \sigma_n^2 \mathbf{B}_2^{-1}). \quad \diamond$$

**Proof** : From (3.10), note the observations  $\mathbf{u} \triangleq (u_1, \dots, u_T)^T$  are conditionally independent given  $\mathbf{h}$ , with conditional distributions  $u_t \sim \mathcal{CN}(\mathbf{h}^T \mathbf{x}_t, \sigma_n^2)$  for  $1 \leq t \leq K$ , and

$u_t \sim \mathcal{CN}(F\mathbf{h}^T \mathbf{x}_t, \sigma_n^2)$  otherwise. The conditional distributions of  $\mathbf{y}_1$  and  $\mathbf{y}_2$  in the lemma follow immediately by substituting these conditional pdfs into (3.13).

From the Neyman-Fisher Theorem [45, pg. 117], to prove sufficiency it suffices to show  $p(\mathbf{u}; F, \sigma_h^2)$  can be factored into a product  $g(\mathbf{y}_1, \mathbf{y}_2, F, \sigma_h^2)f(\mathbf{u})$ , where  $f$  does not depend on  $\mathbf{y}_1, \mathbf{y}_2, F$  or  $\sigma_h^2$ , and  $g$  does not depend on  $\mathbf{u}$ .

To this end, we can express this pdf in terms of the conditional pdf as

$$p(\mathbf{u}; F, \sigma_h^2) = E \left[ p(\mathbf{u}|\mathbf{h}; F, \sigma_h^2) \right],$$

where  $E[\cdot]$  denotes expectation with respect to  $\mathbf{h}$ . Since  $u_1, \dots, u_T$  are conditionally independent given  $\mathbf{h}$ , we can write

$$\begin{aligned} (\pi\sigma_n^2)^T \cdot p(\mathbf{u}; F, \sigma_h^2) &= E \left[ \exp \left( -\frac{1}{\sigma_n^2} \sum_{t=1}^K |u_t - \mathbf{h}^T \mathbf{x}_t|^2 - \frac{1}{\sigma_n^2} \sum_{t=K+1}^T |u_t - F\mathbf{h}^T \mathbf{x}_t|^2 \right) \right] \\ &= E \left[ \exp \left( -\frac{1}{\sigma_n^2} \sum_{t=1}^K \{ |\mathbf{h}^T \mathbf{x}_t|^2 - u_t^* \mathbf{h}^T \mathbf{x}_t - u_t \mathbf{h}^H \mathbf{x}_t^* \} \right) \right. \\ &\quad \left. \times \exp \left( -\frac{1}{\sigma_n^2} \sum_{t=K+1}^T \{ |F\mathbf{h}^T \mathbf{x}_t|^2 - u_t^* F\mathbf{h}^T \mathbf{x}_t - u_t F^* \mathbf{h}^H \mathbf{x}_t^* \} \right) \right] \exp \left( -\frac{|\mathbf{u}|^2}{\sigma_n^2} \right) \\ &= E \left[ \exp \left( \frac{1}{\sigma_n^2} \left\{ 2 \operatorname{Re}[\mathbf{h}^H \mathbf{B}_1 \mathbf{y}_1] + 2 \operatorname{Re}[F^* \mathbf{h}^H \mathbf{B}_2 \mathbf{y}_2] - \sum_{t=1}^K |\mathbf{h}^T \mathbf{x}_t|^2 - \sum_{t=K+1}^T |F\mathbf{h}^T \mathbf{x}_t|^2 \right\} \right) \right] \\ &\quad \times \exp \left( -\frac{|\mathbf{u}|^2}{\sigma_n^2} \right). \end{aligned}$$

Equate the first term with  $(\pi\sigma_n^2)^T g(\mathbf{y}_1, \mathbf{y}_2, F, \sigma_h^2)$ , and the second with  $f(\mathbf{u})$ . Note  $f$  does not depend on  $\mathbf{y}_1, \mathbf{y}_2, F$ , or  $\sigma_h^2$ , while  $g$  depends on  $\mathbf{y}_1, \mathbf{y}_2, F$  and  $\sigma_h^2$  (through the expectation), but not  $\mathbf{u}$ . This completes the proof.

Training sequences for MIMO channel estimation are often chosen to be orthogo-

nal and equal-energy, so  $\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t^H = (PT/N) \mathbf{I}_N$ . In this section, we assume the load impedance (3.8) switches halfway through the training sequence, so  $K = T/2$ , and the training sequences are also equal-energy and orthogonal over the first and last  $K$  symbols, which implies

$$\mathbf{B}_1 = \mathbf{B}_2 = \left( \frac{PT}{2N} \right) \mathbf{I}_N, \quad (3.14)$$

where  $\mathbf{B}_1, \mathbf{B}_2$  are defined in (3.12). For example, this can be achieved by using a normalized discrete Fourier transform (DFT) matrix [16, eq. 10].

With these assumptions, we now derive the maximum-likelihood estimate of the parameter vector

$$\boldsymbol{\theta} \triangleq \begin{bmatrix} F \\ \sigma_h^2 \end{bmatrix}, \quad (3.15)$$

based on the single-packet sufficient statistics (3.13), where  $F$  is defined in (3.11) and  $\sigma_h^2$  in (3.9). This estimate is defined by

$$\boldsymbol{\theta}_{ML} \triangleq \arg \max_{\boldsymbol{\theta}} p(\mathbf{y}_1, \mathbf{y}_2; \boldsymbol{\theta}). \quad (3.16)$$

The following theorem shows that these estimators can easily be calculated from the principal component of a sample covariance matrix.

**Theorem 5 (Single-Packet ML Estimators)** : *Let  $\mathbf{y}_1$  and  $\mathbf{y}_2$  be the sufficient statistics in (3.13), where  $F$  and  $\sigma_h^2$  are unknown constants. Consider the sample covariance*

$$\mathbf{S} \triangleq \frac{1}{N} \begin{bmatrix} \mathbf{y}_1^H \mathbf{y}_1 & \mathbf{y}_2^H \mathbf{y}_1 \\ \mathbf{y}_1^H \mathbf{y}_2 & \mathbf{y}_2^H \mathbf{y}_2 \end{bmatrix}. \quad (3.17)$$

*Let  $\eta_1$  be the largest eigenvalue of  $\mathbf{S}$ , and  $\hat{\mathbf{e}}_1 = [E_1, E_2]^T$  any associated unit eigenvector.*

Then the maximum-likelihood estimates of  $F$  and  $\sigma_h^2$  are given by

$$\hat{\boldsymbol{\theta}}_{ML} = \begin{bmatrix} \hat{F}_{ML} \\ \hat{\sigma}_h^2 \end{bmatrix} = \begin{bmatrix} E_2/E_1 \\ |E_1|^2 \max\{\eta_1 - \sigma^2, 0\} \end{bmatrix}, \quad (3.18)$$

provided  $E_1 \neq 0$ , where  $\sigma^2 \triangleq \sigma_n^2(2N/PT)$ . For  $E_1 = 0$  and  $\eta_1 > \sigma^2$  the likelihood is maximized in the limit as  $F \rightarrow \infty$ .  $\diamond$

**Proof :** From (3.14) and Lemma 2,  $\mathbf{y}_1$  and  $\mathbf{y}_2$  are conditionally independent given  $\mathbf{h}$ , and their conditional distributions are  $\mathbf{y}_1 \sim \mathcal{CN}(\mathbf{h}, \sigma^2 \mathbf{I}_N)$  and  $\mathbf{y}_2 \sim \mathcal{CN}(F\mathbf{h}, \sigma^2 \mathbf{I}_N)$ , where  $\sigma^2 \triangleq \sigma_n^2(2N/PT)$ . Since  $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}_N, \sigma_h^2 \mathbf{I}_N)$  is independent of the noise in (3.13), it follows  $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)^T$  is a zero-mean Gaussian random vector with covariance

$$E[\mathbf{y}\mathbf{y}^H] = \begin{bmatrix} (\sigma_h^2 + \sigma^2)\mathbf{I}_N & \sigma_h^2 F^* \mathbf{I}_N \\ \sigma_h^2 F \mathbf{I}_N & (\sigma_h^2 |F|^2 + \sigma^2)\mathbf{I}_N \end{bmatrix} = \mathbf{C} \otimes \mathbf{I}_N, \quad (3.19)$$

where  $\otimes$  is the Kronecker product and

$$\mathbf{C} \triangleq \mathbf{C}(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_h^2 + \sigma^2 & \sigma_h^2 F^* \\ \sigma_h^2 F & \sigma_h^2 |F|^2 + \sigma^2 \end{bmatrix}. \quad (3.20)$$

It follows

$$\begin{aligned} p(\mathbf{y}_1, \mathbf{y}_2; \boldsymbol{\theta}) &= \frac{1}{\det(\pi[\mathbf{C} \otimes \mathbf{I}_N])} \exp(-\mathbf{y}^H [\mathbf{C} \otimes \mathbf{I}_N]^{-1} \mathbf{y}), \\ &= \det(\pi \mathbf{C})^{-N} \exp(-\mathbf{y}^H [\mathbf{C}^{-1} \otimes \mathbf{I}_N] \mathbf{y}), \\ &= \det(\pi \mathbf{C})^{-N} \exp(-N \text{Tr}[\mathbf{S}\mathbf{C}^{-1}]), \end{aligned} \quad (3.21)$$

where  $\mathbf{S}$  is defined in (3.17), and  $\text{Tr}[\mathbf{A}\mathbf{B}] = \text{Tr}[\mathbf{B}\mathbf{A}]$  [45, pg. 571].

Note  $\mathbf{C}$  can be written in terms of its eigensystem as

$$\mathbf{C} = \mu_1 \mathbf{e}_1 \mathbf{e}_1^H + \mu_2 \mathbf{e}_2 \mathbf{e}_2^H, \quad (3.22)$$

where  $\mu_1 \geq \mu_2$  are the ordered eigenvalues and  $\mathbf{e}_1, \mathbf{e}_2$  are the associated unit eigenvectors. From (3.20), it is easy to verify the following explicit formulas,

$$\begin{aligned} \mu_1 &= \sigma_h^2(1 + |F|^2) + \sigma^2, & \mu_2 &= \sigma^2 \\ \mathbf{e}_1 &= \frac{1}{\sqrt{1 + |F|^2}} \begin{bmatrix} 1 \\ F \end{bmatrix}, & \mathbf{e}_2 &= \frac{1}{\sqrt{1 + |F|^2}} \begin{bmatrix} -F^* \\ 1 \end{bmatrix}. \end{aligned} \quad (3.23)$$

To find maximum-likelihood estimates of  $F$  and  $\sigma_h^2$ , we proceed in two steps: First, we find conditions on  $\mu_1, \mathbf{e}_1, \mathbf{e}_2$  that achieve the maximum in (3.16). (Note the value of  $\mu_2$  is fixed at  $\sigma^2$ .) Second, we use (3.23) to translate these conditions into values of  $F$  and  $\sigma_h^2$ . From (3.20), observe

$$\mathbf{C}^{-1} = \mu_1^{-1} \mathbf{e}_1 \mathbf{e}_1^H + \mu_2^{-1} \mathbf{e}_2 \mathbf{e}_2^H, \quad (3.24)$$

so the log of the likelihood (3.21) can be expressed in terms of the eigensystem as

$$\begin{aligned} \ln p(\mathbf{y}_1, \mathbf{y}_2; \boldsymbol{\theta}) &= -N \ln \det(\pi \mathbf{C}) - N \text{Tr}[\mathbf{S} \mathbf{C}^{-1}] \\ &= -N \ln(\pi \mu_1 \mu_2) - \frac{N \mathbf{e}_1^H \mathbf{S} \mathbf{e}_1}{\mu_1} - \frac{N \mathbf{e}_2^H \mathbf{S} \mathbf{e}_2}{\mu_2} \\ &= -N \ln(\pi \mu_1 \mu_2) + (\mu_2^{-1} - \mu_1^{-1}) N \mathbf{e}_1^H \mathbf{S} \mathbf{e}_1 - \mu_2^{-1} N \text{Tr}[\mathbf{S}], \end{aligned} \quad (3.25)$$

where the last equality follows by observing  $\mathbf{U} \triangleq [\mathbf{e}_1, \mathbf{e}_2]$  is unitary, and

$$\text{Tr}[\mathbf{S}] = \text{Tr}[\mathbf{U}^H \mathbf{S} \mathbf{U}] = \mathbf{e}_1^H \mathbf{S} \mathbf{e}_1 + \mathbf{e}_2^H \mathbf{S} \mathbf{e}_2 . \quad (3.26)$$

From (3.23), the coefficient  $\mu_2^{-1} - \mu_1^{-1}$  in (3.25) is non-negative. Since  $\mathbf{S}$  is Hermitian and positive semi-definite, it has real eigenvalues, say  $\eta_1 \geq \eta_2 \geq 0$ . From the Rayleigh-Ritz theorem,  $\mathbf{e}_1^H \mathbf{S} \mathbf{e}_1 \leq \eta_1$  with equality if and only if  $\mathbf{e}_1$  is an eigenvector of  $\mathbf{S}$  corresponding to  $\eta_1$ . Since  $\text{Tr}[\mathbf{S}] = \eta_1 + \eta_2$ , it follows

$$\begin{aligned} \ln p(\mathbf{y}_1, \mathbf{y}_2; \boldsymbol{\theta}) &\leq -N \ln(\pi \mu_1 \mu_2) - \frac{N \eta_1}{\mu_1} - \frac{N \eta_2}{\mu_2} \\ &= N \left[ \ln \left( \frac{\eta_1}{\mu_1} \right) - \frac{\eta_1}{\mu_1} \right] - N \ln(\pi \eta_1 \mu_2) - \frac{N \eta_2}{\mu_2} , \end{aligned} \quad (3.27)$$

with equality if and only if  $\mu_1 = \mu_2$  or  $\mathbf{e}_1$  is an eigenvector of  $\mathbf{S}$  corresponding to  $\eta_1$ . Note the function  $\ln x - x$  is concave and uniquely maximized at  $x = 1$ . It follows that the bracketed term in (3.27) is maximized over  $\mu_1 \geq \mu_2 = \sigma^2$  by choosing  $\mu_1 = \max\{\eta_1, \sigma^2\}$ .

Finally, we translate these conditions into values of  $F$  and  $\sigma_h^2$ : If  $\eta_1 \leq \sigma^2$ , then  $\mu_1 = \mu_2 = \sigma^2$  and the likelihood (3.25) does not depend on  $F$ . From (3.23), it follows the likelihood is maximized by  $\hat{\sigma}_h^2 = 0$  and any value of  $F$ ; In particular, (3.18) maximizes the likelihood. However, if  $\eta_1 > \sigma^2$ , then  $\mu_1 = \eta_1$  is optimal and hence  $\mathbf{e}_1$  must be an eigenvector  $(E_1, E_2)^T$  of  $\mathbf{S}$  corresponding to  $\eta_1$ . For  $E_1 \neq 0$ , the unique solution of the equations  $\sigma_h^2 = \eta_1$  and  $\mathbf{e}_1 = (E_1, E_2)^T$  in (3.23) is given by (3.18). For  $E_1 = 0$ , no finite  $F$  solves these equations; rather, the solution is approached in the limit as  $F \rightarrow \infty$ .

The theorem above shows the maximum-likelihood estimators can be expressed in terms of eigensystem of the sample covariance  $\mathbf{S}$ . Since  $\mathbf{S}$  is a  $2 \times 2$  Hermitian matrix, it is possible to derive closed-form formulas for the eigensystem. This leads to closed-form

expressions for the estimators, which are given in the following corollary.

**Corollary 1 (Closed-Form ML Estimators)** : Denote the elements of the sample covariance  $\mathbf{S}$  in (3.17) by <sup>1</sup>

$$\mathbf{S} = \frac{1}{N} \begin{bmatrix} \mathbf{y}_1^H \mathbf{y}_1 & \mathbf{y}_2^H \mathbf{y}_1 \\ \mathbf{y}_1^H \mathbf{y}_2 & \mathbf{y}_2^H \mathbf{y}_2 \end{bmatrix} \triangleq \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}. \quad (3.28)$$

The maximum-likelihood estimators in (3.18) can be expressed in closed-form as

$$\hat{F}_{ML} = \frac{S_{22} - S_{11} + \sqrt{(S_{22} - S_{11})^2 + 4|S_{12}|^2}}{2S_{12}}, \quad (3.29)$$

$$\hat{\sigma}_h^2 = \frac{|\hat{F}_{ML}|^2}{|\hat{F}_{ML}|^2 + 1} \max \left\{ S_{11} + S_{12} \hat{F}_{ML} - \sigma^2, 0 \right\}, \quad (3.30)$$

provided  $S_{12} \neq 0$ . ◇

**Proof** : Since  $\mathbf{S}$  is Hermitian,  $S_{11}$  and  $S_{22}$  are real and non-negative and  $S_{12} = S_{21}^*$ . The eigenvalues are given by the two non-negative roots of the polynomial

$$\det[\mathbf{S} - \eta \mathbf{I}] = \eta^2 - (S_{11} + S_{22})\eta + S_{11}S_{22} - |S_{12}|^2,$$

of which the larger is

$$\eta_1 = \frac{S_{22} + S_{11} + \sqrt{(S_{11} - S_{22})^2 + 4|S_{12}|^2}}{2}. \quad (3.31)$$

Next we find a unit eigenvector  $\mathbf{e}_1 = [E_1, E_2]^T$  associated with  $\eta_1$ . Any such vector

---

<sup>1</sup>Note the entry  $S_{ij} = \mathbf{y}_j^H \mathbf{y}_i / N$  corresponds to  $P_{ji}$  in our prequel [81, eq. 31], where  $1 \leq i, j \leq 2$ .

must satisfy  $|E_1|^2 + |E_2|^2 = 1$  and

$$\begin{bmatrix} S_{11} - \eta_1 & S_{12} \\ S_{21} & S_{22} - \eta_1 \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} .$$

It is easy to verify that one solution of these equations is

$$\begin{bmatrix} E_1 \\ E_2 \end{bmatrix} = \frac{1}{\sqrt{|\eta_1 - S_{11}|^2 + |S_{12}|^2}} \begin{bmatrix} S_{12} \\ \eta_1 - S_{11} \end{bmatrix} . \quad (3.32)$$

From (3.18), it follows

$$\hat{F}_{ML} = \frac{\eta_1 - S_{11}}{S_{12}} , \quad \hat{\sigma}_h^2 = \frac{|\eta_1 - S_{11}|^2}{|\eta_1 - S_{11}|^2 + |S_{12}|^2} \max \{ \eta_1 - \sigma^2, 0 \} ,$$

Substituting (3.31) into the first equation gives (3.29); substituting  $\eta_1 = S_{11} + S_{12}\hat{F}_{ML}$  into the second gives (3.30).

We note (3.29) is similar to estimators that arise in the so-called errors-in-variables regression literature, cf. [31, pg. 294, Case 4] and [40, Tab. 1, Cases 1&5]. In Appendix B.1, we also show that  $\hat{F}_{ML}$  equals the ML estimator of  $F$  when both  $\mathbf{h}$  and  $F$  are modeled as unknown constants. Intuitively, this implies that modeling  $\mathbf{h}$  as a zero-mean Gaussian with an unknown variance, and modeling  $\mathbf{h}$  as an unknown constant, both lead to the same ML estimate of  $F$ .

Finally, in order to evaluate the efficiency of these estimators, we now derive the Cramér-Rao Bound for the error covariance of these estimators,

$$\mathbf{C}_{\hat{\theta}} \triangleq E_{\mathbf{y}_1, \mathbf{y}_2; \theta} \left[ \left( \hat{\theta} - \theta \right) \left( \hat{\theta} - \theta \right)^H \right] ,$$

where  $E_{\mathbf{y}_1, \mathbf{y}_2; \boldsymbol{\theta}}$  denotes expectation with respect to the pdf  $p(\mathbf{y}_1, \mathbf{y}_2; \boldsymbol{\theta})$  in (3.27). The Cramér-Rao Bound is given by

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} \geq \mathbf{I}(\boldsymbol{\theta})^{-1},$$

where  $\hat{\boldsymbol{\theta}}$  is any unbiased estimator of  $\boldsymbol{\theta} = [\theta_1, \theta_2]^T = [F, \sigma_h^2]^T$ , and  $\mathbf{I}(\boldsymbol{\theta})$  is the Fisher information matrix (FIM) [45, pg. 529]

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = N \text{Tr} \left[ \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_i^*} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_j} \right], \quad (3.33)$$

where  $\mathbf{C}$  is given in (3.20). Here we use the approach described in [45, Sec. 15.7] to state the FIM in an equivalent form convenient for complex  $\boldsymbol{\theta}$ .

It follows

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{N(1 + |F|^2)}{[\sigma_h^2(1 + |F|^2) + \sigma^2]^2} \begin{bmatrix} \sigma_h^4(1 + \sigma_h^2/\sigma^2) & F\sigma_h^2 \\ F^*\sigma_h^2 & 1 + |F|^2 \end{bmatrix}; \quad (3.34)$$

hence the CRB is given by

$$\mathbf{I}(\boldsymbol{\theta})^{-1} = \frac{\sigma_h^2(1 + |F|^2) + \sigma^2}{N\sigma_h^4(1 + |F|^2)} \begin{bmatrix} \sigma^2(1 + |F|^2) & -F\sigma^2\sigma_h^2 \\ -F^*\sigma^2\sigma_h^2 & \sigma_h^4(\sigma_h^2 + \sigma^2) \end{bmatrix}. \quad (3.35)$$

We are interested primarily in bounds on the mean-squared error of unbiased estimators of  $F$  and  $\sigma_h^2$ . These are given, respectively, by the diagonal entries of (3.36):

$$\mathcal{C}_F(\boldsymbol{\theta}) \triangleq \frac{\sigma^2\sigma_h^2(1 + |F|^2) + \sigma^4}{N\sigma_h^4}, \quad \mathcal{C}_{\sigma_h^2}(\boldsymbol{\theta}) \triangleq \frac{(\sigma_h^2(1 + |F|^2) + \sigma^2)(\sigma_h^2 + \sigma^2)}{N(1 + |F|^2)}. \quad (3.36)$$

### 3.4 Estimators for Multiple Packets

In the last section, we derived estimators for  $F$  and  $\sigma_h^2$  based on a single training packet. In this section, we consider estimators based on multiple packets where the channel evolves in an unknown way from packet to packet.

Suppose the transmitter sends a sequence of  $L$  identical training packets to the receiver. During reception of each packet, the receiver load shifts in the same way as described in (3.8). As in the last section, we assume the channel is constant *within* a packet, but varies from packet to packet in a random way. Under these assumptions, the signal observed during the  $k$ -th packet can be described by a model similar to (3.10):

$$u_{k,t} = \begin{cases} \mathbf{h}_k^T \mathbf{x}_t + n_{k,t}, & 1 \leq t \leq K, \\ F\mathbf{h}_k^T \mathbf{x}_t + n_{k,t}, & K < t \leq T, \end{cases} \quad (3.37)$$

where  $F$  is still defined by (3.11),  $\mathbf{h}_k$  is the channel during the  $k$ -th packet, and the noise variable  $n_{k,t} \sim \mathcal{CN}(0, \sigma_n^2)$  are i.i.d. We can express these observations in a compact matrix form as

$$\mathbf{U}_1 = \mathbf{H}\mathbf{X}_1 + \mathbf{N}_1, \mathbf{U}_2 = F\mathbf{H}\mathbf{X}_2 + \mathbf{N}_2 \quad (3.38)$$

where  $\mathbf{X}_1 \triangleq [\mathbf{x}_1, \dots, \mathbf{x}_K] \in \mathbb{C}^{N \times K}$ ,  $\mathbf{X}_2 \triangleq [\mathbf{x}_{K+1}, \dots, \mathbf{x}_T] \in \mathbb{C}^{N \times (T-K)}$ ,

$$\mathbf{H} \triangleq [\mathbf{h}_1, \dots, \mathbf{h}_L]^T \in \mathbb{C}^{L \times N}, \quad (3.39)$$

and  $\mathbf{U}_1 \in \mathbb{C}^{L \times K}$ ,  $\mathbf{U}_2 \in \mathbb{C}^{L \times (N-K)}$ ,  $\mathbf{N}_1 \in \mathbb{C}^{L \times K}$  and  $\mathbf{N}_2 \in \mathbb{C}^{L \times (N-K)}$  are defined analogously. It follows  $\mathbf{N}_1$  and  $\mathbf{N}_2$  are independent random matrices with i.i.d.  $\mathcal{CN}(0, \sigma_n^2)$  entries. Here  $\mathbf{H}$  models Rayleigh fading path gains which are uncorrelated in space but

not necessarily in time. This implies the columns of  $\mathbf{H}$  are i.i.d. zero-mean Gaussian random vectors with an temporal correlation matrix  $\sigma_h^2 \mathbf{C}_{\mathbf{H}} \in \mathbb{C}^{L \times L}$ . Here we assume the correlation structure of  $\mathbf{H}$  is known except for the power, so  $\mathbf{C}_{\mathbf{H}}$  is known but  $\sigma_h^2$  is unknown. As in the last section, we assume  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are known at the receiver,  $K = T/2$ , and the training sequences are equal-energy and orthogonal over the first and last  $K$  symbols, which implies

$$\mathbf{X}_1 \mathbf{X}_1^H = \mathbf{X}_2 \mathbf{X}_2^H = \left( \frac{PT}{2N} \right) \mathbf{I}_N. \quad (3.40)$$

The goal of this chapter is to derive estimators for  $\mathbf{H}$ ,  $F$  and  $\sigma_h^2$  based on the observations (3.38). As in the last section, we approach the problem in two steps: In this section, we consider joint maximum-likelihood estimation of  $F$  and  $\sigma_h^2$ , treating  $\mathbf{H}$  as a nuisance parameter. In Sec. 3.5, we will explore estimators for  $\mathbf{H}$  given  $F$  and  $\sigma_h^2$  through numerical examples. The following lemma generalizes Lemma 2 to multiple packets.

**Lemma 3 (Multi-Packet Sufficient Statistics)** : *Consider the observations  $\mathbf{U}_1, \mathbf{U}_2$  defined in (3.38), where  $\mathbf{X}_1, \mathbf{X}_2$  are known training sequences and  $F$  and  $\sigma_h^2$  are unknown constants. Then*

$$\mathbf{Y}_1 \triangleq \left( \frac{2N}{PT} \right) \mathbf{U}_1 \mathbf{X}_1^H, \quad \mathbf{Y}_2 \triangleq \left( \frac{2N}{PT} \right) \mathbf{U}_2 \mathbf{X}_2^H, \quad (3.41)$$

are sufficient statistics to estimate  $F$  and  $\mathbf{C}_{\mathbf{H}}$  based on the observation  $\mathbf{U}_1, \mathbf{U}_2$ . Moreover,  $\mathbf{Y}_1 - \mathbf{H}$  and  $\mathbf{Y}_2 - F\mathbf{H}$  are independent random matrices with i.i.d.  $\mathcal{CN}(0, \sigma^2)$  entries, where  $\sigma^2 \triangleq 2N\sigma_n^2/P$ .  $\diamond$

**Proof** : From (3.38) and (3.40), we have  $\mathbf{Y}_1 = \mathbf{H} + (2N/PT)\mathbf{N}_1 \mathbf{X}_1^H$ . Note the rows of  $(2N/PT)\mathbf{N}_1 \mathbf{X}_1^H$  are i.i.d. with covariance  $\sigma_n^2(2N/PT)^2 \mathbf{X}_1 \mathbf{X}_1^H = \sigma^2 \mathbf{I}_N$ , where the last step

follows from (3.40). Similarly,  $\mathbf{Y}_2 = F\mathbf{H} + (2N/PT)\mathbf{N}_2\mathbf{X}_2^H$ , where the last matrix has i.i.d  $\mathcal{CN}(0, \sigma^2)$  entries. It follows  $\mathbf{Y}_1 - \mathbf{H}$  and  $\mathbf{Y}_2 - F\mathbf{H}$  are independent random matrices with i.i.d.  $\mathcal{CN}(0, \sigma^2)$  entries.

From the Neyman-Fisher Theorem [45, pg. 117], to prove sufficiency it suffices to show  $p(\mathbf{U}_1, \mathbf{U}_2; F, \sigma_h^2)$  can be factored into a product  $g(\mathbf{Y}_1, \mathbf{Y}_2, F, \sigma_h^2)f(\mathbf{U}_1, \mathbf{U}_2)$ , where  $f$  does not depend on  $\mathbf{Y}_1, \mathbf{Y}_2, F$  or  $\sigma_h^2$ , and  $g$  does not depend on  $\mathbf{U}_1, \mathbf{U}_2$ .

To this end, we can express this pdf in terms of the conditional pdf as

$$p(\mathbf{U}_1, \mathbf{U}_2; F, \sigma_h^2) = E_{\mathbf{H}} \left[ p(\mathbf{U}_1, \mathbf{U}_2 | \mathbf{H}; F, \sigma_h^2) \right],$$

where  $E_{\mathbf{H}}[\cdot]$  denotes expectation with respect to  $\mathbf{H}$ . Since  $\mathbf{U}_1, \mathbf{U}_2$  are conditionally independent given  $\mathbf{H}$ , we can write

$$\begin{aligned} & (\pi\sigma_n^2)^{LT} \cdot p(\mathbf{U}_1, \mathbf{U}_2; F, \sigma_h^2) \\ = & E_{\mathbf{H}} \left[ \exp \left( -\frac{1}{\sigma_n^2} \|\mathbf{U}_1 - \mathbf{H}\mathbf{X}_1\|_F^2 - \frac{1}{\sigma_n^2} \|\mathbf{U}_2 - F\mathbf{H}\mathbf{X}_2\|_F^2 \right) \right] \\ = & E_{\mathbf{H}} \left[ \exp \left( -\frac{1}{\sigma_n^2} \left\{ -\text{Tr}[\mathbf{U}_1^H \mathbf{H}\mathbf{X}_1] - \text{Tr}[(\mathbf{H}\mathbf{X}_1)^H \mathbf{U}_1] + \text{Tr}[\mathbf{X}_1^H \mathbf{H}^H \mathbf{H}\mathbf{X}_1] - \text{Tr}[\mathbf{U}_2^H F\mathbf{H}\mathbf{X}_1] \right. \right. \right. \\ & \left. \left. - \text{Tr}[(F\mathbf{H}\mathbf{X}_2)^H \mathbf{U}_2] + \text{Tr}[|F|^2 \mathbf{X}_2^H \mathbf{H}^H \mathbf{H}\mathbf{X}_2] \right\} \right) \exp \left( -\frac{1}{\sigma_n^2} \{ \|\mathbf{U}_1\|_F^2 + \|\mathbf{U}_2\|_F^2 \} \right) \right] \\ = & E_{\mathbf{H}} \left[ \exp \left( \frac{1}{\sigma_n^2} \left\{ 2\text{ReTr}[\mathbf{H}^H \mathbf{U}_1 \mathbf{X}_1^H] + 2\text{ReTr}[F^* \mathbf{H}^H \mathbf{U}_2 \mathbf{X}_2^H] - \text{Tr}[\mathbf{H}^H \mathbf{H}\mathbf{X}_1 \mathbf{X}_1^H] \right. \right. \right. \\ & \left. \left. - |F|^2 \text{Tr}[\mathbf{H}^H \mathbf{H}\mathbf{X}_2 \mathbf{X}_2^H] \right\} \right) \exp \left( -\frac{1}{\sigma_n^2} \{ \|\mathbf{U}_1\|_F^2 + \|\mathbf{U}_2\|_F^2 \} \right) \right] \\ = & E_{\mathbf{H}} \left[ \exp \left( \frac{1}{\sigma^2} \left\{ 2\text{ReTr}[\mathbf{H}^H \mathbf{Y}_1] + 2\text{ReTr}[F^* \mathbf{H}^H \mathbf{Y}_2] - (1 + |F|^2) \|\mathbf{H}\|_F^2 \right\} \right) \right] \\ & \times \exp \left( -\frac{1}{\sigma_n^2} \{ \|\mathbf{U}_1\|_F^2 + \|\mathbf{U}_2\|_F^2 \} \right), \end{aligned} \tag{3.42}$$

where  $\| \mathbf{A} \|_F^2 = \text{Tr}[\mathbf{A}^H \mathbf{A}]$  denotes the Frobenius norm. Here the third equality follows from the identities  $2\text{ReTr}[\mathbf{A}] = \text{Tr}[\mathbf{A}] + \text{Tr}[\mathbf{A}^H]$  and  $\text{Tr}[\mathbf{A}\mathbf{B}] = \text{Tr}[\mathbf{B}\mathbf{A}]$ , and the fourth equality follows from (3.40) and the definition of  $\sigma^2$ . In (3.42), denote the first factor by  $(\pi\sigma_n^2)^T g(\mathbf{Y}_1, \mathbf{Y}_2, F, \sigma_h^2)$ , and the second by  $f(\mathbf{U}_1, \mathbf{U}_2)$ . Note  $f$  does not depend on  $\mathbf{Y}_1, \mathbf{Y}_2, F$ , or  $\sigma_h^2$ , while  $g$  depends on  $\mathbf{Y}_1, \mathbf{Y}_2, F$  and  $\sigma_h^2$  (through the expectation), but not  $\mathbf{U}_1, \mathbf{U}_2$ . This completes the proof.

We now present maximum-likelihood estimators for the parameter vector (3.15)

$$\boldsymbol{\theta} \triangleq \begin{bmatrix} F \\ \sigma_h^2 \end{bmatrix}, \quad (3.43)$$

based on the sufficient statistics (3.41), where  $F$  is defined in (3.11) and  $\sigma_h^2$  in (3.9). This estimate is defined by

$$\boldsymbol{\theta}_{ML} \triangleq \arg \max_{\boldsymbol{\theta}} p(\mathbf{Y}_1, \mathbf{Y}_2; \boldsymbol{\theta}). \quad (3.44)$$

The following theorem shows that these estimators can be calculated from the result  $\hat{\mu}$  of a scalar optimization.

**Theorem 6 (Multiple-Packet ML Estimators) :** *Let  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  be the sufficient statistics in (3.41), where  $F$  and  $\sigma_h^2$  are unknown constants. Consider the matrix*

$$\mathbf{S}(\mu) \triangleq \begin{bmatrix} S_{11}(\mu) & S_{12}(\mu) \\ S_{21}(\mu) & S_{22}(\mu) \end{bmatrix}. \quad (3.45)$$

where  $S_{ij}(\mu) \triangleq \text{Tr} \left[ \mu \mathbf{C}_{\mathbf{H}} (\mu \mathbf{C}_{\mathbf{H}} + \sigma^2 \mathbf{I}_L)^{-1} \mathbf{Y}_i \mathbf{Y}_j^H \right]$ ,  $1 \leq i, j \leq 2$ . Define

$$\hat{\mu} \triangleq \arg \max_{\mu \geq 0} [\eta(\mu) - \sigma^2 \ln \det[\mu \mathbf{C}_{\mathbf{H}} + \sigma^2 \mathbf{I}_L]] , \quad (3.46)$$

where  $\eta(\mu)$  is the largest eigenvalue of  $\mathbf{S}(\mu)$ :

$$\eta(\mu) \triangleq \frac{S_{22}(\mu) + S_{11}(\mu) + \sqrt{(S_{11}(\mu) - S_{22}(\mu))^2 + 4|S_{12}(\mu)|^2}}{2} . \quad (3.47)$$

Let  $\hat{\mathbf{e}}_1 = [E_1, E_2]^T$  be any unit eigenvector of  $\mathbf{S}(\hat{\mu})$  corresponding to the eigenvalue  $\eta(\hat{\mu})$ . Then the maximum-likelihood estimates of  $F$  and  $\sigma_h^2$  are given by

$$\hat{\boldsymbol{\theta}}_{ML} = \begin{bmatrix} \hat{F}_{ML} \\ \hat{\sigma}_h^2 \end{bmatrix} = \begin{bmatrix} E_2/E_1 \\ |E_1|^2 \hat{\mu} \end{bmatrix} , \quad (3.48)$$

provided  $E_1 \neq 0$ , where  $\sigma^2 \triangleq \sigma_n^2(2N/PT)$ . For  $E_1 = 0$  and  $\hat{\mu} > 0$  the likelihood is maximized in the limit as  $F \rightarrow \infty$ .  $\diamond$

**Proof :** For any matrix  $\mathbf{A}$ , denote the  $kj$ -th element and  $k$ -th row by  $[\mathbf{A}]_{kj}$  and  $[\mathbf{A}]_k$ , respectively. Let  $\mathbf{C}_{\mathbf{H}} = \mathbf{V}^H \text{diag}[\lambda_1, \dots, \lambda_L] \mathbf{V}$  be an eigendecomposition of  $\mathbf{C}_{\mathbf{H}}$ , where  $\lambda_1 \geq \dots \geq \lambda_L \geq 0$  are eigenvalues of  $\mathbf{C}_{\mathbf{H}}$ , and  $\mathbf{V}$  is a unitary matrix such that  $\mathbf{V} \mathbf{V}^H = \mathbf{V}^H \mathbf{V} = \mathbf{I}_L$ . It follows the elements of  $\mathbf{V} \mathbf{H}$  are independent with  $[\mathbf{V} \mathbf{H}]_{kj} \sim \mathcal{CN}(0, \sigma_h^2 \lambda_k)$ .

For  $1 \leq k \leq L$ , let  $\mathbf{w}_{k1} \triangleq [\mathbf{V} \mathbf{Y}_1]_k$  and  $\mathbf{w}_{k2} \triangleq [\mathbf{V} \mathbf{Y}_2]_k$ . From Lemma 3,  $\mathbf{w}_{k1}$  and  $\mathbf{w}_{k2}$  are conditionally independent given  $[\mathbf{V} \mathbf{H}]_k$ , with conditional distributions  $\mathbf{w}_{k1} \sim \mathcal{CN}([\mathbf{V} \mathbf{H}]_k, \sigma^2 \mathbf{I}_N)$  and  $\mathbf{w}_{k2} \sim \mathcal{CN}(F[\mathbf{V} \mathbf{H}]_k, \sigma^2 \mathbf{I}_L)$ . Since  $[\mathbf{V} \mathbf{H}]_k \sim \mathcal{CN}(\mathbf{0}_N, \sigma_h^2 \lambda_k \mathbf{I}_N)$  is independent of the noise in (3.41), it follows  $\mathbf{w}_k \triangleq (\mathbf{w}_{k1}, \mathbf{w}_{k2})^T$  is a zero-mean Gaussian

random vector with covariance

$$\mathbf{C}_{\mathbf{w}_k} \triangleq E[\mathbf{w}_k^H \mathbf{w}_k] = \begin{bmatrix} (\sigma_h^2 \lambda_k + \sigma^2) \mathbf{I}_N & \sigma_h^2 F^* \lambda_k \mathbf{I}_N \\ \sigma_h^2 F \lambda_k \mathbf{I}_N & (\sigma_h^2 |F|^2 \lambda_k + \sigma^2) \mathbf{I}_N \end{bmatrix} = \mathbf{C}_k \otimes \mathbf{I}_N, \quad (3.49)$$

where

$$\mathbf{C}_k \triangleq \begin{bmatrix} \sigma_h^2 \lambda_k + \sigma^2 & \sigma_h^2 F^* \lambda_k \\ \sigma_h^2 F \lambda_k & \sigma_h^2 |F|^2 \lambda_k + \sigma^2 \end{bmatrix}. \quad (3.50)$$

As in the derivation of (3.23) in the proof of Theorem 5, note  $\mathbf{C}_k$  can be written in terms of its eigensystem as

$$\mathbf{C}_k = \mu_{k1} \mathbf{e}_1 \mathbf{e}_1^H + \mu_2 \mathbf{e}_2 \mathbf{e}_2^H, \quad (3.51)$$

where  $\mu_{k1} \geq \mu_2$  are the ordered eigenvalues and  $\mathbf{e}_1, \mathbf{e}_2$  are the associated unit eigenvectors. From (3.50), it is easy to verify the following explicit formulas,

$$\begin{aligned} \mu_{k1} &= \mu \lambda_k + \sigma^2, & \mu_2 &= \sigma^2 \\ \mathbf{e}_1 &= \frac{1}{\sqrt{1 + |F|^2}} \begin{bmatrix} 1 \\ F \end{bmatrix}, & \mathbf{e}_2 &= \frac{1}{\sqrt{1 + |F|^2}} \begin{bmatrix} -F^* \\ 1 \end{bmatrix}. \end{aligned} \quad (3.52)$$

where  $\mu \triangleq \sigma_h^2(1 + |F|^2)$ . Note only  $\mu_{k1}$  depends on  $k$ . As in the derivation of (3.25) in

the proof of Theorem 5, we have

$$\begin{aligned}
\ln p(\mathbf{w}_{k1}, \mathbf{w}_{k2}; \boldsymbol{\theta}) &= -N \ln(\pi \mu_{k1} \mu_2) + (\mu_2^{-1} - \mu_{k1}^{-1}) N \mathbf{e}_1^H \mathbf{S}_k \mathbf{e}_1 - \mu_2^{-1} N \text{Tr}[\mathbf{S}_k] \\
&= -N \ln(\pi \sigma^2 [\mu \lambda_k + \sigma^2]) + \frac{N \mu \lambda_k}{\sigma^2 [\mu \lambda_k + \sigma^2]} \mathbf{e}_1^H \mathbf{S}_k \mathbf{e}_1 - \sigma^{-2} N \text{Tr}[\mathbf{S}_k] \\
&= B_k + \frac{N}{\sigma^2} \left[ \frac{\mu \lambda_k}{\mu \lambda_k + \sigma^2} \mathbf{e}_1^H \mathbf{S}_k \mathbf{e}_1 - \sigma^2 \ln(\mu \lambda_k + \sigma^2) \right] \tag{3.53}
\end{aligned}$$

where  $B_k$  does not depend on  $\mu$  or  $\mathbf{e}_1$  and

$$\mathbf{S}_k \triangleq \frac{1}{N} \begin{bmatrix} \mathbf{w}_{k1} \mathbf{w}_{k1}^H & \mathbf{w}_{k1} \mathbf{w}_{k2}^H \\ \mathbf{w}_{k2} \mathbf{w}_{k1}^H & \mathbf{w}_{k2} \mathbf{w}_{k2}^H \end{bmatrix} = \frac{1}{N} \begin{bmatrix} [\mathbf{V} \mathbf{Y}_1 \mathbf{Y}_1^H \mathbf{V}^H]_{kk} & [\mathbf{V} \mathbf{Y}_1 \mathbf{Y}_2^H \mathbf{V}^H]_{kk} \\ [\mathbf{V} \mathbf{Y}_2 \mathbf{Y}_1^H \mathbf{V}^H]_{kk} & [\mathbf{V} \mathbf{Y}_2 \mathbf{Y}_2^H \mathbf{V}^H]_{kk} \end{bmatrix}. \tag{3.54}$$

Since  $\mathbf{w}_1, \dots, \mathbf{w}_L$  are independent, the joint probability of  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  is then given by

$$\begin{aligned}
\ln p(\mathbf{Y}_1, \mathbf{Y}_2; \boldsymbol{\theta}) &= \sum_{k=1}^L \ln p(\mathbf{w}_{k1}, \mathbf{w}_{k2}; \boldsymbol{\theta}) \\
&= B + \frac{N}{\sigma^2} \sum_{k=1}^L \left[ \frac{\mu \lambda_k}{\mu \lambda_k + \sigma^2} \mathbf{e}_1^H \mathbf{S}_k \mathbf{e}_1 - \ln(\mu \lambda_k + \sigma^2) \right] \\
&= B + \frac{N}{\sigma^2} \left[ \mathbf{e}_1^H \mathbf{S}(\mu) \mathbf{e}_1 - \sigma^2 \sum_{k=1}^L \ln(\mu \lambda_k + \sigma^2) \right], \tag{3.55}
\end{aligned}$$

where  $B$  does not depend on the parameters and

$$\mathbf{S}(\mu) \triangleq \sum_{k=1}^L \frac{\mu \lambda_k}{\mu \lambda_k + \sigma^2} \mathbf{S}_k \tag{3.56}$$

is the matrix in (3.45). To see this, let  $\Lambda \triangleq \text{diag}(\lambda_1, \dots, \lambda_L)$  and observe

$$\begin{aligned}
[\mathbf{S}(\mu)]_{ij} &\triangleq \sum_{k=1}^L \frac{\mu\lambda_k}{\mu\lambda_k + \sigma^2} [\mathbf{V}\mathbf{Y}_i\mathbf{Y}_j^H\mathbf{V}^H]_{kk} \\
&= \sum_{k=1}^L \left[ \mu\Lambda (\mu\Lambda + \sigma^2\mathbf{I}_L)^{-1} \mathbf{V}\mathbf{Y}_i\mathbf{Y}_j^H\mathbf{V}^H \right]_{kk} \\
&= \sum_{k=1}^L \left[ \mu\mathbf{C}_H (\mu\mathbf{C}_H + \sigma^2\mathbf{I}_L)^{-1} \mathbf{Y}_i\mathbf{Y}_j^H \right]_{kk} \\
&= \text{Tr} \left[ \mu\mathbf{C}_H (\mu\mathbf{C}_H + \sigma^2\mathbf{I}_L)^{-1} \mathbf{Y}_i\mathbf{Y}_j^H \right]. \tag{3.57}
\end{aligned}$$

To find maximum-likelihood estimates of  $F$  and  $\sigma_h^2$ , we proceed in two steps: First, we find conditions on  $\mu$  and  $\mathbf{e}_1$  that achieve the maximum in (3.55). Second, we use (3.52) to translate these conditions into values of  $F$  and  $\sigma_h^2$ .

For each  $\mu$ , the maximum of (3.55) over  $\mathbf{e}_1$  is clearly a unit eigenvector corresponding to the largest eigenvalue of  $\mathbf{S}(\mu)$ . As in the proof of (3.31), it is easily shown this eigenvalue is given by  $\eta(\mu)$ , defined in (3.47). It follows that the maximum-likelihood estimate of  $\mu$  is

$$\hat{\mu} \triangleq \arg \max_{\mu \geq 0} \left[ \eta(\mu) - \sigma^2 \sum_{k=1}^L \ln(\mu\lambda_k + \sigma^2) \right],$$

which equals (3.46), since  $\sum_{k=1}^L \ln(\mu\lambda_k + \sigma^2) = \ln \det[\mu\mathbf{C}_H + \sigma^2\mathbf{I}_L]$ .

Finally, we translate these conditions into values of  $F$  and  $\sigma_h^2$ : If  $\hat{\mu} = 0$ ,  $\mathbf{S}(\hat{\mu})$  vanishes and  $\ln p(\mathbf{Y}_1, \mathbf{Y}_2; \boldsymbol{\theta})$  does not depend on  $F$ . From (3.52), it follows the likelihood is maximized by  $\hat{\sigma}_h^2 = 0$  and any value of  $F$ ; In particular, (3.48) maximizes the likelihood. However, if  $\hat{\mu} > 0$ , then  $\mathbf{S}(\hat{\mu})$  is not zero and  $\mathbf{e}_1$  must be an eigenvector of  $\mathbf{S}(\hat{\mu})$  corresponding to  $\eta(\hat{\mu}) > 0$ . For  $E_1 \neq 0$ , the unique solution of the equations  $\hat{\mu} = \sigma_h^2(1 + |F|^2)$

and  $\mathbf{e}_1 = (E_1, E_2)^T$  in (3.52) is given by (3.48). For  $E_1 = 0$ , no finite  $F$  solves these equations; rather, the solution (and maximum) is approached in the limit as  $F \rightarrow \infty$ .  $\diamond$

The theorem above reduces the problem of calculating the multi-packet estimators to the problem of solving the scalar optimization (3.46). In general, it appears that this optimization must be performed numerically. However, we now show this optimization admits a simple, closed form solution in several scenarios of practical interest.

First consider the case of fast-fading, where  $\mathbf{C}_\mathbf{H} = \mathbf{I}_L$ . In this case, the estimators can be expressed in a simple form similar to those in Corollary 1.

**Corollary 2 (ML Estimators for Fast-Fading)** : *Let  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  be the sufficient statistics in (3.41), where  $F$  and  $\sigma_h^2$  are unknown constants. Consider the matrix*

$$\mathbf{T} \triangleq \frac{1}{L} \begin{bmatrix} \text{Tr} [\mathbf{Y}_1 \mathbf{Y}_1^H] & \text{Tr} [\mathbf{Y}_1 \mathbf{Y}_2^H] \\ \text{Tr} [\mathbf{Y}_2 \mathbf{Y}_1^H] & \text{Tr} [\mathbf{Y}_2 \mathbf{Y}_2^H] \end{bmatrix}. \quad (3.58)$$

If  $\mathbf{C}_\mathbf{H} = \mathbf{I}_L$ , the ML estimators (3.48) can be expressed in closed-form as

$$\hat{F}_{ML} = \frac{T_{22} - T_{11} + \sqrt{(T_{22} - T_{11})^2 + 4|T_{12}|^2}}{2T_{12}}, \quad (3.59)$$

$$\hat{\sigma}_h^2 = \frac{|\hat{F}_{ML}|^2}{|\hat{F}_{ML}|^2 + 1} \max \left\{ T_{11} + T_{12} \hat{F}_{ML} - \sigma^2, 0 \right\}, \quad (3.60)$$

provided  $T_{12} \neq 0$ .  $\diamond$

**Proof** : In Theorem 6, the ML estimators are given in terms of  $\hat{\mu}$  and  $\hat{\mathbf{e}}_1$ , which jointly maximize the function

$$f(\mu, \mathbf{e}_1; \sigma^2) \triangleq \mathbf{e}_1^H \mathbf{S}(\mu) \mathbf{e}_1 - \sigma^2 \sum_{k=1}^L \ln(\mu \lambda_k + \sigma^2). \quad (3.61)$$

For  $\mathbf{C}_H = \mathbf{I}_L$ , we have  $\lambda_1 = \cdots = \lambda_L = 1$ , so

$$\mathbf{S}(\mu) = \frac{L\mu}{\mu + \sigma^2} \mathbf{T}, \quad (3.62)$$

and  $f$  reduces to

$$f(\mu, \mathbf{e}_1; \sigma^2) = L \left[ \frac{\mu}{\mu + \sigma^2} \mathbf{e}_1^H \mathbf{T} \mathbf{e}_1 - \sigma^2 \ln(\mu + \sigma^2) \right].$$

The maximum of  $\mathbf{e}_1^H \mathbf{T} \mathbf{e}_1$  over  $\mathbf{e}_1$  is clearly the largest eigenvalue of  $\mathbf{T}$ , say  $\eta_1$ , and is achieved when  $\mathbf{e}_1$  is any associated eigenvector. By direct differentiation, we observe

$$\frac{\mu\eta_1}{\mu + \sigma^2} - \sigma^2 \ln(\mu + \sigma^2),$$

is maximized by

$$\hat{\mu} = \max \{ \eta_1 - \sigma^2, 0 \}. \quad (3.63)$$

Similar to the proofs of (3.31) and (3.32) in Corollary 1, we can derive closed-form formulas for  $\eta_1$  and  $\mathbf{e}_1$

$$\eta_1 = \frac{T_{22} + T_{11} + \sqrt{(T_{11} - T_{22})^2 + 4|T_{12}|^2}}{2} \quad (3.64)$$

$$\hat{\mathbf{e}}_1 = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} = \frac{1}{\sqrt{|\eta_1 - T_{11}|^2 + |T_{12}|^2}} \begin{bmatrix} T_{12} \\ \eta_1 - T_{11} \end{bmatrix}. \quad (3.65)$$

If  $T_{12} \neq 0$ , substituting  $\hat{\mu}$  and  $\mathbf{e}_1$  into (3.48) yields

$$\hat{F}_{ML} = \frac{\eta_1 - T_{11}}{T_{12}}, \quad \hat{\sigma}_h^2 = \frac{|\eta_1 - T_{11}|^2}{|\eta_1 - T_{11}|^2 + |T_{12}|^2} \max \{ \eta_1 - \sigma^2, 0 \}, \quad (3.66)$$

Substituting (3.64) into the first equation gives (3.59); substituting  $\eta_1 = T_{11} + T_{12}\hat{F}_{ML}$  into the second gives (3.60).

For a general non-singular temporal correlation matrix  $\mathbf{C}_{\mathbf{H}}$ , the ML estimators (3.48) can easily be approximated under high SNR conditions, such that  $\sigma^2 \approx 0$ . In Theorem 6, the ML estimators are given in terms of  $\hat{\mu}$  and  $\hat{\mathbf{e}}_1$ , which jointly maximize the function  $f(\mu, \mathbf{e}_1; \sigma^2)$  in (3.61). If  $\mathbf{C}_{\mathbf{H}}$  is any non-singular matrix, For small  $\sigma^2$ , we can expand this function as

$$\begin{aligned} f(\mu, \mathbf{e}_1; \sigma^2) &= \sum_{k=1}^L \left[ \frac{\mu\lambda_k}{\mu\lambda_k + \sigma^2} \mathbf{e}_1^H \mathbf{S}_k \mathbf{e}_1 - \sigma^2 \ln(\mu\lambda_k + \sigma^2) \right] \\ &= \sum_{k=1}^L \left[ \left(1 - \frac{\sigma^2}{\mu\lambda_k}\right) \mathbf{e}_1^H \mathbf{S}_k \mathbf{e}_1 - \sigma^2 \ln(\mu\lambda_k) \right] + \mathcal{O}(\sigma^4) \\ &= Lg(\mu, \mathbf{e}_1; \sigma^2) - \sigma^2 \ln \det[\mathbf{C}_{\mathbf{H}}] + \mathcal{O}(\sigma^4), \end{aligned} \quad (3.67)$$

where

$$g(\mu, \mathbf{e}; \sigma^2) \triangleq \mathbf{e}^H \left[ \mathbf{T} - \frac{\sigma^2}{\mu} \mathbf{S} \right] \mathbf{e} - \sigma^2 \ln(\mu), \quad (3.68)$$

where  $\mathbf{T}$  is given in (3.58) and

$$\mathbf{S} \triangleq \frac{1}{L} \begin{bmatrix} \text{Tr} [\mathbf{C}_{\mathbf{H}}^{-1} \mathbf{Y}_1 \mathbf{Y}_1^H] & \text{Tr} [\mathbf{C}_{\mathbf{H}}^{-1} \mathbf{Y}_1 \mathbf{Y}_2^H] \\ \text{Tr} [\mathbf{C}_{\mathbf{H}}^{-1} \mathbf{Y}_2 \mathbf{Y}_1^H] & \text{Tr} [\mathbf{C}_{\mathbf{H}}^{-1} \mathbf{Y}_2 \mathbf{Y}_2^H] \end{bmatrix}. \quad (3.69)$$

We define the approximate ‘‘low-noise ML estimators’’ to be the estimators (3.66) with  $\hat{\mu}$  and  $\hat{\mathbf{e}}_1$  replaced by the  $\mathbf{e}, \mu$  that jointly maximize  $g$ . The following result provides a simple characterization of these estimators.

**Theorem 7 (Low-Noise ML Estimators)** : Let  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  be the sufficient statistics in (3.41), where  $F$  and  $\sigma_h^2$  are unknown constants. Suppose  $\mu_o > 0$  and  $\mathbf{e}_o \in \mathbb{C}^2$  satisfy the following two conditions: (1)  $\mathbf{e}_o$  is a unit eigenvector of  $\mathbf{P} \triangleq \mathbf{T} - (\sigma^2/\mu_o)\mathbf{S}$  corresponding to its largest eigenvalue, and (2)  $\mu_o = \mathbf{e}_o^H \mathbf{S} \mathbf{e}_o$ . Then  $\mu_o, \mathbf{e}_o$  are the global maxima of  $g(\mu, \mathbf{e}; \sigma^2)$ , and the low-noise ML estimators are given by

$$\hat{F}_{ML} = \frac{P_{22} - P_{11} + \sqrt{(P_{22} - P_{11})^2 + 4|P_{12}|^2}}{2P_{12}}, \quad (3.70)$$

$$\hat{\sigma}_h^2 = \frac{|\hat{F}_{ML}|^2}{|\hat{F}_{ML}|^2 + 1} \max \left\{ P_{11} + P_{12} \hat{F}_{ML} - \sigma^2, 0 \right\}, \quad (3.71)$$

provided  $P_{12} \neq 0$ . Moreover, the value of  $\mu_o$  is unique.  $\diamond$

**Proof :** If  $\mu$  and  $\mathbf{e}$  maximize  $g(\mu, \mathbf{e}; \sigma^2)$  subject to the constraint  $\mathbf{e}^H \mathbf{e} = 1$ , they must be a critical point of the Lagrangian  $L(\mu, \mathbf{e}; \sigma^2) = g(\mu, \mathbf{e}; \sigma^2) - \eta \mathbf{e}^H \mathbf{e}$  for some real  $\eta$ . It follows  $\mu$  and  $\mathbf{e}$  satisfy

$$\begin{aligned} \mathbf{0} = \nabla_{\mathbf{e}} L &= 2\mathbf{e}^H [\mathbf{T} - (\sigma^2/\mu)\mathbf{S}] - 2\eta \mathbf{e}^H \\ 0 = \frac{\partial L}{\partial \mu} &= -\frac{\sigma^2}{\mu^2} \mathbf{e}^H \mathbf{S} \mathbf{e} + \frac{\sigma^2}{\mu}, \end{aligned} \quad (3.72)$$

for some  $\eta$ . For a given  $\mu$ , the first equation is satisfied only when  $\eta$  is an eigenvalue of  $\mathbf{T} - (\sigma^2/\mu)\mathbf{S}$  and  $\mathbf{e}$  is a corresponding unit eigenvector. From (3.68), it is  $\eta$  must be the largest eigenvalue of  $\mathbf{T} - (\sigma^2/\mu)\mathbf{S}$ , since otherwise  $\mathbf{e}$  would not achieve  $\max_{\mathbf{e}: \mathbf{e}^H \mathbf{e} = 1} g(\mathbf{e}, \mu; \sigma^2)$ . Similarly, for a given  $\mathbf{e}$ , the unique solution of the second equation is  $\mu = \mathbf{e}^H \mathbf{S} \mathbf{e}$ .

We now prove  $\mu_o, \mathbf{e}_o$  is the global maxima of  $g(\mathbf{e}, \mu; \sigma^2)$ . For any  $\mu > 0$ , let  $\eta_\mu$  denote the largest eigenvalue of  $\mathbf{T} - \frac{\sigma^2}{\mu} \mathbf{S}$ , and  $\mathbf{e}_\mu$  be any associated unit eigenvector. We claim

$\mu > \mu'$  implies

$$\mathbf{e}_\mu^H \mathbf{S} \mathbf{e}_\mu \geq \mathbf{e}_{\mu'}^H \mathbf{S} \mathbf{e}_{\mu'} . \quad (3.73)$$

To see this, note the Rayleigh-Ritz Theorem implies

$$\eta_{\mu'} = \mathbf{e}_{\mu'}^H \left[ \mathbf{T} - \frac{\sigma^2}{\mu'} \mathbf{S} \right] \mathbf{e}_{\mu'} \geq \mathbf{e}_\mu^H \left[ \mathbf{T} - \frac{\sigma^2}{\mu'} \mathbf{S} \right] \mathbf{e}_\mu .$$

Since

$$\mathbf{T} - \frac{\sigma^2}{\mu'} \mathbf{S} = \mathbf{T} - \frac{\sigma^2}{\mu} \mathbf{S} + \sigma^2 \left( \frac{1}{\mu} - \frac{1}{\mu'} \right) \mathbf{S} , \quad (3.74)$$

it follows

$$\begin{aligned} & \mathbf{e}_{\mu'}^H \left[ \mathbf{T} - \frac{\sigma^2}{\mu} \mathbf{S} \right] \mathbf{e}_{\mu'} + \sigma^2 \left( \frac{1}{\mu} - \frac{1}{\mu'} \right) \mathbf{e}_{\mu'}^H \mathbf{S} \mathbf{e}_{\mu'} \\ \geq & \mathbf{e}_\mu^H \left[ \mathbf{T} - \frac{\sigma^2}{\mu} \mathbf{S} \right] \mathbf{e}_\mu + \sigma^2 \left( \frac{1}{\mu} - \frac{1}{\mu'} \right) \mathbf{e}_\mu^H \mathbf{S} \mathbf{e}_\mu \\ = & \eta(\mu) + \sigma^2 \left( \frac{1}{\mu} - \frac{1}{\mu'} \right) \mathbf{e}_\mu^H \mathbf{S} \mathbf{e}_\mu , \end{aligned} \quad (3.75)$$

and hence from the Rayleigh-Ritz Theorem

$$\sigma^2 \left( \frac{1}{\mu'} - \frac{1}{\mu} \right) [\mathbf{e}_{\mu'}^H \mathbf{S} \mathbf{e}_{\mu'} - \mathbf{e}_\mu^H \mathbf{S} \mathbf{e}_\mu] \geq \eta(\mu) - \mathbf{e}_{\mu'}^H \left[ \mathbf{T} - \frac{\sigma^2}{\mu} \mathbf{S} \right] \mathbf{e}_{\mu'} \geq 0 , \quad (3.76)$$

thereby proving the claim.

Now let  $\mathbf{e}_o, \mu_o$  satisfy the necessary conditions. For all  $\mu > 0$

$$\begin{aligned}
\eta(\mu) &= \mathbf{e}_\mu^H \left[ \mathbf{T} - \frac{\sigma^2}{\mu} \mathbf{S} \right] \mathbf{e}_\mu \\
&= \mathbf{e}_\mu^H \left[ \mathbf{T} - \frac{\sigma^2}{\mu_o} \mathbf{S} \right] \mathbf{e}_\mu + \sigma^2 \left( \frac{1}{\mu_o} - \frac{1}{\mu} \right) \mathbf{e}_\mu^H \mathbf{S} \mathbf{e}_\mu \\
&\leq \eta(\mu_o) + \sigma^2 \left( \frac{1}{\mu_o} - \frac{1}{\mu} \right) \mathbf{e}_\mu^H \mathbf{S} \mathbf{e}_\mu \\
&= \eta(\mu_o) + \sigma^2 \left( 1 - \frac{\mu_o}{\mu} \right) \theta, \tag{3.77}
\end{aligned}$$

where the first inequality follows from the Rayleigh-Ritz Theorem, and the second from (3.73). From (3.73),  $\theta \geq 1$  if  $\mu \geq \mu_o$ .

Similarly, this inequality also holds for  $\mu < \mu_o$ , since the coefficient of  $\mathbf{e}_\mu^H \mathbf{S} \mathbf{e}_\mu$  above is then negative. Thus, for all  $\mathbf{e}$  and  $\mu$ , we have

$$\begin{aligned}
g(\mathbf{e}, \mu; \sigma^2) &\triangleq \mathbf{e}^H \left[ \mathbf{T} - \frac{\sigma^2}{\mu} \mathbf{S} \right] \mathbf{e} - \sigma^2 \ln(\mu) \\
&\leq \eta_\mu - \sigma^2 \ln(\mu) \\
&\leq \eta_{\mu_o} + \sigma^2 \left( 1 - \frac{\mu_o}{\mu} \right) - \sigma^2 \ln(\mu) \\
&= g(\mathbf{e}_o, \mu_o; \sigma^2) + \sigma^2 \left[ \left( 1 - \frac{\mu_o}{\mu} \right) + \ln \left( \frac{\mu_o}{\mu} \right) \right]. \tag{3.78}
\end{aligned}$$

Recall  $\ln(1+x) \leq x$  for all  $x > -1$ , with equality if and only if  $x = 1$ . For  $x = \mu_o/\mu - 1$ , this implies the bracketed term is negative for all  $\mu \neq \mu_o$ . Thus,  $\mathbf{e}_o, \mu_o$  achieves the global maximum of  $g$  and the value of  $\mu_o$  is unique.  $\diamond$

Temporal correlation generally exists in Rayleigh fading channels, e.g., Clarke's model [9, 21, 86]. To illustrate this correlation in time, we vectorize  $\mathbf{H} \in \mathbb{C}^{L \times N}$  defined in (3.39),

$$\mathbf{h} \triangleq \text{vec } \mathbf{H} \sim \mathcal{CN}(\mathbf{0}_M, \sigma_h^2 \mathbf{I}_N \otimes \mathbf{C}_h), \tag{3.79}$$

where  $\mathbf{C}_h$  is the normalized temporal correlation matrix, such that its diagonal entries are all 1's and other entries are scaled accordingly. If  $\mathbf{Q}$  is known a priori, the channel  $\mathbf{H}$  can be decorrelated in time to generate an independent<sup>2</sup> channel matrix  $\mathbf{H}_d$ ,

$$\mathbf{H}_d \triangleq \mathbf{Q}^H \mathbf{H}, \quad \mathbf{h}_d \triangleq \text{vec } \mathbf{H}_d \sim \mathcal{CN}(\mathbf{0}_M, \sigma_h^2 \mathbf{I}_N \otimes \mathbf{D}), \quad (3.80)$$

where unitary matrix  $\mathbf{Q}$  and diagonal matrix  $\mathbf{D}$  are from the eigensystem of the channel correlation matrix  $\mathbf{C}_h$ , i.e.,

$$\mathbf{C}_h \mathbf{Q} = \mathbf{Q} \mathbf{D}, \quad \mathbf{D} = \text{diag}(d_1, \dots, d_L), \quad (3.81)$$

and  $\text{Tr}[\mathbf{D}] = L$ . Note  $\mathbf{h}_d$  in (3.80) is temporally independent but not necessarily identical, since  $d_k$ 's may differ.

Since  $\mathbf{C}_h$  is Hermitian, its eigen-values are real. We further assume that  $\mathbf{C}_h$  is positive definite, so  $d_k > 0$  for  $1 \leq k \leq L$ . Without loss of generality, we sort  $d_k$ 's in descending order in  $\mathbf{D}$ . We assume  $d_k$ 's are known, because second-order statistics of the fading channel remain the same except an unknown scaling factor in  $\sigma_h^2$ . Thus, the orthonormal eigen-vectors of  $\mathbf{C}_h$  remain unchanged for all passive antenna impedance in  $Z_A$ . We define  $\mathbf{\Lambda}$  for mathematical convenience,

$$\mathbf{\Lambda} \triangleq \mathbf{D} \otimes \sigma_h^2 \mathbf{I}_N = \text{diag}(\lambda_1, \dots, \lambda_M), \quad (3.82)$$

where  $\sigma_h^2 \in \mathbb{R}^+$  is unknown and we define,

$$M \triangleq N \cdot L. \quad (3.83)$$

---

<sup>2</sup>Uncorrelated Gaussian random variables are also independent [45, pg. 558].

Note, due to the Kronecker product, there are only  $L$  distinct diagonal elements in  $\mathbf{\Lambda}$ , each with multiplicity  $N$ .

We combine the vectorized sufficient statistics in (3.41) at the load into a matrix in  $\mathbb{C}^{2 \times M}$ ,

$$\mathbf{V}_s \triangleq \begin{bmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \end{bmatrix} = \begin{bmatrix} \mathbf{h}^T \\ F\mathbf{h}^T \end{bmatrix} + \mathbf{N}_s, \quad (3.84)$$

where we define vectorized sufficient statistics (3.41),

$$\mathbf{y}_1 \triangleq \text{vec } \mathbf{Y}_1 = \mathbf{h} + \mathbf{n}_{s,1}, \quad \mathbf{y}_2 \triangleq \text{vec } \mathbf{Y}_2 = F\mathbf{h} + \mathbf{n}_{s,2}, \quad (3.85)$$

and the noise is i.i.d., such that  $\text{vec } \mathbf{N}_s \sim \mathcal{CN}(\mathbf{0}_M, \sigma_n^2 \mathbf{I}_{2M})$ .

Consider the decorrelated observation (3.84) as

$$\mathbf{V} \triangleq [\mathbf{v}_1, \dots, \mathbf{v}_M] = \mathbf{V}_s (\mathbf{I}_N \otimes \mathbf{Q}^*) = \begin{bmatrix} \mathbf{h}_d^T \\ F\mathbf{h}_d^T \end{bmatrix} + \mathbf{N}. \quad (3.86)$$

It follows that  $\text{vec } \mathbf{N} = [(\mathbf{I}_N \otimes \mathbf{Q}^H) \otimes \mathbf{I}_2] \cdot \text{vec } \mathbf{N}_s \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{2M})$  [19, Th. 2.13]. With a slight abuse of notation, we define the sample covariance using (3.84) and (3.86),

$$\mathbf{S} \triangleq \frac{1}{M} \sum_{k=1}^M \mathbf{v}_k \mathbf{v}_k^H = \frac{1}{M} \mathbf{V} \mathbf{V}^H = \frac{1}{M} \mathbf{V}_s \mathbf{V}_s^H \triangleq \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}. \quad (3.87)$$

where the third equality is achieved by noting  $\mathbf{Q}$  is unitary (3.81) and  $\mathbf{Q}^* \mathbf{Q}^T = \mathbf{I}_L$ . One observes that  $\mathbf{S}$  remains the same regardless decorrelation or not. We write the density

function of (3.86),

$$p(\mathbf{V}; \boldsymbol{\theta}) = \prod_{m=1}^M \det(\pi \mathbf{C}_m)^{-1} \exp \left( - \sum_{m=1}^M \mathbf{v}_m^H \mathbf{C}_m^{-1} \mathbf{v}_m \right), \quad (3.88)$$

where the covariance matrices have the following pattern, for  $1 \leq k \leq M$ ,

$$\begin{aligned} \mathbf{C}_m(\boldsymbol{\theta}) &\triangleq E[\mathbf{v}_m \mathbf{v}_m^H] = \begin{bmatrix} \lambda_m + \sigma^2 & \lambda_m F^* \\ \lambda_m F & \lambda_m |F|^2 + \sigma^2 \end{bmatrix} \\ &= \lambda_m \begin{bmatrix} 1 \\ F \end{bmatrix} \begin{bmatrix} 1 & F^* \end{bmatrix} + \sigma^2 \mathbf{I}_2 = \mu_1 \mathbf{e}_1 \mathbf{e}_1^H + \mu_2 \mathbf{e}_2 \mathbf{e}_2^H, \end{aligned} \quad (3.89)$$

with  $\lambda_m$ 's defined in (3.82), and the eigen-vectors and eigenvalues are, respectively,

$$\begin{aligned} \mathbf{e}_1 &= \frac{1}{\sqrt{1 + |F|^2}} \begin{bmatrix} 1 \\ F \end{bmatrix}, \quad \mathbf{e}_2 = \frac{1}{\sqrt{1 + |F|^2}} \begin{bmatrix} -F^* \\ 1 \end{bmatrix}, \\ \mu_{1,m}(\boldsymbol{\theta}) &= \lambda_m(1 + |F|^2) + \sigma^2, \quad \mu_2 = \sigma^2. \end{aligned} \quad (3.90)$$

We observe that the eigen-vectors of  $\mathbf{C}_m$  are the same, regardless the value of  $\sigma_h^2 \in \mathbb{R}^+$ ; so is its smallest eigenvalue. We find the the log-likelihood function as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &\triangleq \ln p(\boldsymbol{\theta}; \mathbf{V}) = \ln \left[ \prod_{m=1}^M \det(\pi \mathbf{C}_m)^{-1} \exp \left( - \sum_{m=1}^M \mathbf{v}_m^H \mathbf{C}_m^{-1} \mathbf{v}_m \right) \right] \\ &= - \sum_{m=1}^M \ln \det \mathbf{C}_m - \sum_{m=1}^M \mathbf{v}_m^H \mathbf{C}_m^{-1} \mathbf{v}_m + C, \end{aligned} \quad (3.91)$$

where  $C$  is a constant independent of  $\boldsymbol{\theta}$ . More details on this can be found in Appendix B.2. The entries of Fisher information matrix (FIM) have been derived, for  $1 \leq i, j \leq 2$ ,

using [45, pg. 529] and extension of (15.60) in Kay [45, pg. 531],

$$[\mathcal{I}(\boldsymbol{\theta})]_{ij} = \sum_{m=1}^M \text{Tr} \left[ \mathbf{C}_m^{-1} \frac{\partial \mathbf{C}_m}{\partial \theta_i^*} \mathbf{C}_m^{-1} \frac{\partial \mathbf{C}_m}{\partial \theta_j} \right]. \quad (3.92)$$

Using the identities on complex gradients of  $\mathbf{C}_m$  in the Appendix, we derive the FIM as

$$\mathcal{I}(\boldsymbol{\theta}) = N(1 + |F|^2) \sum_{k=1}^L \frac{d_k^2}{[d_k \sigma_h^2 (1 + |F|^2) + \sigma^2]^2} \begin{bmatrix} (\sigma_h^2)^2 \left( \frac{d_k \sigma_h^2}{\sigma^2} + 1 \right) & F \sigma_h^2 \\ F^* \sigma_h^2 & 1 + |F|^2 \end{bmatrix}. \quad (3.93)$$

For any unbiased estimators  $\hat{\boldsymbol{\theta}}$ , the classical Cramér-Rao bound (CRB) is then calculated as the inverse of FIM,

$$E \left[ \left( \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \left( \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right)^H \right] \geq \mathcal{C}(\boldsymbol{\theta}) = \mathcal{I}^{-1}(\boldsymbol{\theta}). \quad (3.94)$$

It appears challenging to derive the CRB in closed-form for general channel correlation  $\mathbf{C}_h$ . However, we next investigate a special case, i.e., i.i.d. fading channels, where both the maximum-likelihood (ML) estimators and the CRB are in closed-form.

### 3.4.1 ML Estimators for Temporally i.i.d. Fading

If the temporal correlation matrix  $\mathbf{C}_h = \mathbf{I}_L$ , the multi-packet ML estimator for  $F$  is trivial extension of the single-packet ones, which are given in (3.29). Note the multi-packet sample covariance  $\mathbf{S}$  in (3.87) should be used in this ML estimator. Similarly, the multi-packet CRB, under i.i.d. fading, is the single-packet CRB in (3.36) scaled by  $1/L$ , i.e., The CRB for  $F$  and  $\sigma_h^2$  are the diagonal entires of (3.36), respectively,

$$\mathcal{C}_F(\boldsymbol{\theta}) \triangleq \frac{\sigma^2 \sigma_h^2 (1 + |F|^2) + \sigma^4}{NL \sigma_h^4}, \quad \mathcal{C}_{\sigma_h^2}(\boldsymbol{\theta}) \triangleq \frac{(\sigma_h^2 (1 + |F|^2) + \sigma^2) (\sigma_h^2 + \sigma^2)}{NL (1 + |F|^2)}. \quad (3.95)$$

Next we consider the general problem where  $\mathbf{C}_h$  is not diagonal. We derive necessary conditions for the ML estimators and numerical methods in finding them.

### 3.4.2 ML Estimators for General $\mathbf{C}_h$

It can be shown that the complex gradient of  $\mathcal{L}(\boldsymbol{\theta})$  is [45, eq. 15.60],

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \theta_j^*} = \sum_{m=1}^M \left[ \mathbf{v}_m^H \mathbf{C}_m^{-1} \frac{\partial \mathbf{C}_m}{\partial \theta_j^*} \mathbf{C}_m^{-1} \mathbf{v}_m - \text{Tr} \left( \mathbf{C}_m^{-1} \frac{\partial \mathbf{C}_m}{\partial \theta_j^*} \right) \right]. \quad (3.96)$$

A necessary condition for joint ML estimators is that the complex gradient vanishes. In the Appendix B.2, the components of the complex gradient are shown to have the following form,

$$\begin{aligned} \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial F^*} &= -F \sum_{m=1}^M \frac{\lambda_m}{\mu_{1,m}} + \frac{1}{\sigma^2} \sum_{m=1}^M \frac{\lambda_m}{\mu_{1,m}^2} \mathbf{v}_m^H \begin{bmatrix} 1 \\ F \end{bmatrix} \begin{bmatrix} -\lambda_m F & \lambda_m + \sigma^2 \end{bmatrix} \mathbf{v}_m, \\ \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \sigma_h^2} &= -(1 + |F|^2) \frac{1}{\sigma_h^2} \sum_{m=1}^M \frac{\lambda_m}{\mu_{1,m}} + \frac{1}{\sigma_h^2} \sum_{m=1}^M \frac{\lambda_m}{\mu_{1,m}^2} \mathbf{v}_m^H \begin{bmatrix} 1 \\ F \end{bmatrix} \begin{bmatrix} 1 & F^* \end{bmatrix} \mathbf{v}_m, \end{aligned} \quad (3.97)$$

where  $\lambda_m$ 's are defined in (3.82). Note the necessary conditions above suggest the joint ML estimators are not solvable in a closed-form without additional assumptions on  $d_k$ 's and  $\sigma^2$ . Thus, we use numerical methods to find these ML estimators. In particular, a variation of the complex Newton-Raphson iteration can be used to find the true MLE [83, eq. 11],

$$\boldsymbol{\theta}_{p+1} = \boldsymbol{\theta}_p - \mathcal{H}^{-1}(\boldsymbol{\theta}_p) \left. \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_p}, \quad (3.98)$$

where the complex Hessian of the log-likelihood function is,

$$\mathcal{H}(\boldsymbol{\theta}) \triangleq \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^* \partial \boldsymbol{\theta}^T} = \begin{bmatrix} \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial F^* \partial F} & \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial F^* \partial \sigma_h^2} \\ \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \sigma_h^2 \partial F} & \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \sigma_h^2 \partial \sigma_h^2} \end{bmatrix}, \quad (3.99)$$

whose entries are derived after some lengthy yet straightforward algebra,

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial F^* \partial F} &= -\sum_{m=1}^M \frac{\lambda_m (\lambda_m + \sigma^2)}{\mu_{1,m}^2} + \frac{1}{\sigma^2} \sum_{m=1}^M \mathbf{v}_m^H \mathbf{A} \mathbf{v}_m, \\ \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial F^* \partial \sigma_h^2} &= -\frac{F \sigma^2}{\sigma_h^2} \sum_{m=1}^M \frac{\lambda_m}{\mu_{1,m}^2} + \frac{1}{\sigma_h^2} \sum_{m=1}^M \frac{\lambda_m}{\mu_{1,m}^3} \mathbf{v}_m^H \begin{bmatrix} 1 \\ F \end{bmatrix} \begin{bmatrix} -2\lambda_m F & \lambda_m - \lambda_m |F|^2 + \sigma^2 \end{bmatrix} \mathbf{v}_m, \\ \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \sigma_h^2 \partial \sigma_h^2} &= \frac{(1 + |F|^2)^2}{(\sigma_h^2)^2} \sum_{m=1}^M \frac{\lambda_m^2}{\mu_{1,m}^2} - \frac{2(1 + |F|^2)}{(\sigma_h^2)^2} \sum_{m=1}^M \frac{\lambda_m^2}{\mu_{1,m}^3} \mathbf{v}_m^H \begin{bmatrix} 1 \\ F \end{bmatrix} \begin{bmatrix} 1 & F^* \end{bmatrix} \mathbf{v}_m, \end{aligned} \quad (3.100)$$

and

$$\mathbf{A} \triangleq \frac{\lambda_m}{\mu_{1,m}^3} \begin{bmatrix} \lambda_m (\lambda_m |F|^2 - \lambda_m - \sigma^2) & -2\lambda_m (\lambda_m + \sigma^2) F^* \\ -2\lambda_m (\lambda_m + \sigma^2) F & -(\lambda_m + \sigma^2) (\lambda_m |F|^2 - \lambda_m - \sigma^2) \end{bmatrix}. \quad (3.101)$$

After some straightforward algebra, we verify that

$$-E[\mathcal{H}(\boldsymbol{\theta})] = \mathcal{I}(\boldsymbol{\theta}), \quad (3.102)$$

where  $\mathcal{I}(\boldsymbol{\theta})$  is the Fisher information matrix given in (3.93), derived using a different approach [45, eq. 15.52]. This serves a confirmation that the complex Hessian expressions derived above are correct. The starting point of iteration-based numerical methods matter, and we propose to start from the ML estimators for the i.i.d. case (3.29). When

it converges (3.98), we have the true ML estimator for  $F$  which we denote as  $\hat{F}_{ML}$ .

Now we give another estimator for  $F$  based on method of moments,

$$\hat{F}_{MM} = \frac{S_{22} - S_{11} + \sqrt{(S_{22} - S_{11})^2 + 4|S_{12}|^2}}{2S_{12}}, \quad (3.103)$$

conditioned on  $S_{12} \neq 0$ , where  $S_{ij}$ 's are entries of the sample covariance (3.28). It can be shown that this MM estimator is also a true ML estimator by treating both  $\mathbf{h}$  and  $F$  as deterministic<sup>3</sup>. One observes that  $\hat{F}_{MM}$  is the same regardless the channel is correlated or not, e.g., see (3.87). The performance of above estimators are evaluated in the next section.

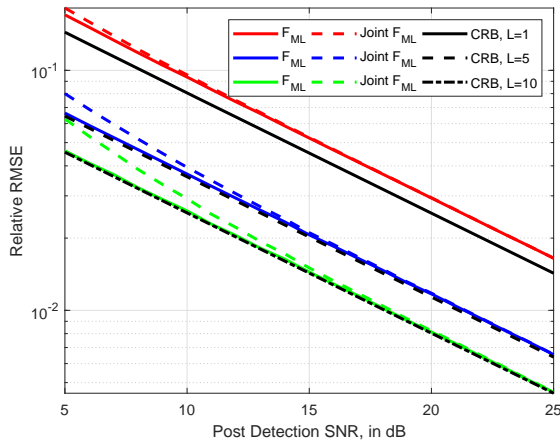
### 3.5 Numerical Results

In this section, we explore the performance of estimators in the previous section through numerical examples. Consider a narrow-band MISO communications system with  $N = 4$  transmit antennas, whose carrier frequency is 2.1 GHz. This frequency is chosen based on the first E-UTRA down-link operating band in LTE specifications [2, pg. 40, Table 5.5-1]. The duration of each data packet equals to a subframe of LTE, i.e.,  $T_s = 1$  ms [24, pg. 5] [1, Fig. 4.1-1]. Block-fading channel is assumed, such that during one data packet, the channel information remains the same, but it generally varies from packet to packet [16].

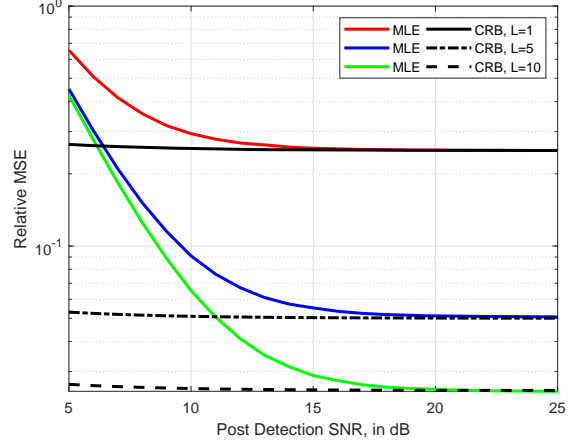
For each data packet, a training sequence precedes data sequence [53, Fig. 1(a)]. We take the two partitions of the training sequence  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$  from a normalized discrete Fourier transform (DFT) matrix of dimension  $K = T/2 = 32$ , e.g., [16, eq. 10].

---

<sup>3</sup>Details see Remark in Sec. 3.3 and Appendix B.1. Although  $\hat{F}_{MM}$  in (3.103) is another ML estimator, we call it the MM (method of moments) estimator to distinguish it from the MLE found via Newton-Raphson iterations (3.98).



(a) RMSE of  $\hat{F}_{ML}$  in i.i.d. Fading.



(b) RMSE of  $\hat{\sigma}_h^2$  under i.i.d. Fading.

Figure 3.2: Relative MSE of ML Estimators in i.i.d. Fading with  $N = 4$ .

In particular, the first part  $\mathbf{X}_1$  is chosen as the first  $N$  rows, while  $\mathbf{X}_2$  the next  $N$  rows, and  $\mathbf{X}_i \mathbf{X}_i^H = K \mathbf{I}_N$  for  $i = 1, 2$ . The unknown antenna impedance is that of a dipole  $Z_A = 73 + j42.5 \Omega$ . The load impedance (3.8) is  $Z_1 = 50 \Omega$  for the first  $K = 32$  symbols of each training sequence, and  $Z_2 = 60 + j20 \Omega$  for the remaining  $T - K = 32$  symbols. From (3.11), it follows  $F = 0.9646 - j0.1032$ .

From (3.5) with  $Z_L = Z_1$ , (3.9) and (3.14), we define the average post-detection signal-to-noise ratio (SNR) of a received symbol as

$$\gamma \triangleq E[\rho_{sym}] = \frac{R_1}{\sigma_n^2 |Z_A + Z_1|^2} \text{Tr} (E[\mathbf{x}\mathbf{x}^H] E[\mathbf{g}^* \mathbf{g}^T]) = \frac{\sigma_h^2 \cdot P}{\sigma_n^2}. \quad (3.104)$$

Next we numerically explore the performance of the true ML estimators in (3.18) for i.i.d. Rayleigh fading channels, i.e.,  $\mathbf{C}_h = \mathbf{I}_L$ .

In Fig 3.2a, the relative root MSE (RMSE) of the ML estimator  $\hat{F}_{ML}$  in (3.29) approaches its corresponding Cramér-Rao bound (CRB) within a dB to for one packet. This gap vanishes with a sufficient number of training packets, e.g.,  $L = 5$ . Also plotted

is the joint ML estimator derived under the hybrid estimation framework [81, eq. 29]. At low SNR,  $\hat{F}_{ML}$  beats the joint ML estimator, despite the latter assumes more knowledge, i.e., knowing  $\sigma_h^2$ . Thus, more knowledge does not guarantee better estimation accuracy unless used optimally, and the hybrid framework is not the optimal formulation of this problem in terms of estimation efficiency.

In Fig. 3.2b, we plot the MSE of the MLE  $\hat{\sigma}_h^2$  in (3.18) and its CRB (3.95). Observe that  $\hat{\sigma}_h^2$  becomes efficient at high SNR for all  $L$  plotted. Also, the relative CRB's saturate at  $1/(N \cdot L)$  as SNR increases asymptotically, i.e.,  $\rho \rightarrow \infty$ . This is confirmed by its expression in (3.95). Note  $\sigma_h^2$  is a nuisance parameter, which has to be estimated in this classical estimation framework. We next explore the performance of  $F$  estimators under correlated fading channels.

Assume Clarke's model, and the normalized channel correlation matrix is [9, 21, 86],

$$\mathbf{C}_h = \begin{bmatrix} R[0] & R[-1] & \cdots & R[-L+1] \\ R[1] & R[0] & \cdots & R[-L+2] \\ \vdots & \ddots & \ddots & \vdots \\ R[L-1] & R[L-2] & \cdots & R[0] \end{bmatrix}, \quad (3.105)$$

where  $R[l] = J_0(2\pi f_d T_s |l|)$ ,  $J_0(\cdot)$  is the zeroth-order Bessel function of the first kind,  $T_s = 1 \text{ ms}$  is the sampling interval, and  $l$  is the sample difference. The fading frequency (maximum Doppler frequency) is  $f_d \triangleq v/\lambda$ , where  $v$  is the velocity of the fast moving scatterer and  $\lambda$  the wave-length of the carrier frequency.

LTE supports peak data rates of 100/50 Mbps (downlink/uplink) for velocity up to 350 km/h; for velocity between 350 and 500 km/h, a high-speed railways version of LTE, i.e., LTE-R, maintains reliable communication at lower capacity [38, Table 1].

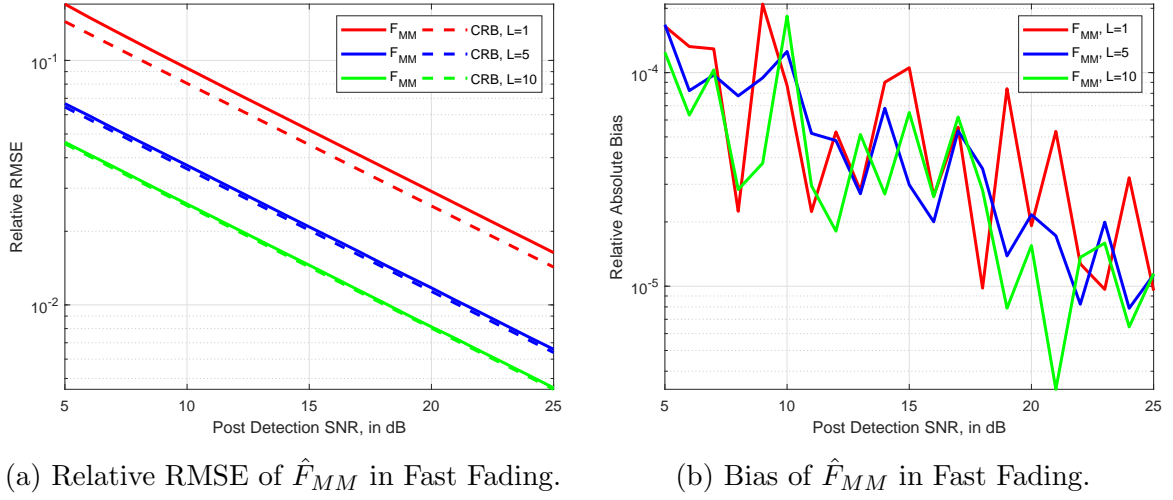


Figure 3.3: Relative RMSE and Absolute Bias of  $\hat{F}_{MM}$  in Fast Fading,  $N = 4$ .

Consider first a fast fading scenario, i.e., a high-speed environment where the velocity is at  $v = 300$  km/h [24, pg. 11]. The fading frequency is calculated as  $f_d = 583.3$ Hz. Note the normalized correlation matrix  $\mathbf{C}_h$  under this setting is,

$$\mathbf{C}_h = \begin{bmatrix} 1.0000 & -0.3971 & 0.2856 & -0.1720 & 0.0556 \\ \vdots & \ddots & \ddots & \ddots & \vdots \end{bmatrix},$$

and its eigenvalues are,  $[0.5527, 0.5766, 0.6267, 1.0822, 2.1618]$ . The correlation matrix  $\mathbf{C}_h$  is real symmetric Toeplitz, so its first row contains all the distinct entries, and other rows are structured repetitions of the first row. These eigenvalues indicate the decorrelated channels are not identical but comparable in fast fading, with  $L$  orders of temporal diversity from  $L$  packets.

Under this fading, we plot in Fig. 3.3a the RMSE of  $\hat{F}_{MM}$  in (3.103), compared to its general CRB (3.94). Since the channel is correlated, the  $\hat{F}_{MM}$  is generally sub-optimal to the true MLE. We observe the  $\hat{F}_{MM}$  is about 1 dB away from the CRB for  $L = 1$ .

They eventually achieve the CRB as  $L$  increases further to 5 and 10. Note  $\hat{F}_{MM}$  can be interpreted as discarding the temporal correlation over packets and view the  $L$  channels i.i.d.. Thus, for  $L = 1$  the  $\hat{F}_{MM}$  is also true ML estimator (3.29). For  $L = 5, 10$ , the little gap between the RMSE of  $\hat{F}_{MM}$  and its CRB does not justify the extra computational complexity in finding the  $\hat{F}_{ML}$ , e.g., via Newton-Raphson (3.98). Thus, we defer plotting of  $\hat{F}_{ML}$  till later; see Figs. 3.5a and 3.5b.

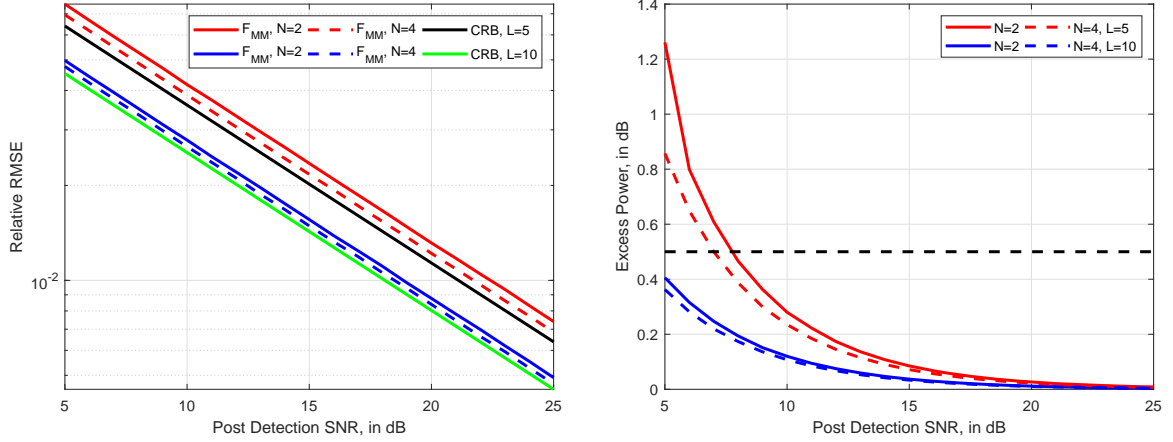
We plot the relative absolute bias of  $\hat{F}_{MM}$  in Fig. 3.3b, i.e.,  $|E[\hat{F}_{MM} - F]|/|F|$ . The random fluctuation at all  $L$  and SNR suggests its true bias is buried below these curves. Thus, although it remains unclear if  $\hat{F}_{MM}$  is unbiased, its bias (if any) is tiny and can be ignored practically.

Next we investigate a more correlated case. When the fading is moderate, e.g.,  $v = 50$  km/h and  $f_d = 97.2$  Hz. The correlation matrix  $\mathbf{C}_h$  and its eigenvalues are, respectively,

$$\mathbf{C}_h = \begin{bmatrix} 1.0000 & 0.9089 & 0.6602 & 0.3210 & -0.0199 \\ \vdots & \ddots & \ddots & \ddots & \vdots \end{bmatrix},$$

and  $[2.3661 \times 10^{-8}, 7.0552 \times 10^{-4}, 0.0646, 1.3589, 3.5757]$ . Under this fading conditions, 5 correlated channels provide about 2 orders of diversity.

In Figs. 3.4a and 3.4b, the MM estimator  $\hat{F}_{MM}$  are plotted with four combinations of  $N = 2, 4$  and  $L = 5, 10$ . The general trend is that for a given  $L$  and total transmit power constraint in (3.14), having 4 transmit antennas is better than having 2 both in terms of RMSE and excess power. This is due to the fact that transmit antennas provide i.i.d. fading and hence spatial diversity. The CRB (3.94) under correlated fading generally differs for different  $L$  or  $N$ . However, for the particular cases in Fig. 3.4a, this difference is less than 0.3% at low SNR, and vanishes as SNR increases. So we only plot the  $N = 2$



(a) Relative RMSE of  $\hat{F}_{MM}$  and CRB.

(b) Excess Power due to Mismatch in  $\hat{Z}_L$ .

Figure 3.4: Effects of Transmit Diversity under Medium Fading,  $N = 2, 4$ .

CRB's. The  $\hat{F}_{MM}$  is within a fraction of a dB to its CRB for all values of  $N$ ,  $L$ , and SNR.

In Fig. 3.4b, we plot the excess (transmit) power needed due to mismatch in  $\hat{Z}_L$ , i.e.,

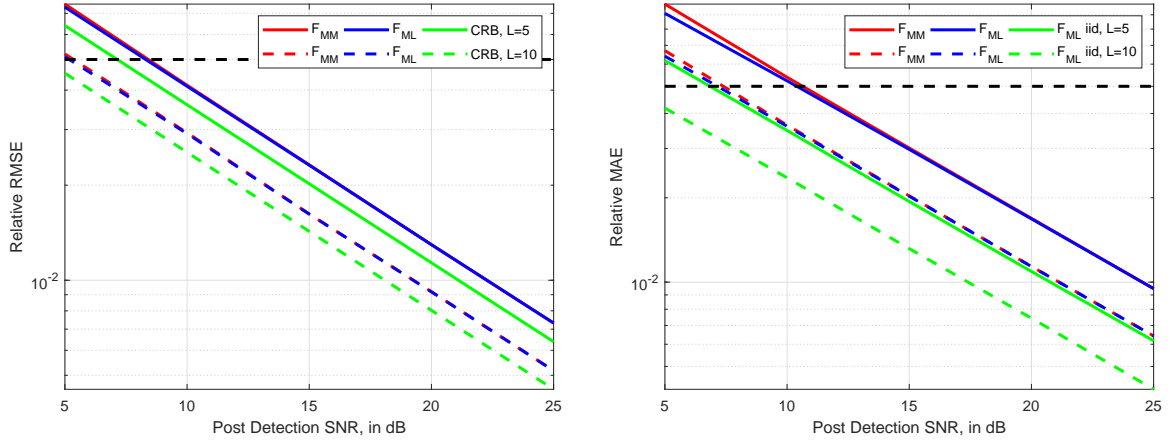
$$-10 \log_{10} \left( 4R_A \cdot E \left[ \hat{R}_L / |Z_A + \hat{Z}_L|^2 \right] \right), \quad (3.106)$$

where the load is conjugately matched to an estimate of  $Z_A$ ,

$$\hat{Z}_L \triangleq \hat{Z}_A^*, \quad \hat{Z}_A = \frac{Z_2 \sqrt{R_1/R_2} \cdot \hat{F}_{MM} - Z_1}{1 - \sqrt{R_1/R_2} \cdot \hat{F}_{MM}}, \quad (3.107)$$

and  $\hat{Z}_A$  is calculated via the invariance principle of MLE, given  $K = T/2$ . If an excess power of 0.5 dB or smaller is desired, a RMSE of 5% suffices to achieve this goal. This can be used as a rule of thumb on how accurate  $F$  needs to be estimated.

When the fading is extremely slow, e.g.,  $v = 5$  km/h and  $f_d = 9.72$  Hz. The highly



(a) Relative RMSE of  $F$  Estimators for  $N = 4$ . (b) Relative MAE of  $F$  Estimators for  $N = 1$ .

Figure 3.5: Benefits of  $\hat{F}_{ML}$  over  $\hat{F}_{MM}$  in Slow Fading,  $N = 1, 4$ .

correlated  $\mathbf{C}_h$  and its eigenvalues are, respectively,

$$\mathbf{C}_h = \begin{bmatrix} 1.0000 & 0.9991 & 0.9963 & 0.9916 & 0.9851 \\ \vdots & \ddots & \ddots & \ddots & \vdots \end{bmatrix},$$

and  $[2.172 \times 10^{-14}, 6.501 \times 10^{-10}, 6.098 \times 10^{-6}, 0.0186, 4.981]$ . These temporal channels are highly correlated, even between the first and last channels. Essentially 5 temporal packets results in only one order of diversity. This is a worse case for temporal diversity.

Next, we explore the benefits of  $\hat{F}_{ML}$  over  $F_{MM}$  under slow fading. Specifically,  $L = 5, 10$  packets are chosen, and  $N = 4$  spatial diversity is available at the transmitter in Fig. 3.5a. Here the  $\hat{F}_{ML}$ , found via Newton-Raphson iteration<sup>4</sup> in (3.98), exhibits negligible improvement over the simple  $\hat{F}_{MM}$ . From (3.82) and previous numerical results, another rule of thumbs is that, to be 1 dB within the CRB, a total of 4 orders of diversity,

<sup>4</sup>To improve numerical stability, we here find  $\hat{F}_{ML}$  by assuming  $\hat{\sigma}_h^2$  is known and only estimates of  $F$  are iteratively updated using the upper left entry of the complex Hessian (3.99). This serves a lower bound on the ML estimator not knowing  $\sigma_h^2$ .

temporal and/or spatial, is needed.

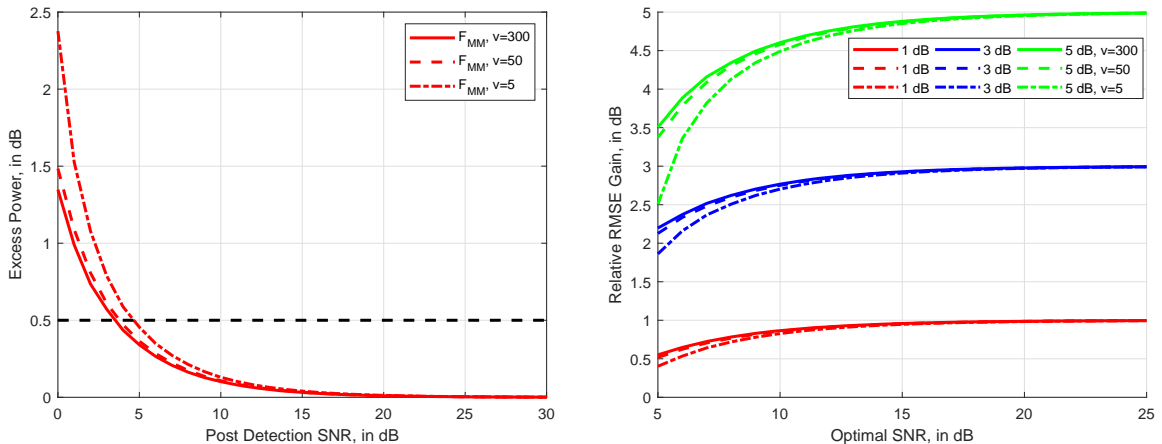
The classical CRB for  $F$  given  $\sigma_h^2$  is the inverse of the upper left entry of the FIM in (3.93), which we denote as  $\mathcal{C}_{F|\sigma_h^2}(\boldsymbol{\theta})$ . It has been shown that  $\mathcal{C}_{F|\sigma_h^2}(\boldsymbol{\theta})$  is a lower bound on  $\mathcal{C}_F(\boldsymbol{\theta})$ , the upper left entry of (3.94). This is intuitive, as  $\mathcal{C}_{F|\sigma_h^2}(\boldsymbol{\theta})$  assumes more knowledge, i.e., knowing  $\sigma_h^2$ , than  $\mathcal{C}_F(\boldsymbol{\theta})$ . However, for all points considered in Fig. 3.5a, these two CRB's differ by about 0.3% at low SNR and merge at high SNR. Thus, the 1 dB gap between  $\hat{F}_{MM}$  (or  $\hat{F}_{ML}$ ) and the CRB is because CRB is loose for finite sample size. Thus, in terms of estimating  $F$ , little benefit is gained by knowing  $\sigma_h^2$  when when enough (spatial) diversity exists, e.g.,  $N = 4$ . This condition is often satisfied, since modern wireless communication systems tend to deploy multiple antennas at the down-link transmitter, e.g. base stations for LTE and 5G and access points for WLAN [32, 50, 84].

When only one transmit antenna is used, in Fig. 3.5b relative Mean Absolute Error (MAE) is plotted instead of the RMSE, due to the heavy tails of the single-packet  $\hat{F}_{ML}$  [80, Sec. V-A]. Here  $\hat{F}_{ML}$  provides about 1 dB of gain at low SNR, but converges to  $\hat{F}_{MM}$  as SNR increases beyond 10 dB. Compared to the i.i.d. fading, high correlation costs about 3 dB in general.

Once antenna impedance has been estimated using (3.107) and the load is matched to this estimate. Starting from the next packet, the receiver could perform the minimum mean-square error (MMSE) estimator for channel estimation [16]. It can be shown that the relative MSE for this estimator over  $L$  packets is,

$$\frac{J_L(\sigma_h^2)}{\sigma_h^2} = \frac{\text{Tr} \left[ (\mathbf{C}_h^{-1} + \mathbf{I}_L)^{-1} \right]}{L} \frac{\sigma_{n,sym}^2}{\sigma_{n,sym}^2 + PT\sigma_h^2/N}, \quad (3.108)$$

where  $\mathbf{C}_h$  and  $\sigma_h^2$  from (3.79) are assumed known. The factor  $\text{Tr} \left[ (\mathbf{C}_h^{-1} + \mathbf{I}_L)^{-1} \right] / L$  to



(a) Excess Power under Different Fading.

(b) Channel Estimation Gain in dB.

Figure 3.6: Benefits of  $\hat{F}_{MM}$  on MMSE Channel Estimation,  $L = 10$ ,  $N = 4$ .

the one packet MMSE, i.e.,  $J_1$  given in (1.34). is the benefit from multi-packet smoothing. We fix the total transmit power such that for a perfectly matched receiver, the receive SNR (3.104) is

$$\gamma_{opt} = \frac{P\sigma_{h,opt}^2}{\sigma_{n,sym}^2}, \quad (3.109)$$

which we call the *optimal SNR*. Then we calculate the gain for MMSE channel estimation using (3.107) compared to the originally mismatched receiver. We consider three cases of mismatch, where the power loss (to the optimal SNR) is 1, 3, and 5 dB, respectively. As shown in Fig. 3.6b, there is always a benefit in using our impedance estimation algorithm for all scenarios considered. These original power losses are compensated partially at low SNR and nearly completely at high SNR.

Hassib and Hochwald derived a lower bound on ergodic capacity when channel estimation errors exist [37, Th. 1]. In particular, if the MMSE estimate in (1.28) is treated as correct during data transmission for one packet, then the ergodic capacity is lower

bounded by [37, eq. 21]

$$C_l = E \left[ \log_2 \left( 1 + \gamma_{\text{eff}} \frac{\mathbf{h}^H \mathbf{h}}{N} \right) \right], \quad (3.110)$$

where the distribution is over  $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$  and the effective SNR is derived as (1.43),

$$\gamma_{\text{eff}} \triangleq \frac{P(\sigma_h^2 - J_1)}{\sigma_{n,\text{sym}}^2 + P \cdot J_1} = \gamma \cdot \frac{1}{1 + (1 + 1/\gamma)N/T}, \quad (3.111)$$

where  $\gamma = P\sigma_h^2/\sigma_{n,\text{sym}}^2$  is defined in (3.104). It has been shown that the capacity lower bound in (3.115) can be simplified into a closed-form [71, eq. 20],

$$C_l = \log_2(e) e^{N/\gamma_{\text{eff}}} \sum_{k=1}^N E_k \left( \frac{N}{\gamma_{\text{eff}}} \right), \quad (3.112)$$

where  $E_n(a)$  is the exponential integral for nonnegative integer  $n$  and  $\text{Re}\{a\} > 0$ ,

$$E_n(a) = \int_1^\infty \frac{e^{-at}}{t^n} dt. \quad (3.113)$$

It can be shown for  $n \geq 2$ ,  $E_n(a)$  is a function of  $E_1(a)$ , e.g.,

$$\begin{aligned} E_2(a) &= \int_1^\infty \frac{e^{-at}}{t^2} dt = - \int_1^\infty e^{-at} d\frac{1}{t} = - \frac{e^{-at}}{t} \Big|_1^\infty + \int_1^\infty \frac{1}{t} de^{-at} \\ &= e^{-a} - a \cdot E_1(a), \\ E_3(a) &= \int_1^\infty \frac{e^{-at}}{t^3} dt = -\frac{1}{2} \int_1^\infty e^{-at} d\frac{1}{t^2} = -\frac{1}{2} \left( \frac{e^{-at}}{t^2} \Big|_1^\infty - \int_1^\infty \frac{1}{t^2} de^{-at} \right) \\ &= \frac{1}{2} [e^{-a} - a \cdot E_2(a)], \\ E_4(a) &= \int_1^\infty \frac{e^{-at}}{t^4} dt = -\frac{1}{3} \int_1^\infty e^{-at} d\frac{1}{t^3} = -\frac{1}{3} \left( \frac{e^{-at}}{t^3} \Big|_1^\infty - \int_1^\infty \frac{1}{t^3} de^{-at} \right) \\ &= \frac{1}{3} [e^{-a} - a \cdot E_3(a)]. \end{aligned} \quad (3.114)$$

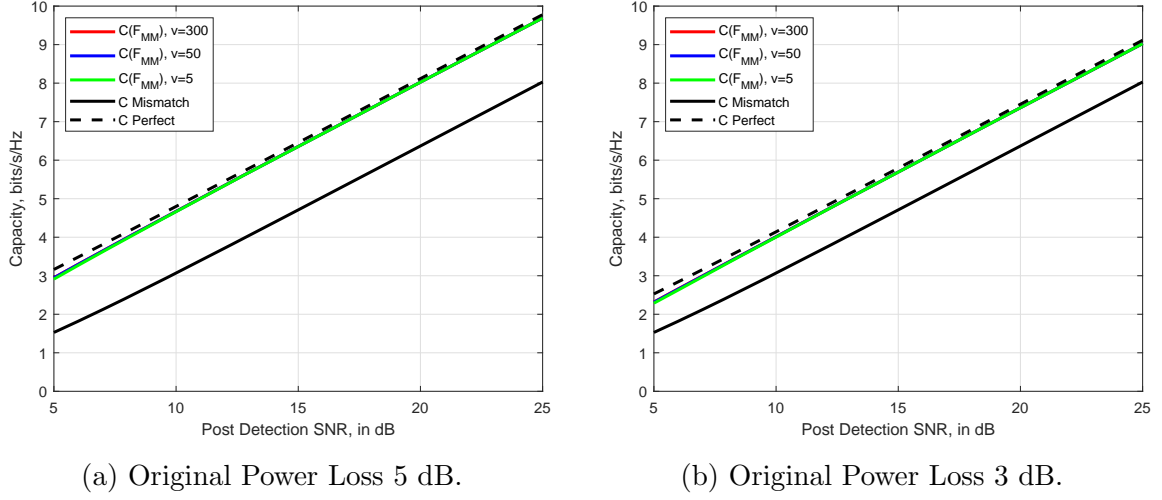


Figure 3.7: Ergodic Capacity in (3.115),  $L = 10$ ,  $N = 4$ .

With above simplifications of  $E_n(a)$  and after some straightforward algebra, the capacity lower bound for  $N = 4$  transmit antennas can be written as,

$$C_l = \frac{\log_2 e}{2} \left[ \left( \frac{N}{\gamma_{\text{eff}}} - 1 \right)^2 + \frac{8}{3} + \tau \left( \frac{N}{\gamma_{\text{eff}}} \right) \exp \left( \frac{N}{\gamma_{\text{eff}}} \right) E_1 \left( \frac{N}{\gamma_{\text{eff}}} \right) \right], \quad (3.115)$$

where  $\tau(x)$ , a function of  $x \in \mathbb{R}^+$ , is defined as

$$\tau \triangleq 2 - x \cdot (2 + x^2 - x). \quad (3.116)$$

Note in Figs. 3.7a and 3.7b, the load impedance is originally mismatched, with 5 dB and 3 dB power loss, respectively. The horizontal axis is the post-detection SNR (3.104) for the mismatched receiver. After our proposed antenna impedance estimation (3.107), the receiver quickly compensates its impedance mismatch (3.115) and improves (ergodic) capacity, which is shown as  $C(\hat{F}_{MM})$  in Figs. 3.7a and 3.7b. The solid black line represents the capacity of the mismatched receiver, while the black dash line is the

capacity upper bound, where the load is optimally matched and no channel estimation errors occur. One observes that  $C(\hat{F}_{MM})$  hones in this capacity upper bound for all SNR and fading conditions plotted. Compared to the mismatched receiver, the capacity gain is about 50% at low SNR and about 20% at high SNR for the 5 dB power loss case.

### 3.6 Conclusion

In this chapter, we consider antenna impedance estimation at the receiver of a MISO system. A classical estimation framework is developed incorporating Rayleigh fading assumption. We then derive in closed-form the maximum-likelihood (ML) estimator for  $F$  under i.i.d. fading and a method of moments (MM) estimator in correlated channels. We show this MM estimator is also another ML estimator by treating the fading channels as deterministic. Furthermore, necessary conditions and a Newton-Raphson iteration algorithm are derived to find the ML estimators under correlated channels.

Numerical results demonstrate that the MM estimator approaches its corresponding Cramér-Rao bounds (CRB) given sufficient degrees of diversity, spatial and/or temporal. A rule of thumb is to approach the CRB within 1 dB, 4 degrees of diversity is needed. This MM estimator also enjoys negligible bias for all SNR and fading conditions considered in this chapter. These findings suggest a fast PCA-based algorithm, i.e., MLE in i.i.d. fading and the MM based estimator in correlated fading, that estimates antenna impedance in real-time for all Rayleigh fading channels. We also found that a relative RMSE (or MAE) of 5% or less in estimating  $F$  leads to an excess power of 0.5 dB or smaller.

We demonstrate, via numerical examples, the gains of channel estimation accuracy due to our proposed antenna impedance estimation algorithm. Ergodic capacity is boosted by this same algorithm, compared to an originally mismatched receiver, and hones in a

capacity upper bound. Thus, this proposed algorithm enables accurate impedance estimation and mismatch compensation quickly, in a matter of milliseconds.

In the next chapter, we investigate the most general and thus more important problem, i.e., antenna impedance estimation at multi-antenna receivers. In particular, we extend mathematical solutions developed in this chapter to its MIMO counterpart.

# Chapter 4

## Antenna Impedance Estimation at MIMO Receivers

In previous chapters, we considered impedance estimation at single-antenna receivers. However, modern receivers are often equipped with multiple antennas for multiplexing and/or diversity benefits. Therefore, in this chapter, we investigate the general and more important problem of estimating the antenna impedance matrix at MIMO receivers.

### 4.1 Introduction

This chapter considers antenna impedance estimation algorithms using training data for multiple-input, multiple-output (MIMO) communication systems. We assume the receiver switches its impedance in a predetermined fashion during each training sequence. In i.i.d. Rayleigh fading channels, the maximum-likelihood (ML) estimator is derived for the impedance matrix as a function of the top block eigen-vector of the sample covariance matrix. This ML estimator is shown to be a method of moments (MM) when the fading

channel is temporally correlated. Fundamental lower bounds, e.g., Cramér-Rao bounds, on these estimators are derived and important properties of these estimators, e.g., bias and efficiency, are explored through numerical simulations. The trade-off between channel and impedance estimation is demonstrated empirically.

The rest of the chapter is organized as follows. We present our system model in Sec. 4.2, derive a set of maximum-likelihood estimators for the MIMO antenna impedance and channel covariance matrix in Sec. 4.3, and derive method of moment (MM) estimators of these matrices under multiple packets scenarios and discuss ML estimators in Sec. 4.4. We then explore important properties of the estimators through numerical examples in Sec. 4.5, and summarize our conclusions in Sec. 4.6.

## 4.2 System Model

Consider a narrowband multiple-input multiple-output (MIMO) communications link with  $M$  receive antennas and  $N$  transmit antennas. The receiver model is illustrated in Fig. 4.1. This circuit model is identical to the ones widely used to model a scenario, where amplifier noise dominates [29, 51, 79]. This model is also a special case of the more general and complex models, which include additional noise sources, e.g., sky-noise and downstream noise [26, 27, 43].

In Fig. 4.1, we model the antenna array by its Thevenin equivalent,

$$\mathbf{v} = \mathbf{Z}_A \mathbf{i} + \mathbf{v}_o, \quad (4.1)$$

where  $\mathbf{v}, \mathbf{i} \in \mathbb{C}^M$  are the voltage across, and current into, the antenna array terminals. In particular, the antenna impedance is a symmetric matrix in  $\mathbb{C}^{M \times M}$  due to the reciprocity

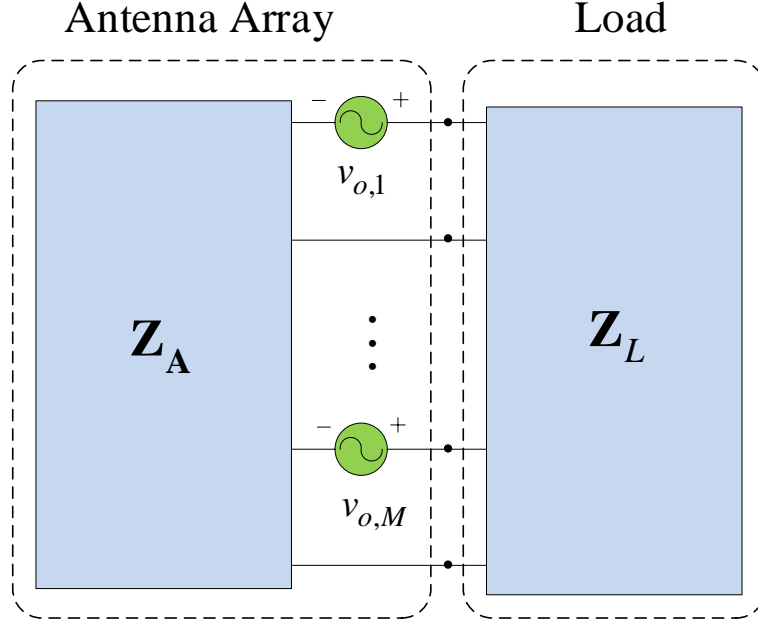


Figure 4.1: Circuit model of a multiple-antenna receiver

theorem of electromagnetics [11],

$$\mathbf{Z}_A = \mathbf{R}_A + j\mathbf{X}_A, \quad (4.2)$$

where  $\mathbf{R}_A$  and  $\mathbf{X}_A$  are the resistance and reactance matrices, respectively. The incident electromagnetic field induces open-circuit voltage  $\mathbf{v}_o \in \mathbb{C}^M$  in (4.1). Under flat-fading conditions, the open-circuit voltage  $\mathbf{v}_o$  is modeled as [27]

$$\mathbf{v}_o = \mathbf{G}\mathbf{x}, \quad (4.3)$$

where  $\mathbf{x} \in \mathbb{C}^N$  is the transmitted symbol and  $\mathbf{G} \in \mathbb{C}^{M \times N}$  is the matrix of fading path gains. Similar to the previous two chapters, we consider a Rayleigh fading environment,

where transmit antennas are sufficiently separated. Thus, columns of  $\mathbf{G}$  are modeled as i.i.d. zero-mean, complex Gaussian random vectors,  $\mathbf{g}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{g}})$ . As shown in Fig. 3.1, noisy voltage signal  $\mathbf{v}_L \in \mathbb{C}^M$  across load impedance  $\mathbf{Z}_L \in \mathbb{C}^{M \times M}$  is observed [29, 51, 79],

$$\mathbf{v}_L = \mathbf{Z}_L (\mathbf{Z}_A + \mathbf{Z}_L)^{-1} \mathbf{G} \mathbf{x} + \mathbf{n}_L , \quad (4.4)$$

where the noise  $\mathbf{n}_L \in \mathbb{C}^M$  is a zero-mean, circularly-symmetric, complex Gaussian random vector with covariance  $E[\mathbf{n}_L \mathbf{n}_L^H] = \boldsymbol{\Sigma}_L$ , which is hereafter denoted by  $\mathbf{n}_L \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_L)$ .

As mentioned in the previous chapter, performance of estimators typically depend on the signal-to-noise ratio (SNR) in estimation theory, which is conventionally defined, for example, as  $\text{Tr}[\mathbf{v}_L \mathbf{v}_L^H \boldsymbol{\Sigma}_L^{-1}]$  [27, Sec. II-A]. In circuit theory, however, power depends on both voltage and current [43, eq. 22]. This estimation-theory SNR formula does not correctly predict the ratio of the physical signal power and noise power in the receiver front-end. For a given  $\mathbf{v}_o$ , this ratio of physical signal power to noise power across the load is given by

$$\rho \triangleq \frac{1}{\sigma_n^2} \text{Tr} \left[ \mathbf{R}_L^{1/2} (\mathbf{Z}_A + \mathbf{Z}_L)^{-1} \mathbf{v}_o \mathbf{v}_o^H (\mathbf{Z}_A + \mathbf{Z}_L)^{-H} \mathbf{R}_L^{1/2} \right] , \quad (4.5)$$

where  $\sigma_n^2$  represents the noise power at the output of the amplifier and  $\mathbf{R}_L = \text{Re}\{\mathbf{Z}_L\}$  is the load resistance. As in the previous chapter, we correct this discrepancy by defining  $\boldsymbol{\Sigma}_L$  in a way that ensures the SNR and physical power ratio  $\rho$  coincide:

$$\boldsymbol{\Sigma}_L \triangleq \sigma_n^2 \mathbf{Z}_L \mathbf{R}_L^{-1} \mathbf{Z}_L^H .$$

With this definition, it is convenient to redefine the observed signal as

$$\mathbf{w} \triangleq \mathbf{R}_L^{1/2} \mathbf{Z}_L^{-1} \mathbf{v}_L = \mathbf{R}_L^{1/2} (\mathbf{Z}_A + \mathbf{Z}_L)^{-1} \mathbf{G} \mathbf{x} + \mathbf{n} , \quad (4.6)$$

where  $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_M)$  represents physical noise power referred to the amplifier output. This signal model (4.6) correctly connects the estimator performance to physical signal-to-noise power ratio. As mentioned before in the previous chapter, this connection is essential to accurately predict the impact of impedance mismatch at MIMO receivers on important system-level metrics, such as channel capacity.

Suppose the channel gain matrix and antenna impedance matrix are unknown to the receiver. As in the previous two chapters, our objective is to jointly estimate these two matrices using observations of known training sequences. Suppose the transmitter sends a known training sequence of length  $T$ , i.e.,  $\mathbf{x}_1, \dots, \mathbf{x}_T \in \mathbb{C}^N$  to the receiver, during which the receiver synchronously shifts its impedance as  $\mathbf{Z}_{L,1}, \dots, \mathbf{Z}_{L,T}$ . Also assume both the fading path gain  $\mathbf{G}$  and impedance  $\mathbf{Z}_A$  remain fixed during each transmission. The received observations take the following form,

$$\mathbf{w}_t = \mathbf{R}_{L,t}^{1/2} (\mathbf{Z}_A + \mathbf{Z}_{L,t})^{-1} \mathbf{G} \mathbf{x}_t + \mathbf{n}_t , \quad (4.7)$$

where  $t = 1, 2, \dots, T$  and the additive noises  $\mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_M)$  are independent and identically distributed (i.i.d.).

We again assume the load impedance takes on two possible matrices,

$$\mathbf{Z}_L = \begin{cases} \mathbf{Z}_1 , & 1 \leq t \leq K , \\ \mathbf{Z}_2 , & K \leq t \leq T . \end{cases} \quad (4.8)$$

Similar to the previous two chapters, we assume  $\mathbf{Z}_L = \mathbf{Z}_1$  is the load impedance used to receive the transmitted data, and is matched to our best estimate of  $\mathbf{Z}_A$ ; additionally  $\mathbf{Z}_L = \mathbf{Z}_2$  is an impedance variation introduced in order to make  $\mathbf{Z}_A$  observable. To estimate  $\mathbf{Z}_A$ ,  $\mathbf{Z}_1 \neq \mathbf{Z}_2$  is required.

Note that in order to perform optimal detection of the transmitted symbols in (4.6), an accurate estimate of the entire matrix coefficient of  $\mathbf{x}$  is ideal, not simply the fading path gain matrix  $\mathbf{G}$ . This motivates the definition of an effective channel matrix that communication algorithms need,

$$\mathbf{H} \triangleq \mathbf{R}_1^{1/2} (\mathbf{Z}_A + \mathbf{Z}_1)^{-1} \mathbf{G} \in \mathbb{C}^{M \times N}, \quad (4.9)$$

whose columns are also i.i.d. zero-mean, complex Gaussian,  $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_h)$ , and

$$\boldsymbol{\Sigma}_h = \mathbf{R}_1^{1/2} (\mathbf{Z}_A + \mathbf{Z}_1)^{-1} \boldsymbol{\Sigma}_g (\mathbf{Z}_A + \mathbf{Z}_1)^{-H} \mathbf{R}_1^{1/2}. \quad (4.10)$$

With this choice of load impedance in (4.8) and definition of  $\mathbf{H}$  (4.9), we can express the observations in a simpler, bilinear form. The voltage observation (4.6) at the load is then,

$$\mathbf{w}_t = \begin{cases} \mathbf{H}\mathbf{x}_t + \mathbf{n}_t, & 1 \leq t \leq K \\ \mathbf{F}\mathbf{H}\mathbf{x}_t + \mathbf{n}_t, & K + 1 \leq t \leq T \end{cases} \quad (4.11)$$

where  $\mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_M)$  are independent over time  $t$  and we define  $\mathbf{F} \in \mathbb{C}^{M \times M}$  as a one-to-one mapping of  $\mathbf{Z}_A$  for mathematical convenience, conditioned on  $\mathbf{Z}_1 \neq \mathbf{Z}_2$ ,

$$\mathbf{F} = \mathbf{R}_2^{1/2} (\mathbf{Z}_2 + \mathbf{Z}_A)^{-1} (\mathbf{Z}_1 + \mathbf{Z}_A) \mathbf{R}_1^{-1/2}. \quad (4.12)$$

We present this chapter as a generalization of the previous chapter to MIMO receivers. Here the goal is to derive maximum-likelihood (ML) estimators for  $\mathbf{H}$ ,  $\mathbf{Z}_A$  and  $\Sigma_{\mathbf{h}}$  based on the observations (4.11). From the invariance principle of maximum-likelihood estimation (MLE) [45, pg. 185], knowing the MLE of  $\mathbf{F}$  is equivalent to knowing that of  $\mathbf{Z}_A$  and vice versa. Theoretically it suffices to derive estimators for  $\mathbf{H}$ ,  $\mathbf{F}$  and  $\Sigma_{\mathbf{h}}$ . Specifically, we follow the two-step procedure as described in the last chapter: First, we consider joint maximum-likelihood estimation of  $\mathbf{F}$  and  $\Sigma_{\mathbf{h}}$ , treating  $\mathbf{H}$  as a nuisance parameter. Second, given estimates of  $\mathbf{F}$  and  $\Sigma_{\mathbf{h}}$ , we then estimate  $\mathbf{H}$  using minimum mean-squared error estimation. Again we focus exclusively on estimators for  $\mathbf{F}$  and  $\Sigma_{\mathbf{h}}$  in the next two sections; estimators for  $\mathbf{H}$  will be explored through numerical examples in Sec. 4.5.

### 4.3 Maximum-Likelihood Estimators

In this section, we derive maximum-likelihood (ML) estimators for  $\mathbf{F}$  and  $\Sigma_{\mathbf{h}}$  based on observations in (4.11). It is often convenient to find sufficient statistics before deriving the ML estimators.

We write (4.11) in matrix form, after defining  $\mathbf{X}_1 \triangleq [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K] \in \mathbb{C}^{N \times K}$  and  $\mathbf{X}_2 \triangleq [\mathbf{x}_{K+1}, \mathbf{x}_{K+2}, \dots, \mathbf{x}_T] \in \mathbb{C}^{N \times (T-K)}$ ,

$$\mathbf{W}_1 = \mathbf{H}\mathbf{X}_1 + \mathbf{N}_1, \quad \mathbf{W}_2 = \mathbf{F}\mathbf{H}\mathbf{X}_2 + \mathbf{N}_2, \quad (4.13)$$

where  $\mathbf{N}_1$  and  $\mathbf{N}_2$  are defined analogously, independent and have i.i.d. entries  $\mathcal{CN}(0, \sigma_n^2)$ .

The known training sequences for MIMO channel estimation are often equal-energy and orthogonal. We further assume  $K = T/2$  and equal-energy and orthogonal training

over the first and last  $K$  symbols i.e.,

$$\mathbf{X}_1 \mathbf{X}_1^H = \mathbf{X}_2 \mathbf{X}_2^H = \frac{PK}{N} \mathbf{I}_N, \quad (4.14)$$

This can be achieved by using a normalized discrete Fourier transform (DFT) matrix, e.g., [16, eq. 10]. We present a sufficient statistic in the next lemma.

**Lemma 4 (Sufficient Statistic)** *Consider the observations  $\mathbf{W}_1$  and  $\mathbf{W}_2$  defined in (4.13) and known training sequences in (4.14). Then*

$$\mathbf{Y}_1 = \left( \frac{2N}{PT} \right) \mathbf{W}_1 \mathbf{X}_1^H, \quad \mathbf{Y}_2 = \left( \frac{2N}{PT} \right) \mathbf{W}_2 \mathbf{X}_2^H, \quad (4.15)$$

*are sufficient for estimating unknown matrices  $\mathbf{F}$  and  $\Sigma_{\mathbf{h}}$ . Moreover,  $\mathbf{Y}_1 - \mathbf{H}$  and  $\mathbf{Y}_2 - \mathbf{FH}$  are independent random matrices with i.i.d.  $\mathcal{CN}(0, \sigma^2)$  entries, where  $\sigma^2 \triangleq 2N\sigma_n^2/PT$ .  $\diamond$*

**Proof** From (4.13) and (4.14), we have  $\mathbf{Y}_1 = \mathbf{H} + \left( \frac{2N}{PT} \right) \mathbf{N}_1 \mathbf{X}_1^H$ . To show the entries of the last matrix are i.i.d., we vectorize it,

$$\left( \frac{2N}{PT} \right) \text{vec}(\mathbf{N}_1 \mathbf{X}_1^H) = \left( \frac{2N}{PT} \right) (\mathbf{X}_1^* \otimes \mathbf{I}_M) \text{vec} \mathbf{N}_1 \in \mathbb{C}^{MN}, \quad (4.16)$$

which is zero-mean and has covariance matrix  $\left( \frac{2N}{PT} \right)^2 (\mathbf{X}_1^* \mathbf{X}_1^T \otimes \mathbf{I}_M) \sigma_n^2 \mathbf{I}_{MN} = \sigma^2 \mathbf{I}_{MN}$ , where by definition  $\sigma^2 = 2N\sigma_n^2/PT$ . Note an identity of Kronecker product  $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec} \mathbf{B}$  is used [19]. This shows that  $\mathbf{Y}_1 - \mathbf{H}$  is a random matrices with i.i.d.  $\mathcal{CN}(0, \sigma^2)$  entries. Similarly,  $\mathbf{Y}_2 - \mathbf{FH} = \left( \frac{2N}{PT} \right) \mathbf{N}_2 \mathbf{X}_2^H$  is also a random matrices with i.i.d.  $\mathcal{CN}(0, \sigma^2)$  entries. The independence between these two matrices follow from that noises are independent over time (4.13).

From the Neyman-Fisher theorem [45, pg. 117], to prove sufficiency of (4.15) it suffices

to show that  $p(\mathbf{W}_1, \mathbf{W}_2; \mathbf{F}, \boldsymbol{\Sigma}_h)$  factors into a product  $g(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{F}, \boldsymbol{\Sigma}_h) f(\mathbf{W}_1, \mathbf{W}_2)$ , where  $f$  does not depend on  $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{F}, \boldsymbol{\Sigma}_h$  and  $g$  does not depend on  $\mathbf{W}_1, \mathbf{W}_2$ . We prove this using the conditional pdf

$$p(\mathbf{W}_1, \mathbf{W}_2; \mathbf{F}, \boldsymbol{\Sigma}_h) = E_{\mathbf{H}} [p(\mathbf{W}_1, \mathbf{W}_2 | \mathbf{H}; \mathbf{F}, \boldsymbol{\Sigma}_h)] , \quad (4.17)$$

where the expectation  $E_{\mathbf{H}}[\cdot]$  is with respect to  $\mathbf{H}$  (4.9). Since  $\mathbf{W}_1$  and  $\mathbf{W}_2$  are conditionally independent given  $\mathbf{H}$ , we have

$$\begin{aligned} & (\pi\sigma_n^2)^{NT} p(\mathbf{W}_1, \mathbf{W}_2; \mathbf{F}, \boldsymbol{\Sigma}_h) \\ &= E_{\mathbf{H}} \left[ \exp \left( -\frac{1}{\sigma_n^2} \|\mathbf{W}_1 - \mathbf{H}\mathbf{X}_1\|^2 - \frac{1}{\sigma_n^2} \|\mathbf{W}_2 - \mathbf{F}\mathbf{H}\mathbf{X}_2\|^2 \right) \right] \\ &= E_{\mathbf{H}} \left[ \exp \left( \frac{\text{Tr} [\mathbf{W}_1^H \mathbf{H}\mathbf{X}_1]}{\sigma_n^2} + \frac{\text{Tr} [\mathbf{X}_1^H \mathbf{H}^H \mathbf{W}_1]}{\sigma_n^2} - \frac{\text{Tr} [\mathbf{X}_1^H \mathbf{H}^H \mathbf{H}\mathbf{X}_1]}{\sigma_n^2} \right. \right. \\ &\quad \left. \left. + \frac{\text{Tr} [\mathbf{W}_2^H \mathbf{F}\mathbf{H}\mathbf{X}_2]}{\sigma_n^2} + \frac{\text{Tr} [\mathbf{X}_2^H \mathbf{H}^H \mathbf{F}^H \mathbf{W}_2]}{\sigma_n^2} - \frac{\text{Tr} [\mathbf{X}_2^H \mathbf{H}^H \mathbf{F}^H \mathbf{F}\mathbf{H}\mathbf{X}_2]}{\sigma_n^2} \right) \right] \\ &\quad \exp \left( -\frac{1}{\sigma_n^2} \|\mathbf{W}_1\|^2 - \frac{1}{\sigma_n^2} \|\mathbf{W}_2\|^2 \right) \\ &= E_{\mathbf{H}} \left[ \exp \left( \frac{2 \text{Re Tr} [\mathbf{H}^H \mathbf{W}_1 \mathbf{X}_1^H]}{\sigma_n^2} - \frac{\text{Tr} [\mathbf{H}^H \mathbf{H}\mathbf{X}_1 \mathbf{X}_1^H]}{\sigma_n^2} + \frac{2 \text{Re Tr} [\mathbf{H}^H \mathbf{F}^H \mathbf{W}_2 \mathbf{X}_2^H]}{\sigma_n^2} \right. \right. \\ &\quad \left. \left. - \frac{\text{Tr} [\mathbf{H}^H \mathbf{F}^H \mathbf{F}\mathbf{H}\mathbf{X}_2 \mathbf{X}_2^H]}{\sigma_n^2} \right) \right] \exp \left( -\frac{1}{\sigma_n^2} \|\mathbf{W}_1\|^2 - \frac{1}{\sigma_n^2} \|\mathbf{W}_2\|^2 \right) \\ &= E_{\mathbf{H}} \left[ \exp \left( \frac{2 \text{Re Tr} [\mathbf{H}^H \mathbf{Y}_1 + \mathbf{H}^H \mathbf{F}^H \mathbf{Y}_2]}{\sigma^2} - \frac{\text{Tr} [\mathbf{H}^H \mathbf{H} + \mathbf{H}^H \mathbf{F}^H \mathbf{F}\mathbf{H}]}{\sigma^2} \right) \right] \\ &\quad \exp \left( -\frac{1}{\sigma_n^2} \|\mathbf{W}_1\|^2 - \frac{1}{\sigma_n^2} \|\mathbf{W}_2\|^2 \right) , \quad (4.18) \end{aligned}$$

where  $\|\mathbf{A}\|^2 = \text{Tr}[\mathbf{A}^H \mathbf{A}]$  denotes the Frobenius norm. Also, the third equality follows from the identities  $2 \text{Re Tr}[\mathbf{A}] = \text{Tr}[\mathbf{A}] + \text{Tr}[\mathbf{A}^H]$  and  $\text{Tr}[\mathbf{A}\mathbf{B}] = \text{Tr}[\mathbf{B}\mathbf{A}]$ , and the fourth

equality follows from (4.14) and the definition of  $\sigma^2$ . In (4.18), denote the first factor by  $(\pi\sigma_n^2)^{NT} g(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{F}, \boldsymbol{\Sigma}_h)$  and the second by  $f(\mathbf{W}_1, \mathbf{W}_2)$ . Note  $g$  depends on  $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{F}$  and  $\boldsymbol{\Sigma}_h$  (through the expectation) but not on  $\mathbf{W}_1, \mathbf{W}_2$ . And  $f$  depends on  $\mathbf{W}_1, \mathbf{W}_2$  only, not  $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{F}, \boldsymbol{\Sigma}_h$ . This completes the proof.

Based on the sufficient statistics in (4.15), we want to estimate the following complex parameters

$$\boldsymbol{\theta} \triangleq \text{vec} \begin{bmatrix} \mathbf{F} & \boldsymbol{\Sigma}_h \end{bmatrix}, \quad (4.19)$$

where  $\mathbf{F}$  is defined in (4.12) and  $\boldsymbol{\Sigma}_h$  in (4.10). Here we present the maximum-likelihood (ML) estimator, such that

$$\boldsymbol{\theta}_{ML} \triangleq \arg \max_{\boldsymbol{\theta}} p(\mathbf{Y}_1, \mathbf{Y}_2; \boldsymbol{\theta}). \quad (4.20)$$

The next theorem shows the ML estimator can be calculated via block eigen-decomposition using the sufficient statistic given in (4.15).

**Theorem 8 (Single-Packet ML Estimators)** *Let  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  be the sufficient statistics in (4.15). Suppose  $\mathbf{F}$  and  $\boldsymbol{\Sigma}_h$  are unknown. Consider the sample covariance matrix,*

$$\mathbf{S} \triangleq \frac{1}{N} \begin{bmatrix} \mathbf{Y}_1 \mathbf{Y}_1^H & \mathbf{Y}_1 \mathbf{Y}_2^H \\ \mathbf{Y}_2 \mathbf{Y}_1^H & \mathbf{Y}_2 \mathbf{Y}_2^H \end{bmatrix} \in \mathbb{C}^{2M \times 2M}. \quad (4.21)$$

*The eigen-decomposition of  $\mathbf{S}$  can be written as*

$$\mathbf{S} \mathbf{U}_s = \mathbf{U}_s \text{diag}(\mu_1, \dots, \mu_{2M}), \quad (4.22)$$

*where  $\text{diag}(\cdot)$  denotes a square diagonal matrix with its input as diagonal entries, and*

the eigen-values  $\mu_k \geq 0$  are in descending order. Define the unitary eigen-vector matrix  $\mathbf{U}_s$  as a 2 by 2 block matrix, i.e.,

$$\mathbf{U}_s \triangleq \begin{bmatrix} \mathbf{U}_{s11} & \mathbf{U}_{s12} \\ \mathbf{U}_{s21} & \mathbf{U}_{s22} \end{bmatrix}, \quad (4.23)$$

where  $\mathbf{U}_{sij} \in \mathbb{C}^{M \times M}$  and  $i, j = 1, 2$ . Then, the maximum-likelihood estimate of  $\boldsymbol{\theta}$  is,

$$\hat{\boldsymbol{\theta}}_{ML} \triangleq \text{vec} \begin{bmatrix} \hat{\mathbf{F}}_{ML} & \hat{\boldsymbol{\Sigma}}_{\mathbf{h}} \end{bmatrix}, \quad (4.24)$$

where  $\sigma^2 \triangleq 2N\sigma_n^2/PT$  and, conditioned on  $\mathbf{U}_{s11}$  is non-singular,

$$\hat{\mathbf{F}}_{ML} = \mathbf{U}_{s21}\mathbf{U}_{s11}^{-1}, \quad \hat{\boldsymbol{\Sigma}}_{\mathbf{h}} = \mathbf{U}_{s11} (\text{diag}(\mu_1, \dots, \mu_M) - \sigma^2 \mathbf{I}_M)^+ \mathbf{U}_{s11}^H. \quad (4.25)$$

Here  $(\cdot)^+$  is an element-wise operator on real matrices, such that  $[(\mathbf{A})^+]_{ij} \triangleq \max\{[\mathbf{A}]_{ij}, 0\}$ .

**Proof** Consider the sufficient statistic in (4.15), and define

$$\mathbf{V} \triangleq \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{H} + \mathbf{N}_1 \\ \mathbf{F}\mathbf{H} + \mathbf{N}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1 + \mathbf{n}_{1,1} & \cdots & \mathbf{h}_N + \mathbf{n}_{1,N} \\ \mathbf{F}\mathbf{h}_1 + \mathbf{n}_{2,1} & \cdots & \mathbf{F}\mathbf{h}_N + \mathbf{n}_{2,N} \end{bmatrix}. \quad (4.26)$$

Due to uncoupled transmit antennas, the  $N$  columns of  $\mathbf{H}$  are also i.i.d. zero-mean, circularly-symmetric, complex Gaussian random vectors, i.e.,  $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}_M, \boldsymbol{\Sigma}_{\mathbf{h}})$  for all  $1 \leq k \leq N$ , where  $\boldsymbol{\Sigma}_{\mathbf{h}} \triangleq E[\mathbf{h}_k \mathbf{h}_k^H]$ . Thus, the prior information for channel  $\mathbf{H}$  can be written as,

$$p(\mathbf{H}) = \frac{1}{\det(\pi \mathbf{I}_N \otimes \boldsymbol{\Sigma}_{\mathbf{h}})} \exp \left[ -(\text{vec } \mathbf{H})^H (\mathbf{I}_N \otimes \boldsymbol{\Sigma}_{\mathbf{h}}^{-1}) \text{vec } \mathbf{H} \right]. \quad (4.27)$$

The covariance matrix of  $\mathbf{V}$  is defined as,

$$\mathbf{\Sigma} \triangleq \frac{1}{N} E[\mathbf{V}\mathbf{V}^H] = \begin{bmatrix} \mathbf{\Sigma}_h + \sigma^2 \mathbf{I}_M & \mathbf{\Sigma}_h \mathbf{F}^H \\ \mathbf{F} \mathbf{\Sigma}_h & \mathbf{F} \mathbf{\Sigma}_h \mathbf{F}^H + \sigma^2 \mathbf{I}_M \end{bmatrix}. \quad (4.28)$$

Since  $\mathbf{\Sigma}$  is Hermitian, it decomposes into the following block eigen-system [62],

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{I}_M \\ \mathbf{F} \end{bmatrix} \mathbf{\Sigma}_h \begin{bmatrix} \mathbf{I}_M & \mathbf{F}^H \end{bmatrix} + \sigma^2 \mathbf{I}_{2M} = \mathbf{B}_1 \mathbf{D}_1 \mathbf{B}_1^H + \mathbf{B}_2 \mathbf{D}_2 \mathbf{B}_2^H, \quad (4.29)$$

where  $\mathbf{D}_i \in \mathbb{C}^{M \times M}$  and  $\mathbf{B}_i \in \mathbb{C}^{2M \times M}$  (for  $i = 1, 2$ ) are the block eigenvalues and orthonormal block eigenvectors of  $\mathbf{\Sigma}$ ,

$$\begin{aligned} \mathbf{B}_1 &= \begin{bmatrix} \mathbf{I}_M \\ \mathbf{F} \end{bmatrix} \mathbf{A}_1^{-\frac{1}{2}}, & \mathbf{B}_2 &= \begin{bmatrix} -\mathbf{F}^H \\ \mathbf{I}_M \end{bmatrix} \mathbf{A}_2^{-\frac{1}{2}}, \\ \mathbf{D}_1 &= \mathbf{A}_1^{\frac{1}{2}} \mathbf{\Sigma}_h \mathbf{A}_1^{\frac{1}{2}} + \sigma^2 \mathbf{I}_M, & \mathbf{D}_2 &= \sigma^2 \mathbf{I}_M, \end{aligned} \quad (4.30)$$

and  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are Hermitian matrices defined as,

$$\mathbf{A}_1 \triangleq \mathbf{F}^H \mathbf{F} + \mathbf{I}_M, \quad \mathbf{A}_2 \triangleq \mathbf{F} \mathbf{F}^H + \mathbf{I}_M. \quad (4.31)$$

Note both  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are positive definite; also  $\mathbf{F}^H \mathbf{A}_2 = \mathbf{A}_1 \mathbf{F}^H$  and  $\mathbf{F} \mathbf{A}_1 = \mathbf{A}_2 \mathbf{F}$ .

Since eigenvalues of  $\mathbf{D}_i$  are also eigenvalues of  $\mathbf{\Sigma}$  [62, Th. 1.1], and observe from (4.29) that  $\mathbf{D}_2$  is already diagonal, we write down the (scalar) eigenvalue decomposition of  $\mathbf{\Sigma}$ ,

$$\mathbf{\Sigma} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H \triangleq \begin{bmatrix} \mathbf{B}_1 \mathbf{Q} & \mathbf{B}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_1 & \mathbf{0}_{M \times M} \\ \mathbf{0}_{M \times M} & \mathbf{\Lambda}_2 \end{bmatrix} \begin{bmatrix} \mathbf{Q}^H \mathbf{B}_1^H \\ \mathbf{B}_2^H \end{bmatrix}, \quad (4.32)$$

where  $\mathbf{D}_1\mathbf{Q} = \mathbf{Q}\mathbf{\Lambda}_1$ ,  $\mathbf{\Lambda}_1 = \text{diag}(\lambda_1, \dots, \lambda_M)$  is diagonal,  $\mathbf{Q} \in \mathbb{C}^{M \times M}$  is unitary, and  $\mathbf{\Lambda}_2 = \sigma^2\mathbf{I}_M$ . Note  $\lambda_j \geq \sigma^2$  for all  $1 \leq j \leq M$ . In other words, the covariance matrix  $\mathbf{\Sigma}$  has  $2M$  real eigenvalues, where the largest  $M$  eigenvalues are greater than or equal to  $\sigma^2$ , while the smallest  $M$  eigenvalues are exactly  $\sigma^2$ .

Under the assumptions above, we assume  $\sigma_n^2$  is known but  $\mathbf{\Lambda}_1$  and  $\mathbf{U}$  are unknown. If we find the joint ML estimator for  $\mathbf{\Lambda}_1$  and  $\mathbf{U}$ , the ML estimators for  $\boldsymbol{\theta}$  (4.19) follows by the invariance principle of MLE [45, Th. 7.4, pg. 185]. From (4.32), we firstly find the inverse of covariance  $\mathbf{\Sigma}$  using its scalar eigen system,

$$\mathbf{\Sigma}^{-1} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^H. \quad (4.33)$$

We then write the density function of  $\mathbf{V} \triangleq [\mathbf{v}_1, \dots, \mathbf{v}_N]$  in (4.26) as,

$$\begin{aligned} p(\mathbf{V}; \boldsymbol{\theta}) &= \det(\pi\mathbf{\Sigma})^{-N} \exp\left(-\sum_{i=1}^N \mathbf{v}_i^H \mathbf{\Sigma}^{-1} \mathbf{v}_i\right) \\ &= \det(\pi\mathbf{\Sigma})^{-N} \exp(-N \text{Tr}[\mathbf{S}\mathbf{\Sigma}^{-1}]) = \left(\pi \prod_{j=1}^{2M} \lambda_j\right)^{-N} \exp\left(-N \sum_{j=1}^{2M} \frac{\mathbf{u}_j^H \mathbf{S} \mathbf{u}_j}{\lambda_j}\right) \\ &= \left(\pi \prod_{j=1}^{2M} \lambda_j\right)^{-N} \exp\left[N \sum_{j=1}^M \left(\frac{1}{\sigma^2} - \frac{1}{\lambda_j}\right) \mathbf{u}_j^H \mathbf{S} \mathbf{u}_j - \frac{N \text{Tr}(\mathbf{S})}{\sigma^2}\right], \end{aligned} \quad (4.34)$$

where in the second the sample covariance matrix defined in (4.21) can be equivalently written in terms of  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_N]$ ,

$$\mathbf{S} = \frac{1}{N} \mathbf{V}\mathbf{V}^H = \sum_{i=1}^N \mathbf{v}_i \mathbf{v}_i^H, \quad (4.35)$$

and in the third equality we define  $\mathbf{U} \triangleq [\mathbf{u}_1, \dots, \mathbf{u}_N]$ . Since we have  $\lambda_j \geq \sigma^2$  for  $1 \leq j \leq M$ , the coefficients  $\frac{1}{\sigma^2} - \frac{1}{\lambda_j}$  are non-negative and in decreasing order. The likelihood

function is maximized, when  $\mathbf{u}_j$  is chosen to be the eigenvector corresponding to the  $j$ th largest eigenvalue of the sample covariance  $\mathbf{S}$ . Assume the eigenvalues of  $\mathbf{S}$  are ordered in descending order,  $[\mu_1, \dots, \mu_{2M}]$ , the maximum of (4.34) is,

$$\begin{aligned} \max_{\mathbf{u}_j} p(\mathbf{V}) &= \left( \pi \prod_{j=1}^{2M} \lambda_j \right)^{-N} \exp \left( -N \sum_{j=1}^M \frac{\mu_j}{\lambda_j} - \frac{N}{\sigma^2} \sum_{k=M+1}^{2M} \mu_k \right) \\ &\leq \left( \pi \sigma^{2M} \prod_{j=1}^M \mu_j \right)^{-N} \exp \left( -NM - \frac{N}{\sigma^2} \sum_{k=M+1}^{2M} \mu_k \right), \end{aligned} \quad (4.36)$$

where the last step is because  $xe^{-x}$  is uniquely maximized at  $x = 1$ , and equality is achieved by letting

$$\hat{\lambda}_j = \max\{\mu_j, \sigma^2\}, \quad 1 \leq j \leq M. \quad (4.37)$$

Now it remains to show that (4.34) can be achieved by actual estimates of  $\mathbf{F}$  and  $\Sigma_{\mathbf{h}}$ . Assume that the  $\mathbf{S}\mathbf{U}_{\mathbf{s}} = \mathbf{U}_{\mathbf{s}} \text{diag}(\mu_1, \dots, \mu_{2M})$ , where

$$\mathbf{U}_{\mathbf{s}} \triangleq \begin{bmatrix} \mathbf{U}_{\mathbf{s}11} & \mathbf{U}_{\mathbf{s}12} \\ \mathbf{U}_{\mathbf{s}21} & \mathbf{U}_{\mathbf{s}22} \end{bmatrix}. \quad (4.38)$$

Then we verify that the following estimates achieve the maximum in (4.34),

$$\hat{\Sigma} = \mathbf{U}_{\mathbf{s}} \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_M, \sigma^2, \dots, \sigma^2) \mathbf{U}_{\mathbf{s}}^H. \quad (4.39)$$

Thus, the joint ML estimators for  $\mathbf{F}$  and  $\Sigma_{\mathbf{h}}$  are, conditioned on  $\mathbf{U}_{\mathbf{s}11}$  is non-singular,

$$\hat{\mathbf{F}}_{ML} = \mathbf{U}_{\mathbf{s}21} \mathbf{U}_{\mathbf{s}11}^{-1}, \quad \hat{\Sigma}_{\mathbf{h}} = \mathbf{U}_{\mathbf{s}11} (\text{diag}(\mu_1, \dots, \mu_M) - \sigma^2 \mathbf{I}_M)^+ \mathbf{U}_{\mathbf{s}11}^H. \quad (4.40)$$

The derivation for  $\hat{\mathbf{F}}_{ML} = \mathbf{U}_{\mathbf{s}21} \mathbf{U}_{\mathbf{s}11}^{-1}$  is straightforward from (4.29), as  $\mathbf{F}$  only depends

on the eigen-vector matrix and not the eigen-values. The derivation of  $\hat{\Sigma}_{\mathbf{h}}$  follows also (4.29),

$$\begin{aligned} \hat{\Sigma} - \sigma^2 \mathbf{I}_{2M} &= \begin{bmatrix} \mathbf{I}_M \\ \hat{\mathbf{F}} \end{bmatrix} \hat{\Sigma}_{\mathbf{h}} \begin{bmatrix} \mathbf{I}_M & \hat{\mathbf{F}}^H \end{bmatrix} \\ \begin{bmatrix} \mathbf{I}_M & \hat{\mathbf{F}}^H \end{bmatrix} (\hat{\Sigma} - \sigma^2 \mathbf{I}_{2M}) \begin{bmatrix} \mathbf{I}_M \\ \hat{\mathbf{F}} \end{bmatrix} &= (\mathbf{I}_M + \hat{\mathbf{F}}^H \hat{\mathbf{F}}) \hat{\Sigma}_{\mathbf{h}} (\mathbf{I}_M + \hat{\mathbf{F}}^H \hat{\mathbf{F}}) \\ \hat{\Sigma}_{\mathbf{h}} &= (\mathbf{I}_M + \hat{\mathbf{F}}^H \hat{\mathbf{F}})^{-1} \begin{bmatrix} \mathbf{I}_M & \hat{\mathbf{F}}^H \end{bmatrix} (\hat{\Sigma} - \sigma^2 \mathbf{I}_{2M}) \begin{bmatrix} \mathbf{I}_M \\ \hat{\mathbf{F}} \end{bmatrix} (\mathbf{I}_M + \hat{\mathbf{F}}^H \hat{\mathbf{F}})^{-1}, \end{aligned} \quad (4.41)$$

which reduces to (4.40) after plugging in  $\hat{\mathbf{F}}_{ML}$ 's formula and some simplification. This completes the proof.

In order to study the efficiency of the joint ML estimators for  $\mathbf{F}$  and  $\Sigma_{\mathbf{h}}$ , a natural next step is to find tight fundamental lower bounds on these estimators. Two fundamental lower bounds are investigated next.

It has been shown that the (complex) Fisher information matrix (FIM) on  $\boldsymbol{\theta}$  is given by [82, eq. 37], which extended results on real parameters to complex ones [45, eq. 15.52],

$$[\mathcal{I}(\boldsymbol{\theta})]_{ij} = N \text{Tr} \left[ \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i^*} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \right], \quad (4.42)$$

where  $\Sigma$  is given in (4.29). The error covariance matrix  $\mathbf{C}_{\hat{\boldsymbol{\theta}}}$  of any unbiased estimator  $\hat{\boldsymbol{\theta}}$  is lower bounded by the Cramér-Rao bound (CRB), i.e., the inverse of  $\mathcal{I}(\boldsymbol{\theta})$ ,

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} \triangleq E_{\mathbf{Y}_1, \mathbf{Y}_2; \boldsymbol{\theta}} \left[ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^H \right] \geq \mathcal{I}(\boldsymbol{\theta})^{-1}, \quad (4.43)$$

where the expectation is with respect to pdf in (4.34) and  $\mathbf{A} \geq \mathbf{B}$  means  $\mathbf{A} - \mathbf{B}$  is positive semi-definite. The formula in (4.43) is unlikely to simplify further without additional assumptions. Hence we evaluate the CRB numerically in simulations.

Another useful lower bound is the Miller-Chang bound (MCB) [56]. The formula for this bound is given below, but details of its derivation are given in the Appendix. It can be shown that square of Frobenius norm of any unbiased estimator for  $\mathbf{F}$ , for all  $\mathbf{H} \in \mathbb{C}^{M \times N}$  (5.5), is lower bounded by the MCB,

$$\begin{aligned} \mathcal{M}(\mathbf{F}) &= \sigma^2 \text{Tr} \left( E \left[ (\mathbf{H}^* \mathbf{H}^T)^{-1} \right] \right) \text{Tr} (\mathbf{F} \mathbf{F}^H + \mathbf{I}_M) \\ &= \begin{cases} \infty, & N \leq M, \\ \frac{\sigma^2}{N-M} \text{Tr} (\boldsymbol{\Sigma}_{\mathbf{h}}^{-1}) \text{Tr} (\mathbf{F} \mathbf{F}^H + \mathbf{I}_M), & N > M, \end{cases} \end{aligned} \quad (4.44)$$

where  $N > M$  means the number of antennas at the transmitter is greater than that at the receiver, and the final expression follows from that  $\mathbf{H}^* \mathbf{H}^T$  is a complex  $M \times M$  Wishart matrix of degree  $N$  (4.27) and the mean of its inverse is derived by Maiwald and Kraus [55, eq. 39]. Note the independence between columns of  $\mathbf{H}$  is essential. If  $N \leq M$ , the inverse mean of  $\mathbf{H}^* \mathbf{H}^T$  is unbounded, then one packet is likely insufficient for any unbiased estimator of  $\mathbf{F}$  to have finite error in Frobenius norm. This motivates finding estimators when observations from multiple packets are available.

## 4.4 Estimators for Multiple Packets

In the last section, we derived the maximum-likelihood estimators for  $\mathbf{F}$  and  $\boldsymbol{\Sigma}_{\mathbf{h}}$  using observations of training sequences from one packet. In this section, we consider estimators based on multiple packets, where channel varies from packet to packet.

As in the previous chapter, suppose the transmitter sends  $L$  identical training packets to the receiver. During transmission of each training packet, the receiver shifts its load impedance as described in (4.8). Similar to previous chapters and last section, block fading is assumed, i.e., the channel remains constant within a packet but randomly varies from packet to packet. Similarly to (4.11) the signal observations at the  $l$ -th packet can be described as

$$\mathbf{w}_{l,t} = \begin{cases} \mathbf{H}_l \mathbf{x}_t + \mathbf{n}_{l,t}, & 1 \leq t \leq K \\ \mathbf{F} \mathbf{H}_l \mathbf{x}_t + \mathbf{n}_{l,t}, & K + 1 \leq t \leq T \end{cases} \quad (4.45)$$

where the random noise vectors  $\mathbf{n}_{l,t} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_M)$  are independent over packets  $1 \leq l \leq L$  and time  $1 \leq t \leq T$ . We can express above observations in a compact matrix form, with a slight abuse of notation<sup>1</sup>,

$$\mathbf{W}_1 \triangleq \mathbf{H}(\mathbf{I}_L \otimes \mathbf{X}_1) + \mathbf{N}_1, \quad \mathbf{W}_2 \triangleq \mathbf{F} \mathbf{H}(\mathbf{I}_L \otimes \mathbf{X}_2) + \mathbf{N}_2, \quad (4.46)$$

where  $\mathbf{W}_1 \in \mathbb{C}^{M \times LK}$ ,  $\mathbf{W}_2 \in \mathbb{C}^{M \times L(T-K)}$ ,  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are defined above (4.13),  $\mathbf{N}_1$  and  $\mathbf{N}_2$  are independent random matrices with i.i.d. entries  $\mathcal{CN}(0, \sigma_n^2)$ , and we define the multi-packet channel as, again slightly abusing notation,

$$\mathbf{H} \triangleq [\mathbf{H}_1 \cdots \mathbf{H}_L]. \quad (4.47)$$

Here  $\mathbf{H}_l \in \mathbb{C}^{M \times N}$  is the channel matrix for the  $l$ -th packet, whose columns are spatially i.i.d. complex Gaussian  $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_h)$  across transmit antennas but temporally correlated across packets. If the normalized channel correlation is  $\mathbf{C}_h$ , then the space-time correla-

---

<sup>1</sup>Ideally, we would use notations like  $\mathbf{W}_{mp,1}$ ,  $\mathbf{H}_{mp}$ ,  $\mathbf{N}_{mp,1}$  and etc to distinguish them from their single-packet counterparts. We hereafter drop the subscript  $mp$  for simplicity when confusion is unlikely to occur.

tion of  $\mathbf{H}$  is better shown as

$$\text{vec } \mathbf{H} \sim \mathcal{CN}(\mathbf{0}_{MNL}, \mathbf{C}_h \otimes \mathbf{I}_N \otimes \boldsymbol{\Sigma}_h) . \quad (4.48)$$

In this section, we assume  $\mathbf{C}_h$  is known.

Similar to the previous section, the goal of this section is to derive estimators for both  $\mathbf{F}$  and  $\boldsymbol{\Sigma}_h$ , or  $\boldsymbol{\theta}$  as defined in (4.19), treating  $\mathbf{H}$  as a nuisance parameter. Then, we explore estimators for  $\mathbf{H}$  given  $\mathbf{F}$  and  $\boldsymbol{\Sigma}_h$  through numerical examples. The following lemma generalizes Lemma 4 to multiple packets.

**Lemma 5 (Multi-Packet Sufficient Statistic)** *Consider the observations  $\mathbf{W}_1$  and  $\mathbf{W}_2$  defined in (4.46) and known training sequences in (4.14). Then*

$$\mathbf{Y}_1 = \left(\frac{2N}{PT}\right) \mathbf{W}_1 (\mathbf{I}_L \otimes \mathbf{X}_1^H) , \quad \mathbf{Y}_2 = \left(\frac{2N}{PT}\right) \mathbf{W}_2 (\mathbf{I}_L \otimes \mathbf{X}_2^H) , \quad (4.49)$$

*are sufficient for estimating unknown matrices  $\mathbf{F}$  and  $\boldsymbol{\Sigma}_h$ . Moreover,  $\mathbf{Y}_1 - \mathbf{H}$  and  $\mathbf{Y}_2 - \mathbf{FH}$  are independent random matrices with i.i.d.  $\mathcal{CN}(0, \sigma^2)$  entries, where  $\sigma^2 \triangleq 2N\sigma_n^2/PT$ .  $\diamond$*

**Proof** From (4.13) and (4.14), we have  $\mathbf{Y}_1 = \mathbf{H} + \left(\frac{2N}{PT}\right) \mathbf{N}_1 (\mathbf{I}_L \otimes \mathbf{X}_1^H)$ . To show the entries of the last matrix are i.i.d., we vectorize it,

$$\left(\frac{2N}{PT}\right) \text{vec} [\mathbf{N}_1 \cdot (\mathbf{I}_L \otimes \mathbf{X}_1^H)] = \left(\frac{2N}{PT}\right) [(\mathbf{I}_L \otimes \mathbf{X}_1^*) \otimes \mathbf{I}_M] \text{vec } \mathbf{N}_1 \in \mathbb{C}^{MNL} , \quad (4.50)$$

which is zero-mean and has covariance matrix  $\left(\frac{2N}{PT}\right)^2 (\mathbf{I}_L \otimes \mathbf{X}_1^* \mathbf{X}_1^T \otimes \mathbf{I}_M) \sigma_n^2 \mathbf{I}_{MNL} = \sigma^2 \mathbf{I}_{MNL}$ .

Note Kronecker product is associative and  $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec } \mathbf{B}$  is used [19].

This shows that  $\mathbf{Y}_1 - \mathbf{H}$  is a random matrices with i.i.d.  $\mathcal{CN}(0, \sigma^2)$  entries. Similarly,

$\mathbf{Y}_2 - \mathbf{FH} = \left(\frac{2N}{PT}\right) \mathbf{N}_2 (\mathbf{I}_L \otimes \mathbf{X}_2^H)$  is also a random matrices with i.i.d.  $\mathcal{CN}(0, \sigma^2)$  entries.

The independence between these two matrices follow from that noises are independent over time and across packets (4.45).

From the Neyman-Fisher theorem [45, pg. 117], to prove sufficiency of (4.15) it suffices to show that  $p(\mathbf{W}_1, \mathbf{W}_2; \mathbf{F}, \Sigma_{\mathbf{h}})$  factors into a product  $g(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{F}, \Sigma_{\mathbf{h}}) f(\mathbf{W}_1, \mathbf{W}_2)$ , where  $f$  does not depend on  $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{F}, \Sigma_{\mathbf{h}}$  and  $g$  does not depend on  $\mathbf{W}_1, \mathbf{W}_2$ . We prove this using the conditional pdf

$$p(\mathbf{W}_1, \mathbf{W}_2; \mathbf{F}, \Sigma_{\mathbf{h}}) = E_{\mathbf{H}} [p(\mathbf{W}_1, \mathbf{W}_2 | \mathbf{H}; \mathbf{F}, \Sigma_{\mathbf{h}})] , \quad (4.51)$$

where the expectation  $E_{\mathbf{H}}[\cdot]$  is with respect to  $\mathbf{H}$ . Since  $\mathbf{W}_1$  and  $\mathbf{W}_2$  are conditionally

independent given  $\mathbf{H}$ , we have

$$\begin{aligned}
& (\pi\sigma_n^2)^{MLT} p(\mathbf{W}_1, \mathbf{W}_2; \mathbf{F}, \boldsymbol{\Sigma}_h) \\
= & E_{\mathbf{H}} \left[ \exp \left( -\frac{1}{\sigma_n^2} \|\mathbf{W}_1 - \mathbf{H}(\mathbf{I}_L \otimes \mathbf{X}_1)\|^2 - \frac{1}{\sigma_n^2} \|\mathbf{W}_2 - \mathbf{F}\mathbf{H}(\mathbf{I}_L \otimes \mathbf{X}_2)\|^2 \right) \right] \\
= & E_{\mathbf{H}} \left[ \exp \left( \frac{\text{Tr}[\mathbf{W}_1^H \mathbf{H}(\mathbf{I}_L \otimes \mathbf{X}_1)]}{\sigma_n^2} + \frac{\text{Tr}[(\mathbf{I}_L \otimes \mathbf{X}_1^H) \mathbf{H}^H \mathbf{W}_1]}{\sigma_n^2} \right. \right. \\
& \left. \left. - \frac{\text{Tr}[(\mathbf{I}_L \otimes \mathbf{X}_1^H) \mathbf{H}^H \mathbf{H}(\mathbf{I}_L \otimes \mathbf{X}_1)]}{\sigma_n^2} + \frac{\text{Tr}[\mathbf{W}_2^H \mathbf{F}\mathbf{H}(\mathbf{I}_L \otimes \mathbf{X}_2)]}{\sigma_n^2} \right. \right. \\
& \left. \left. + \frac{\text{Tr}[(\mathbf{I}_L \otimes \mathbf{X}_2^H) \mathbf{H}^H \mathbf{F}^H \mathbf{W}_2]}{\sigma_n^2} - \frac{\text{Tr}[(\mathbf{I}_L \otimes \mathbf{X}_2^H) \mathbf{H}^H \mathbf{F}^H \mathbf{F}\mathbf{H}(\mathbf{I}_L \otimes \mathbf{X}_2)]}{\sigma_n^2} \right) \right] \\
& \exp \left( -\frac{1}{\sigma_n^2} \|\mathbf{W}_1\|^2 - \frac{1}{\sigma_n^2} \|\mathbf{W}_2\|^2 \right) \\
= & E_{\mathbf{H}} \left[ \exp \left( \frac{2 \text{Re Tr}[\mathbf{H}^H \mathbf{W}_1(\mathbf{I}_L \otimes \mathbf{X}_1^H)]}{\sigma_n^2} - \frac{\text{Tr}[\mathbf{H}^H \mathbf{H}(\mathbf{I}_L \otimes \mathbf{X}_1 \mathbf{X}_1^H)]}{\sigma_n^2} \right. \right. \\
& \left. \left. + \frac{2 \text{Re Tr}[\mathbf{H}^H \mathbf{F}^H \mathbf{W}_2(\mathbf{I}_L \otimes \mathbf{X}_2^H)]}{\sigma_n^2} - \frac{\text{Tr}[\mathbf{H}^H \mathbf{F}^H \mathbf{F}\mathbf{H}(\mathbf{I}_L \otimes \mathbf{X}_2 \mathbf{X}_2^H)]}{\sigma_n^2} \right) \right] \\
& \exp \left( -\frac{1}{\sigma_n^2} \|\mathbf{W}_1\|^2 - \frac{1}{\sigma_n^2} \|\mathbf{W}_2\|^2 \right) \\
= & E_{\mathbf{H}} \left[ \exp \left( \frac{2 \text{Re Tr}[\mathbf{H}^H \mathbf{Y}_1 + \mathbf{H}^H \mathbf{F}^H \mathbf{Y}_2]}{\sigma^2} - \frac{\text{Tr}[\mathbf{H}^H \mathbf{H} + \mathbf{H}^H \mathbf{F}^H \mathbf{F}\mathbf{H}]}{\sigma^2} \right) \right] \\
& \exp \left( -\frac{1}{\sigma_n^2} \|\mathbf{W}_1\|^2 - \frac{1}{\sigma_n^2} \|\mathbf{W}_2\|^2 \right), \tag{4.52}
\end{aligned}$$

where  $\|\mathbf{A}\|^2 = \text{Tr}[\mathbf{A}^H \mathbf{A}]$  denotes the Frobenius norm. Also, the third equality follows from the identities  $2 \text{Re Tr}[\mathbf{A}] = \text{Tr}[\mathbf{A}] + \text{Tr}[\mathbf{A}^H]$  and  $\text{Tr}[\mathbf{A}\mathbf{B}] = \text{Tr}[\mathbf{B}\mathbf{A}]$ , and the fourth equality follows from (4.14) and the definition of  $\sigma^2$ . In (4.52), denote the first factor by  $(\pi\sigma_n^2)^{MLT} g(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{F}, \boldsymbol{\Sigma}_h)$  and the second by  $f(\mathbf{W}_1, \mathbf{W}_2)$ . Note  $g$  depends on  $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{F}$  and  $\boldsymbol{\Sigma}_h$  (through the expectation) but not on  $\mathbf{W}_1, \mathbf{W}_2$ . And  $f$  depends on  $\mathbf{W}_1, \mathbf{W}_2$  only, not  $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{F}, \boldsymbol{\Sigma}_h$ . This completes the proof.

Note  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  in (4.49) are a sufficient statistic regardless what the correlation matrix  $\mathbf{C}_h$  is. But  $\mathbf{C}_h$  will play a role in the density function after the expectation over  $\mathbf{H}$ . As the last section, our ultimate goal is to find the maximum-likelihood (ML) estimators for  $\boldsymbol{\theta}$ ,

$$\boldsymbol{\theta} \triangleq \text{vec} \begin{bmatrix} \mathbf{F} & \boldsymbol{\Sigma}_h \end{bmatrix}, \quad (4.53)$$

where  $\mathbf{F}$  is defined in (4.12) and  $\boldsymbol{\Sigma}_h$  in (4.10). Using the multi-packet sufficient statistics in (4.49), the multi-packet ML estimators for  $\boldsymbol{\theta}$  shall satisfy the following optimal criteria,

$$\boldsymbol{\theta}_{ML} \triangleq \arg \max_{\boldsymbol{\theta}} p(\mathbf{Y}_1, \mathbf{Y}_2; \boldsymbol{\theta}). \quad (4.54)$$

However, as we learned from the last chapter, these ML estimators are unlikely in closed-form in general. Thus, we defer discussion on the ML estimators but first seek another set of estimators, i.e., the *method of moments* estimators [45, Ch. 9].

**Lemma 6 (Method of Moments Estimators)** *Let  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  be the sufficient statistics in (4.49). Suppose  $\mathbf{F}$  and  $\boldsymbol{\Sigma}_h$  are unknown. Consider the sample covariance matrix,*

$$\mathbf{S}_{mp} \triangleq \frac{1}{NL} \begin{bmatrix} \mathbf{Y}_1 \mathbf{Y}_1^H & \mathbf{Y}_1 \mathbf{Y}_2^H \\ \mathbf{Y}_2 \mathbf{Y}_1^H & \mathbf{Y}_2 \mathbf{Y}_2^H \end{bmatrix} \in \mathbb{C}^{2M \times 2M}. \quad (4.55)$$

*The eigen-decomposition of  $\mathbf{S}$  can be written as*

$$\mathbf{S}_{mp} \mathbf{U}_s = \mathbf{U}_s \text{diag}(\mu_1, \dots, \mu_{2M}), \quad (4.56)$$

*where  $\text{diag}(\cdot)$  denotes a square diagonal matrix with its input as diagonal entries, and the eigen-values  $\mu_k \geq 0$  are in descending order. Define the unitary eigen-vector matrix*

$\mathbf{U}_s$  as a 2 by 2 block matrix, i.e.,

$$\mathbf{U}_s \triangleq \begin{bmatrix} \mathbf{U}_{s11} & \mathbf{U}_{s12} \\ \mathbf{U}_{s21} & \mathbf{U}_{s22} \end{bmatrix}, \quad (4.57)$$

where  $\mathbf{U}_{sij} \in \mathbb{C}^{M \times M}$  and  $i, j = 1, 2$ . Then, the method of moments (MM) estimates of  $\mathbf{F}$  and  $\Sigma_{\mathbf{h}}$  are, respectively, conditioned on  $\mathbf{U}_{s11}$  is non-singular,

$$\hat{\mathbf{F}}_{MM} = \mathbf{U}_{s21} \mathbf{U}_{s11}^{-1}, \quad \hat{\Sigma}_{\mathbf{h}MM} = \mathbf{U}_{s11} (\text{diag}(\mu_1, \dots, \mu_M) - \sigma^2 \mathbf{I}_M)^+ \mathbf{U}_{s11}^H, \quad (4.58)$$

where  $\sigma^2 \triangleq 2N\sigma_n^2/PT$  and the function  $(\cdot)^+$  is defined in Theorem 8.

**Proof** The multi-packet sufficient statistics can be collected as

$$\mathbf{V}_s \triangleq \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{H} \\ \mathbf{F}\mathbf{H} \end{bmatrix} + \mathbf{N}_s \in \mathbb{C}^{2M \times NL}, \quad (4.59)$$

where the noise is i.i.d., that is  $\text{vec } \mathbf{N}_s \sim \mathcal{CN}(\mathbf{0}_{2MNL}, \sigma^2 \mathbf{I}_{2MNL})$ , as proven in Lemma 5.

It is straightforward to show that

$$E[\mathbf{S}_{mp}] = \frac{1}{NL} E[\mathbf{V}_s \mathbf{V}_s^H] = \begin{bmatrix} \Sigma_{\mathbf{h}} + \sigma^2 \mathbf{I}_M & \Sigma_{\mathbf{h}} \mathbf{F}^H \\ \mathbf{F} \Sigma_{\mathbf{h}} & \mathbf{F} \Sigma_{\mathbf{h}} \mathbf{F}^H + \sigma^2 \mathbf{I}_M \end{bmatrix}. \quad (4.60)$$

Apparently, the second moments of sufficient statistics  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  are functions of unknown parameters  $\boldsymbol{\theta}$ , or  $\mathbf{F}$  and  $\Sigma_{\mathbf{h}}$  (4.19),

$$E[\mathbf{S}_{mp}] = \mathbf{T}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I}_M \\ \mathbf{F} \end{bmatrix} \Sigma_{\mathbf{h}} \begin{bmatrix} \mathbf{I}_M & \mathbf{F}^H \end{bmatrix} + \sigma^2 \mathbf{I}_{2M}, \quad (4.61)$$

where  $\mathbf{T} : \mathbb{C}^{2M^2} \rightarrow \mathbb{C}^{2M \times 2M}$  denotes a mapping. Then, from basic principles of MM estimation [45, Sec. 9.4], we find  $\hat{\boldsymbol{\theta}}_{MM}$  by the inverse of aforementioned mapping,

$$\hat{\boldsymbol{\theta}}_{MM} = \mathbf{T}^{-1}(\mathbf{S}_{mp}) . \quad (4.62)$$

The formula of MM estimates in (4.58) follows directly from the proof of Theorem 8. This completes the proof.

The MM estimators are provably consistent, easy to determine, and does not require knowing  $\mathbf{C}_{\mathbf{h}}$ . However, they are generally sub-optimal to the corresponding ML estimators [45, Ch. 9]. Next we show a special case where the ML estimators and the MM estimators coincide, and then discuss how to find the ML estimators in general fading conditions if  $\mathbf{C}_{\mathbf{h}}$  is known.

**Corollary 3 (ML Estimators for Fast Fading)** *If the Rayleigh fading channel is temporally i.i.d., that is  $\mathbf{C}_{\mathbf{h}} = \mathbf{I}_L$  in (4.48), then the multi-packet MM estimators given in (4.58) for  $\mathbf{F}$  and  $\boldsymbol{\Sigma}_{\mathbf{h}}$  satisfy (4.54) and hence are the maximum-likelihood estimators.*

**Proof** Conditioned on  $\mathbf{C}_{\mathbf{h}} = \mathbf{I}_L$ , the channel matrix  $\mathbf{H}$  in (4.48) would satisfy,

$$\text{vec } \mathbf{H} \sim \mathcal{CN}(\mathbf{0}_{MNL}, \mathbf{I}_{NL} \otimes \boldsymbol{\Sigma}_{\mathbf{h}}) . \quad (4.63)$$

Note the similarity between the distribution of this multi-packet  $H$  and the distribution the single-packet channel matrix follows in (4.27). The proof follows Theorem 8, except the number of i.i.d. columns in  $\mathbf{H}$  is  $NL$  rather  $N$ .

Next we discuss finding the ML estimator  $\hat{\boldsymbol{\theta}}_{ML}$  under general fading conditions. We assume the temporal correlation  $\mathbf{C}_{\mathbf{h}}$  is known, and  $L$  packets can be decorrelated by its

eigen-vector matrix  $\mathbf{Q}$ , i.e.,

$$\mathbf{Q}^H \mathbf{C}_h \mathbf{Q} = \mathbf{D} \triangleq \text{diag}(d_1, \dots, d_L) . \quad (4.64)$$

Note  $\mathbf{C}_h$  is normalized with 1's on its diagonal, i.e.,  $\text{Tr}[\mathbf{C}_h] = \text{Tr}[\mathbf{D}] = L$ . Consider the decorrelated observation, i.e.,

$$\mathbf{V} \triangleq \mathbf{V}_s (\mathbf{Q}^* \otimes \mathbf{I}_N) = \begin{bmatrix} \mathbf{H}_d \\ \mathbf{F}\mathbf{H}_d \end{bmatrix} + \mathbf{N} \in \mathbb{C}^{2M \times NL} , \quad (4.65)$$

where  $\text{vec } \mathbf{N} = [(\mathbf{Q}^* \otimes \mathbf{I}_N) \otimes \mathbf{I}_{2M}] \text{vec } \mathbf{N}_s \sim \mathcal{CN}(\mathbf{0}_{2MNL}, \sigma^2 \mathbf{I}_{2MNL})$  is i.i.d. and,

$$\text{vec } \mathbf{H}_d \sim \mathcal{CN}(\mathbf{0}_{MNL}, \mathbf{D} \otimes \mathbf{I}_N \otimes \Sigma_h) . \quad (4.66)$$

To understand this, for each  $1 \leq k \leq L$ , we have  $N$  i.i.d. complex Gaussian random vectors that follow  $\mathcal{CN}(\mathbf{0}_M, d_k \Sigma_h)$ , where  $d_k$  are defined in (4.64).

The pdf of  $\mathbf{V} \triangleq [\mathbf{v}_1, \dots, \mathbf{v}_{NL}]$  is

$$\begin{aligned} p(\mathbf{V}; \boldsymbol{\theta}) &= \prod_{k=1}^L \det(\pi \Sigma_k)^{-N} \exp \left( - \sum_{i=1}^N \mathbf{v}_{(k-1)N+i}^H \Sigma_k^{-1} \mathbf{v}_{(k-1)N+i} \right) \\ &= \prod_{k=1}^L \det(\pi \Sigma_k)^{-N} \cdot \exp \left( -N \text{Tr}[\mathbf{S}_k \Sigma_k^{-1}] \right) , \end{aligned} \quad (4.67)$$

where we define for  $1 \leq k \leq L$ ,

$$\mathbf{S}_k \triangleq \frac{1}{N} \sum_{i=1}^N \mathbf{v}_{(k-1)N+i} \mathbf{v}_{(k-1)N+i}^H , \quad (4.68)$$

and

$$\boldsymbol{\Sigma}_k = \begin{bmatrix} d_k \boldsymbol{\Sigma}_{\mathbf{h}} + \sigma^2 \mathbf{I}_M & d_k \boldsymbol{\Sigma}_{\mathbf{h}} \mathbf{F}^H \\ d_k \mathbf{F} \boldsymbol{\Sigma}_{\mathbf{h}} & d_k \mathbf{F} \boldsymbol{\Sigma}_{\mathbf{h}} \mathbf{F}^H + \sigma^2 \mathbf{I}_M \end{bmatrix}. \quad (4.69)$$

In order to find the ML estimators, we define the log-likelihood function as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &\triangleq \ln p(\boldsymbol{\theta}; \mathbf{V}) = \ln \left[ \prod_{k=1}^L \det(\pi \boldsymbol{\Sigma}_k)^{-N} \cdot \exp(-N \operatorname{Tr}[\mathbf{S}_k \boldsymbol{\Sigma}_k^{-1}]) \right] \\ &= C - \sum_{k=1}^L (\ln \det \boldsymbol{\Sigma}_k + N \operatorname{Tr}[\mathbf{S}_k \boldsymbol{\Sigma}_k^{-1}]), \end{aligned} \quad (4.70)$$

where  $C$  is a constant independent of  $\boldsymbol{\theta}$ . A necessary condition for joint ML estimators is that the complex gradient vanishes, i.e.,

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} = \mathbf{0}. \quad (4.71)$$

However, it is unclear if a closed-form solution to this condition exists. Thus, we use numerical methods to find these ML estimators. In particular, a variation of the complex Newton-Raphson<sup>2</sup> iteration can be used to find the true MLE [83, eq. 11],

$$\boldsymbol{\theta}_{p+1} = \boldsymbol{\theta}_p - \mathcal{H}^{-1}(\boldsymbol{\theta}_p) \left. \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_p}, \quad (4.72)$$

where the complex Hessian of the log-likelihood function is,

$$\mathcal{H}(\boldsymbol{\theta}) \triangleq \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^* \partial \boldsymbol{\theta}^T}. \quad (4.73)$$

---

<sup>2</sup>This is called the quasi-Newton method in a tutorial on Wirtinger calculus (or  $\mathbb{C}\mathbb{R}$  calculus) [47].

The multi-packet FIM follows directly from (4.42),

$$[\mathcal{I}_{mp}(\boldsymbol{\theta})]_{ij} = N \cdot \sum_{k=1}^L \text{Tr} \left[ \boldsymbol{\Sigma}_k^{-1} \frac{\partial \boldsymbol{\Sigma}_k}{\partial \theta_i^*} \boldsymbol{\Sigma}_k^{-1} \frac{\partial \boldsymbol{\Sigma}_k}{\partial \theta_j} \right], \quad (4.74)$$

where  $\boldsymbol{\Sigma}_k$  is the  $k$ -th covariance matrix defined in (4.69). Similarly, the error covariance matrix  $\mathbf{C}_{\hat{\boldsymbol{\theta}}}$  of any unbiased estimator  $\hat{\boldsymbol{\theta}}$  is lower bounded by the Cramér-Rao bound (CRB), which is the inverse of  $\mathcal{I}_{mp}(\boldsymbol{\theta})$ ,

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} \triangleq E_{\mathbf{Y}_1, \mathbf{Y}_2; \boldsymbol{\theta}} \left[ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^H \right] \geq \mathcal{I}_{mp}(\boldsymbol{\theta})^{-1}, \quad (4.75)$$

where the expectation is with respect to pdf in (4.67) and  $\mathbf{A} \geq \mathbf{B}$  means  $\mathbf{A} - \mathbf{B}$  is positive semi-definite.

For any estimator of  $\mathbf{F}$ , we find an estimator for antenna impedance via (4.12), i.e.,

$$\hat{\mathbf{Z}}_{\mathbf{A}} = \left( \mathbf{Z}_1 - \mathbf{Z}_2 \mathbf{R}_2^{-1/2} \hat{\mathbf{F}}_{MM} \mathbf{R}_1^{1/2} \right) \left( \mathbf{R}_2^{-1/2} \hat{\mathbf{F}}_{MM} \mathbf{R}_1^{1/2} - \mathbf{I}_M \right)^{-1}. \quad (4.76)$$

However, due to the reciprocity theorem of electromagnetics [11, pg. 144],  $\mathbf{Z}_{\mathbf{A}}$  is symmetric and so should any reasonable estimate of it. Here we replace  $\hat{\mathbf{Z}}_{\mathbf{A}}$  by its nearest symmetric matrix, i.e.,

$$\tilde{\mathbf{Z}}_{\mathbf{A}} \triangleq \frac{1}{2} \left( \hat{\mathbf{Z}}_{\mathbf{A}} + \hat{\mathbf{Z}}_{\mathbf{A}}^T \right). \quad (4.77)$$

Based on this new estimate, the receiver matches its load impedance for minimum noise-figure, which reduces to maximum power transfer under our noise model [25, eq. 10],  $\hat{\mathbf{Z}}_L = \tilde{\mathbf{Z}}_{\mathbf{A}}^*$ . Consequently, we calculate an excess (transmit) power needed for this matching

compared to the truly optimal one, i.e.,  $\mathbf{Z}_{L,opt} = \mathbf{Z}_A^*$ ,

$$10 \log_{10} \left( \text{Tr} [(4\mathbf{R}_A)^{-1} \boldsymbol{\Sigma}_g] / E \text{Tr} \left[ \left( \tilde{\mathbf{Z}}_A^* + \mathbf{Z}_A \right)^{-H} \tilde{\mathbf{R}}_A \left( \tilde{\mathbf{Z}}_A^* + \mathbf{Z}_A \right)^{-1} \boldsymbol{\Sigma}_g \right] \right), \quad (4.78)$$

where  $\tilde{\mathbf{R}}_A \triangleq \text{Re}\{\tilde{\mathbf{Z}}_A\}$ .

In the next section, we compare the performance of estimators derived in this chapter against their corresponding lower bounds, and explore the potential benefits of these estimators on system-level metrics, such as channel capacity.

## 4.5 Numerical Results

In this section, we explore the performance of estimators in the previous section through numerical examples. Consider a narrow-band MIMO communications system with  $N = 4$  transmit antennas and  $M = 2$  receive antennas, whose carrier frequency is 2.1 GHz. This frequency is chosen based on the first E-UTRA down-link operating band in LTE specifications [82]. The duration of each data packet equals to a subframe of LTE, i.e.,  $T_s = 1$  ms. Block-fading channel is assumed, such that during one data packet, the channel information remains the same, but it generally varies from packet to packet [16].

For each data packet, a training sequence precedes data sequence [54, Fig. 1(a)]. We take the two partitions of the training sequence  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$  in (4.14) from a normalized discrete Fourier transform (DFT) matrix of dimension  $K = T/2 = 32$ , e.g., [16, eq. 10]. In particular, the first part  $\mathbf{X}_1$  is chosen as the first  $N$  rows, while  $\mathbf{X}_2$  the next  $N$  rows, and  $\mathbf{X}_i \mathbf{X}_i^H = K \mathbf{I}_N$  for  $i = 1, 2$ . The unknown antenna impedance is that of a uniform

linear array (ULA) [27], i.e.,

$$\mathbf{Z}_A = \begin{bmatrix} 72.8521 + j1.6869 & -15.7457 - j27.8393 \\ -15.7457 - j27.8393 & 72.8521 + j1.6869 \end{bmatrix} \Omega. \quad (4.79)$$

The load impedance is  $\mathbf{Z}_1 = 50\mathbf{I}_M \Omega$  for the first  $K = 32$  symbols of each training sequence, and  $\mathbf{Z}_2 = (50 + j20)\mathbf{I}_M + 10 \times \mathbf{1} \Omega$  for the remaining  $T - K = 32$  symbols, where  $\mathbf{1}$  is the  $M$  by  $M$  all one matrix. From (4.12), it follows that

$$\mathbf{F} = \begin{bmatrix} 0.9804 - j0.1613 & 0.0261 - j0.0334 \\ 0.0261 - j0.0334 & 0.9804 - j0.1613 \end{bmatrix}. \quad (4.80)$$

In this section, we explore important properties of the estimators derived in previous sections. The average post-detection SNR of a received symbol is defined from (4.6) as [16, Sec. VIII],

$$\gamma \triangleq \frac{E \text{Tr} [\mathbf{H}\mathbf{x}\mathbf{x}^H \mathbf{H}^H]}{E \text{Tr} [\mathbf{n}_L \mathbf{n}_L^H]} = \frac{P\sigma_H^2}{\sigma_n^2}, \quad (4.81)$$

where  $\sigma_n^2$  is the noise variance at each port of the  $M$ -port receiver and  $\sigma_H^2$  is the mean of diagonal entries of  $\mathbf{\Sigma}_h$  in (4.27),

$$\sigma_H^2 \triangleq \frac{1}{M} \text{Tr}[\mathbf{\Sigma}_h]. \quad (4.82)$$

As shown in Fig. 4.2, the relative root mean-square error (RMSE) is plotted against SNR (4.81). The ML estimator in  $\hat{\mathbf{F}}_{ML}$ , for a given  $L = 5$  packets, becomes efficient as the number of transmit antenna increases, i.e., more spatial diversity. We also observe that the Miller-Chang bound (MCB) is tighter than the CRB and touches the RMSE for all values of  $L$  and SNR plotted in Fig. 4.2. Although the ML estimators are asymptotically

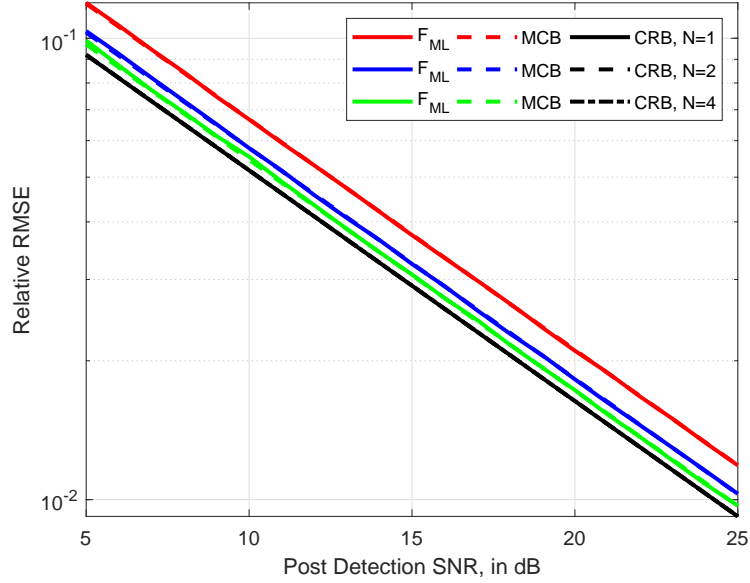


Figure 4.2: Relative MSE of  $\hat{\mathbf{F}}_{ML}$  versus SNR in i.i.d. Fading,  $L = 5$ .

unbiased and efficient, i.e., it achieves its corresponding CRB, the MCB (if exists) better predicts the RMSE of  $\hat{\mathbf{F}}_{ML}$  for finite sample size in  $L$ . For different  $N$  the CRB generally have different values as indicated by the formulas of Fisher information matrix in (4.42), but their numerical evaluations seem indistinguishable in Fig. 4.2.

Next, we investigate the performance of estimators derived previously under different Rayleigh fading conditions, i.e., fast, medium, and slow fading [82]. In particular, Clarke's model is assumed [9, 21, 86] and the normalized channel correlation matrix is,

$$\mathbf{C}_h = \begin{bmatrix} R[0] & R[-1] & \cdots & R[-L+1] \\ R[1] & R[0] & \cdots & R[-L+2] \\ \vdots & \ddots & \ddots & \vdots \\ R[L-1] & R[L-2] & \cdots & R[0] \end{bmatrix}, \quad (4.83)$$

where  $R[l] = J_0(2\pi f_d T_s |l|)$ ,  $J_0(\cdot)$  is the zeroth-order Bessel function of the first kind,

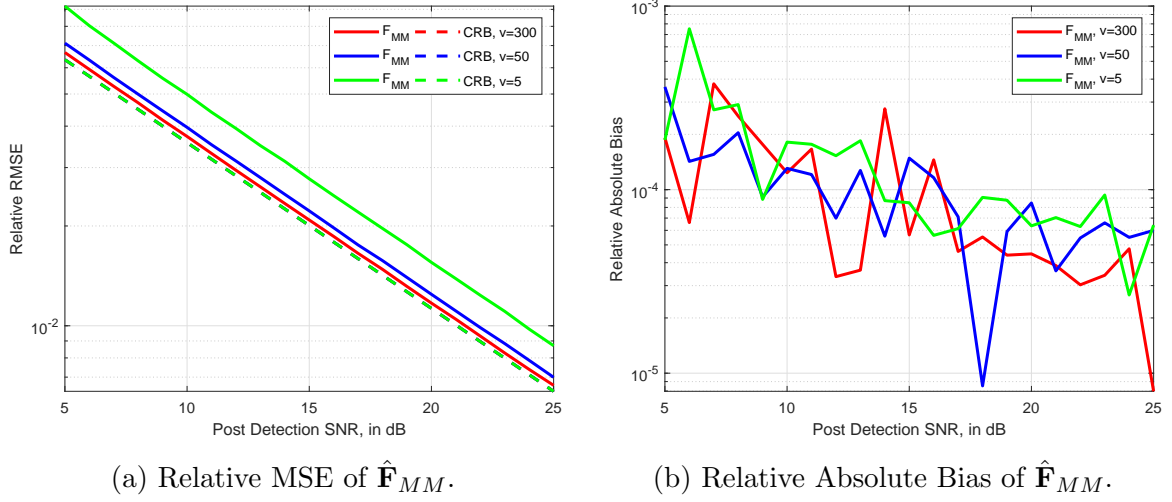


Figure 4.3: Properties of  $\hat{\mathbf{F}}_{MM}$  for a 4 by 2 MIMO,  $L = 10$

$T_s = 1$  ms is the sampling interval, and  $l$  is the sample difference. The fading frequency (maximum Doppler frequency) is  $f_d \triangleq v/\lambda$ , where  $v$  is the velocity of the fastest moving scatterer and  $\lambda$  the wave-length of the carrier frequency.

In Fig. 4.3a, we plot the relative RMSE of the method of moments (MM) estimators  $\hat{\mathbf{F}}_{MM}$  (4.58), for a MIMO with  $N = 4$  transmit and  $M = 2$  receiver antennas. The velocity of the fastest moving scatterer is  $v = 300, 50$  and  $5$  km/h, which represents a fast, medium, and slow fading scenario, respectively. The MM estimator  $\hat{\mathbf{F}}_{MM}$  is about 3 dB away from its CRB under slow fading, and this gap narrows to less 1 dB for medium and fast fading. Thus, a faster fading results in improved estimation accuracy, which is always lower bounded by their corresponding CRB. This is intuitively reasonable as fast fading means more temporal diversity. Similar to Fig. 4.2, the CRB depends very little on fading conditions; for the three drastically different cases considered, their evaluated CRB's are indistinguishable.

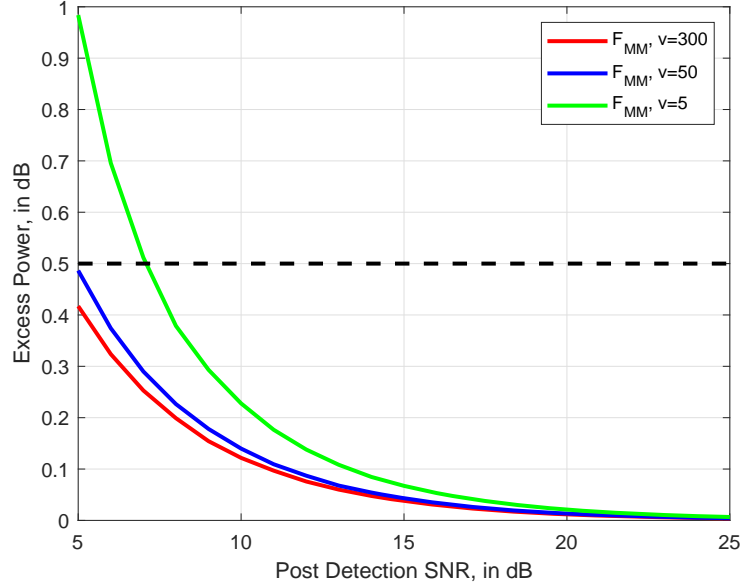


Figure 4.4: Excess Power of  $\hat{\mathbf{F}}_{MM}$  for 4 by 2 MIMO.

Plotted in Fig. 4.3b is the relative absolute bias of these MM estimators, i.e.,

$$\frac{E\|\hat{\mathbf{F}}_{MM} - \mathbf{F}\|_1}{\|\mathbf{F}\|_1} = \frac{\sum \left( E \left| \text{vec} \left( \hat{\mathbf{F}}_{MM} - \mathbf{F} \right) \right| \right)}{\sum |\text{vec } \mathbf{F}|}, \quad (4.84)$$

where  $\|\cdot\|_1$  is the entry-wise  $\ell_1$  norm, and  $\sum(\cdot)$  represents the summation of all entries of a vector. Random fluctuation is observed for all data points plotted. This suggests the true bias, if any, is below these curves. Although this figure does not show  $\hat{\mathbf{F}}_{MM}$  is unbiased, its bias is small and can be ignored practically.

Next we evaluate the excess power defined in (4.78). A faster fading channel results in a smaller excess power. This means the transmitter may save power for an intended receive SNR, due to an improved impedance match between antenna and load. For example, the gain between our fast and slow fading cases is about 3 dB at low SNR. If a 0.5 dB excess power is considered good match in practice, then this goal is achieved at relatively low

SNR for all fading conditions. Further, the excess power vanishes to 0 dB at high SNR, suggesting the receiver is capable of compensating for its impedance mismatch quickly, e.g., over  $L = 10$  packets or 10 ms. Next we give two examples which demonstrates the benefits of this impedance estimation algorithm in terms of ergodic capacity.

As derived by Hassibi and Hochwald, a lower bound on (ergodic) capacity exists, which incorporates the MMSE channel estimation error [37, eq. 21], i.e.,

$$C_l = E \left[ \log_2 \det \left( \mathbf{I}_M + \gamma_{\text{eff}} \cdot \frac{1}{N} \boldsymbol{\Sigma}_{\tilde{\mathbf{h}}} \mathbf{H}_w \mathbf{H}_w^H \right) \right], \quad (4.85)$$

where  $\text{vec } \mathbf{H}_w \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{MN})$ ,  $\boldsymbol{\Sigma}_{\tilde{\mathbf{h}}}$  is the normalized version of  $\boldsymbol{\Sigma}_{\mathbf{h}}$  such that  $\text{Tr}[\boldsymbol{\Sigma}_{\tilde{\mathbf{h}}}] / M = 1$ , and with (4.81) the effective SNR is defined as,

$$\gamma_{\text{eff}} = \frac{P\sigma_H^2}{\sigma_n^2} \frac{PT\sigma_H^2}{PT\sigma_H^2 + N(P\sigma_H^2 + \sigma_n^2)} = \gamma \frac{1}{1 + (1 + 1/\gamma)N/T}. \quad (4.86)$$

Shin and Lee derived an upper bound for this ergodic capacity in closed-form, putting the expectation between  $\log_2(\cdot)$  and  $\det(\cdot)$  by Jensen's inequality [71, Th. III.2], i.e.,

$$C_l \leq \log_2 \left( \sum_{k=0}^M \left[ \left( \frac{\gamma_{\text{eff}}}{N} \right)^k k! \sigma_k(\boldsymbol{\Sigma}_{\tilde{\mathbf{h}}}) \cdot \sigma_k(\mathbf{I}_N) \right] \right), \quad (4.87)$$

where  $M \leq N$  is assumed and  $\sigma_k(\mathbf{A})$  denotes the sum of all the  $k$ -rowed principal minor determinants of a square matrix  $\mathbf{A}$  [41, pg. 17]. In particular, we have [71, Th. II.3],

$$\sigma_k(\mathbf{I}_N) = \binom{N}{k} = \frac{N!}{k!(N-k)!}. \quad (4.88)$$

Consider a 4 by 2 MIMO system again, i.e.,  $N = 4$  and  $M = 2$ . This ergodic capacity

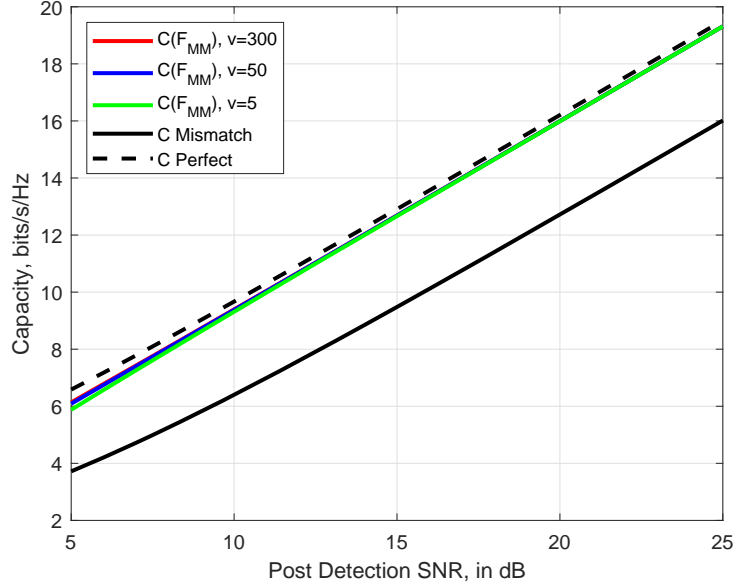


Figure 4.5: Ergodic Capacity over SNR for 4 by 2 MIMO.

upper bound boils down to

$$C_l \leq \log_2 \left[ 1 + 2\gamma_{\text{eff}} + \frac{3}{4}\gamma_{\text{eff}}^2 \cdot \det(\Sigma_{\hat{\mathbf{h}}}) \right]. \quad (4.89)$$

Although calculating an upper bound instead of the exact ergodic capacity is less than ideal, it should qualitatively demonstrate the capacity boost using our proposed antenna impedance estimation algorithm.

In Fig. 4.5, ergodic capacity upper bound (4.89) are plotted against SNR of an originally mismatched receiver. The power loss due to mismatch is chosen as 5dB. After applying our algorithm and matching to the estimate of  $\mathbf{Z}_{\mathbf{A}}$  (4.77), a significant gain on ergodic capacity  $C(\hat{\mathbf{F}}_{MM})$  is observed, compared to the mismatched receiver (the black solid line). This gain ranges from about 50% at low SNR to 20% at high SNR. The black dash line represents an upper bound on ergodic capacity, where the receiver is always

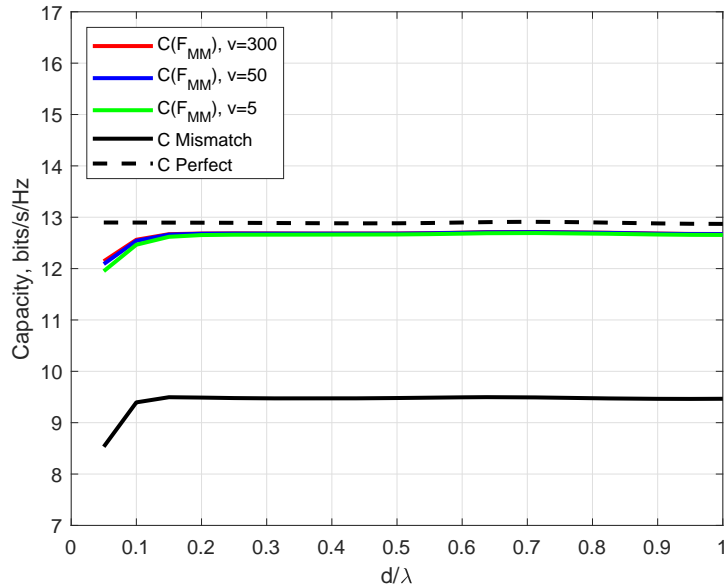


Figure 4.6: Ergodic Capacity over  $d/\lambda$  for 4 by 2 MIMO, SNR = 10dB.

optimally matched and observes the channel without errors. This upper bound, although unachievable by any practical system, is closed in by  $C(\hat{\mathbf{F}}_{MM})$  to around 1 dB or less for all SNR and fading conditions considered. Also note at low SNR, faster fading leads to a marginally capacity boost, which vanishes as SNR increases.

Plotted in Fig. 4.6 is the ergodic capacity upper bound in (4.89) versus antenna element-separation  $d/\lambda$ . The SNR for the originally mismatched receiver is fixed at 10 dB, while other settings remain identical as in Fig. 4.5. Similar observations are also made here, as the  $C(\hat{\mathbf{F}}_{MM})$ 's hone in the practically unachievable upper bound (the black dash line) within a fraction of 1 bit/s/Hz. This upper bound seems to depend very little on antenna spacing, yet the other capacity curves tend to drop for closely spaced arrays. Compared to the mismatched receiver, our algorithm improves capacity by over 30% for all data points.

## 4.6 Conclusion

In this chapter, we derived the maximum-likelihood (ML) and method of moments (MM) estimators for MIMO antenna impedance using training sequences in various fading conditions. In particular, under i.i.d. fading, the ML estimator is derived as the ratio of the top block eigen-vector of the sample covariance matrix. We also derive two fundamental lower bounds on these estimators, and explore the performance of these estimators through numerical examples. The ML and MM estimators become efficient (to CRB) when sufficient spatial and/or temporal diversity exists. A typical rule of thumb is the number of diversity is four times the number of receive antennas. Additionally, trade-off between channel correlation and impedance estimation accuracy is investigated. Our numerical results indicate that the MIMO antenna impedance can be accurately estimated in a matter of milliseconds. This estimate is able to compensate partially at low SNR and almost all power loss due to mismatch. In the example of ergodic capacity, if the original mismatch at the receiver is significant, large capacity boost can be observed in general.

# Chapter 5

## Likelihood Ratio Tests for Detecting Antenna Impedance Variations

### 5.1 Introduction

In previous chapters, we have developed estimation techniques that learn in real-time the unknown antenna impedance using observations of training sequences in multiple-input, multiple-output (MIMO) receivers. However, it remains unclear what triggers the receiver to perform this estimation. In other words, the following question needs to be answered: *How does the receiver know that antenna impedance has changed and impedance estimation becomes necessary?* In this chapter, we address this problem by formulating it as a hypothesis test. In this test, we decide if the covariance matrices of observations remain the same or not, over two consecutive blocks of packets. In particular, we apply well known likelihood ratio tests to tackle this problem.

The rest of this chapter is organized as follows. We present the system model in Sec. 5.2, and formulate this problem using hypothesis testing and develop a detector

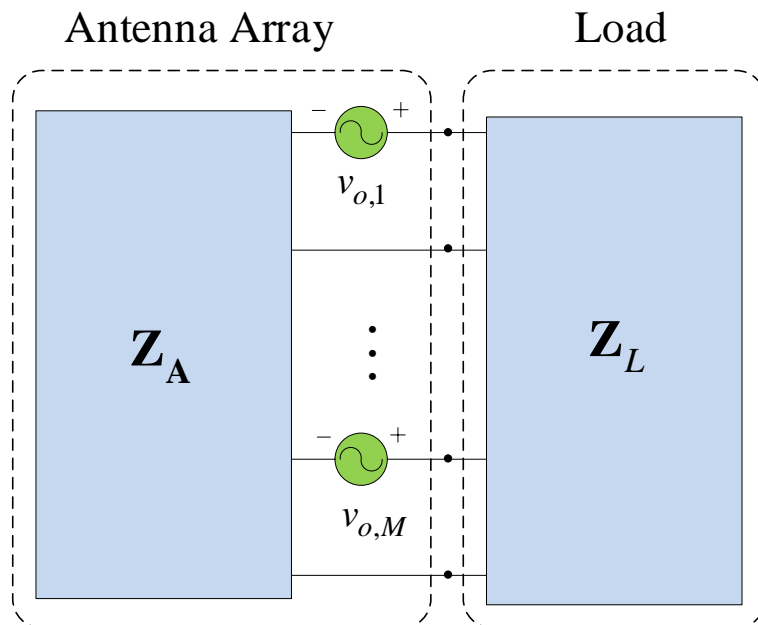


Figure 5.1: Circuit model of a multiple-antenna receiver

using likelihood ratio test in Sec. 5.3. The performance of this detector is then explored via numerical examples in Sec. 5.4. We summarize our conclusions and point out future directions in Sec. 5.5.

## 5.2 System Model

Consider a narrowband multiple-input multiple-output (MIMO) communications link with  $M$  receive antennas and  $N$  transmit antennas. The receiver model is illustrated in Fig. 5.1. This circuit model is identical to the ones widely used to model a scenario, where amplifier noise dominates [29, 51, 79]. More complex models are considered in [26, 27].

In Fig. 5.1, we model the antenna array by its Thevenin equivalent,

$$\mathbf{v} = \mathbf{Z}_A \mathbf{i} + \mathbf{v}_o, \quad (5.1)$$

where  $\mathbf{v}, \mathbf{i} \in \mathbb{C}^M$  are the voltage across, and current into, the antenna array terminals. In particular, the antenna impedance is a symmetric matrix in  $\mathbb{C}^{M \times M}$ ,

$$\mathbf{Z}_A = \mathbf{R}_A + j\mathbf{X}_A, \quad (5.2)$$

and  $\mathbf{R}_A$  and  $\mathbf{X}_A$  are the resistance and reactance, respectively. The incident electromagnetic field induces open-circuit voltage  $\mathbf{v}_o \in \mathbb{C}^M$  in (5.1). Under flat-fading conditions, the open-circuit voltage  $\mathbf{v}_o$  is modeled as [27]

$$\mathbf{v}_o = \mathbf{G} \mathbf{x}, \quad (5.3)$$

where  $\mathbf{x} \in \mathbb{C}^N$  is the transmitted symbol and  $\mathbf{G} \in \mathbb{C}^{M \times N}$  is the fading path gain. With sufficiently separated transmit antennas, the columns of  $\mathbf{G}$  is i.i.d. zero-mean, complex Gaussian, i.e.,  $\mathbf{g}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{g}})$ . As shown in Fig. 5.1, noisy voltage observation  $\mathbf{v}_L$  across load impedance  $\mathbf{Z}_L \in \mathbb{C}^{M \times M}$  is assumed [29, 51, 79],

$$\mathbf{v}_L = \mathbf{R}_L^{1/2} (\mathbf{Z}_A + \mathbf{Z}_L)^{-1} \mathbf{G} \mathbf{x} + \mathbf{n}_L, \quad (5.4)$$

where the observation noise  $\mathbf{n}_L \in \mathbb{C}^M$  is a zero-mean, circularly-symmetric, complex Gaussian random vector with variance  $E[\mathbf{n}_L \mathbf{n}_L^H] = \sigma_{n,s}^2 \mathbf{I}_M$ , which is hereafter denoted by  $\mathbf{n}_L \sim \mathcal{CN}(\mathbf{0}, \sigma_{n,s}^2 \mathbf{I}_M)$ . The subscript  $s$  represents the noise for each received symbol.

Suppose the transmitter is a base station, with  $N$  uncoupled transmit antennas. The

transmitter sends a known training sequence of length  $T$ , i.e.,  $\mathbf{x}_1, \dots, \mathbf{x}_T \in \mathbb{C}^N$  to the receiver. Also assume block fading, such that the fading path gain  $\mathbf{G}$  remains fixed during each packet, of which the training sequence is a part [16]. An effective channel matrix that communication algorithms need is defined as,

$$\mathbf{H} \triangleq \mathbf{R}_L^{1/2} (\mathbf{Z}_A + \mathbf{Z}_L)^{-1} \mathbf{G} \in \mathbb{C}^{M \times N} . \quad (5.5)$$

The columns of  $\mathbf{H}$  are also i.i.d. zero-mean, complex Gaussian, i.e.,  $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_h)$ , where

$$\boldsymbol{\Sigma}_h = \mathbf{R}_L^{1/2} (\mathbf{Z}_A + \mathbf{Z}_L)^{-1} \boldsymbol{\Sigma}_g (\mathbf{Z}_A + \mathbf{Z}_L)^{-H} \mathbf{R}_L^{1/2} . \quad (5.6)$$

Consider observations of  $T$  training symbols at the load. We write (5.4) in matrix form, after defining  $\mathbf{X} \triangleq [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T] \in \mathbb{C}^{N \times T}$ ,

$$\mathbf{V}_L = \mathbf{H}\mathbf{X} + \mathbf{N}_L . \quad (5.7)$$

It is reasonable to assume that the base station (transmitter) has more antennas than the mobile device (receiver), i.e.,  $N \geq M$ . Conditioned on  $T \geq N$ , and  $\mathbf{X}\mathbf{X}^H \in \mathbb{C}^{N \times N}$  is full-rank, the following statistic is sufficient for estimating  $\mathbf{H}$ , which has been shown in Lemma 4 in Chapter 4,

$$\mathbf{Y} = \mathbf{V}_L \mathbf{X}^H (\mathbf{X}\mathbf{X}^H)^{-1} . \quad (5.8)$$

The training sequences for MIMO channel estimation are often equal-energy and orthogonal, i.e.,  $\mathbf{X}\mathbf{X}^H = (PT/N)\mathbf{I}_N$ . This can be achieved by using a normalized discrete Fourier transform (DFT) matrix, e.g., [16, eq. 10]. Then,  $\mathbf{Y}$  becomes,

$$\mathbf{Y} = \mathbf{H} + \mathbf{N} , \quad (5.9)$$

where  $\mathbf{n} \triangleq \text{vec } \mathbf{N} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{MN})$ ,  $P/N$  is the total transmitted power for each training symbol, and the noise variance is,

$$\sigma_n^2 \triangleq \frac{N\sigma_{n,s}^2}{PT}. \quad (5.10)$$

### 5.3 Antenna Impedance Variation Detection

In this section, we utilize well known detection techniques to sense variations of antenna impedance over time. The goal is to detect changes in antenna impedance due to time-varying user interference in the near-field and then apply our previously developed antenna impedance estimation algorithms to compensate these changes.

We consider two consecutive blocks of packets, with each block containing  $L$  packets. Thus,  $2L$  consecutive packets are utilized for this detection task. We assume the antenna impedance  $\mathbf{Z}_A$  remains fixed during each block of packets, but abrupt changes may occur from block to block. This is a tractable approximation of the possibly continuously time-varying antenna impedance.

In the remainder of this section, we use resources readily available at the receiver, e.g., training sequences allocated to channel estimation, and formulate this detection problem using hypothesis testing. Specifically, we apply likelihood ratio tests for equality of two covariance matrices of training sequences to detect variations in antenna impedance matrix  $\mathbf{Z}_A$ .

Firstly, we consider a scenario where the receiver does not know *a priori* the signal covariance matrices. This is the worse-case scenario for this detection task. We later comment on appropriate techniques to use when one of these two covariance matrices is known.

Collect the sufficient statistics (5.9) of received training sequences over the  $L$  packets for both blocks,

$$\mathbf{Y}_1 \triangleq \begin{bmatrix} \mathbf{Y}_1 & \cdots & \mathbf{Y}_L \end{bmatrix}, \quad \mathbf{Y}_2 \triangleq \begin{bmatrix} \mathbf{Y}_{L+1} & \cdots & \mathbf{Y}_{2L} \end{bmatrix}, \quad (5.11)$$

where, for  $1 \leq l \leq 2L$ ,

$$\mathbf{Y}_l = \mathbf{H}_l + \mathbf{N}_l \in \mathbb{C}^{M \times N}. \quad (5.12)$$

The covariance matrix of  $\mathbf{Y}_i$  is defined as, where  $i = 1, 2$ ,

$$\boldsymbol{\Sigma}_i \triangleq \frac{1}{NL} E [\mathbf{Y}_i \mathbf{Y}_i^H] = \boldsymbol{\Sigma}_{\mathbf{h},i} + \sigma_n^2 \mathbf{I}_M. \quad (5.13)$$

Note the signal covariance matrix is a function of the antenna impedance matrix, i.e.,

$$\boldsymbol{\Sigma}_{\mathbf{h},i} = \mathbf{R}_L^{1/2} (\mathbf{Z}_{\mathbf{A},i} + \mathbf{Z}_L)^{-1} \boldsymbol{\Sigma}_{\mathbf{g}}(\mathbf{Z}_{\mathbf{A},i}) (\mathbf{Z}_{\mathbf{A},i} + \mathbf{Z}_L)^{-H} \mathbf{R}_L^{1/2}, \quad (5.14)$$

where the subscript  $i = 1, 2$  of  $\mathbf{Z}_{\mathbf{A},i}$  represents the possibly different antenna impedance matrices in these two blocks. The load impedance  $\mathbf{Z}_L$  is assumed fixed. The covariance of fading path gains  $\boldsymbol{\Sigma}_{\mathbf{g}}$  in general may depend on  $\mathbf{Z}_{\mathbf{A},i}$ , e.g., under black block radiation [48, eq. 3.8].

Thus, we formulate the detection of antenna impedance variations by the following hypothesis testing,

$$\begin{aligned} H_0 &: \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2, \\ H_1 &: \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2. \end{aligned} \quad (5.15)$$

If the null hypothesis  $H_0$  is true, then the antenna impedance matrix remains identical in the two blocks considered, i.e.,  $\mathbf{Z}_{\mathbf{A},1} = \mathbf{Z}_{\mathbf{A},2}$ . Otherwise, if the alternative hypothesis  $H_1$  is true, a change in antenna impedance occurs, i.e.,  $\mathbf{Z}_{\mathbf{A},1} \neq \mathbf{Z}_{\mathbf{A},2}$ , which needs to be estimated for mismatch compensation.

Bartlett's modified likelihood ratio test<sup>1</sup> has been widely used for the hypothesis test given in (5.15), e.g., [44, Sec. 2.4] [63],

$$\lambda_L \triangleq \frac{\det(\mathbf{S}_1)^L \det(\mathbf{S}_2)^L}{\det(\mathbf{S}_1 + \mathbf{S}_2)^{2L}} \cdot 4^{M \times L}, \quad (5.16)$$

where  $\det(\cdot)$  is the determinant operator, and the sample covariance matrices are defined as, for  $i = 1, 2$ ,

$$\mathbf{S}_i = \frac{1}{NL} \mathbf{Y}_i \mathbf{Y}_i^H. \quad (5.17)$$

If the test statistic  $\lambda_L$  in (5.16) is less than or equal to a predetermined threshold, say  $\theta$ , then the null hypothesis  $H_0$  is rejected, i.e., the detector thinks a change has occurred. Conditioned on  $H_1$  is true, i.e., there is a change in the covariance matrix, this change has been correctly detected. But if  $H_0$  is true, i.e., no change has occurred, we encounter an event of false alarm. For fixed  $M$  and  $L \rightarrow \infty$  and under the null hypothesis  $H_0$  in (5.15), a scaled natural logarithm of the test statistic  $\lambda_L$  converges to a chi-square distribution [44, eq. 2.12]. This result could be used to select  $\theta(\alpha)$  based on a desired false alarm probability  $\alpha \in (0, 1)$ . Furthermore, if  $\Sigma_1$  is assumed known, we could use another test given by Bartlett [13, Sec. IIIa]. This can be justified by that antenna impedance has been accurately estimated and the signal covariance matrix is known.

Note the test in (5.16) is derived for i.i.d. observations from complex Gaussian distri-

---

<sup>1</sup>Note the discrepancies between the test statistic in (5.16) and [44, eq. 2.11] are due to our assumptions of complex Gaussian with known zero-mean, yet they consider real Gaussian, whose mean is unknown and estimated.

butions,  $\mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_i)$  for  $i = 1, 2$ . We would like to learn, under correlated Rayleigh fading, how its performance deteriorates if temporal correlation is ignored. In the next section, we investigate these aspects of this test statistic given in (5.16) via numerical examples.

## 5.4 Numerical Results

In this section, we explore the performance of the test statistics given in the previous section through numerical examples. Consider a narrow-band MIMO communications system with  $N = 4$  transmit antennas and  $M = 2$  receive antennas, whose carrier frequency is 2.1 GHz. This frequency is chosen based on the first E-UTRA down-link operating band in LTE specifications [82]. The duration of each data packet equals to a subframe of LTE, i.e.,  $T_s = 1$  ms. Block-fading channel is assumed, such that during one data packet, the channel information remains the same, but it generally varies from packet to packet [16].

For each data packet, a training sequence precedes data sequence [54, Fig. 1(a)]. We take the training sequence  $\mathbf{X}$  in (5.7) from a normalized discrete Fourier transform (DFT) matrix of dimension  $T = 64$ , e.g., [16, eq. 10]. Specifically,  $\mathbf{X}$  is chosen as the first  $N$  rows of this DFT matrix, and  $\mathbf{X}\mathbf{X}^H = \frac{PT}{N}\mathbf{I}_N$ . Before a change occurs, the antenna impedance is that of a uniform linear array (ULA) with half-wavelength inter-element spacing [27], i.e.,

$$\mathbf{Z}_A = \begin{bmatrix} 72.8521 + j1.6869 & -15.7457 - j27.8393 \\ -15.7457 - j27.8393 & 72.8521 + j1.6869 \end{bmatrix} \Omega . \quad (5.18)$$

In other words, we assume  $\mathbf{\Sigma}_1$  in (5.13) is that calculated using (5.18) (and its correlation matrix) with the noise variance  $\sigma_n^2$  defined in (5.10). If a change occurs, the antenna

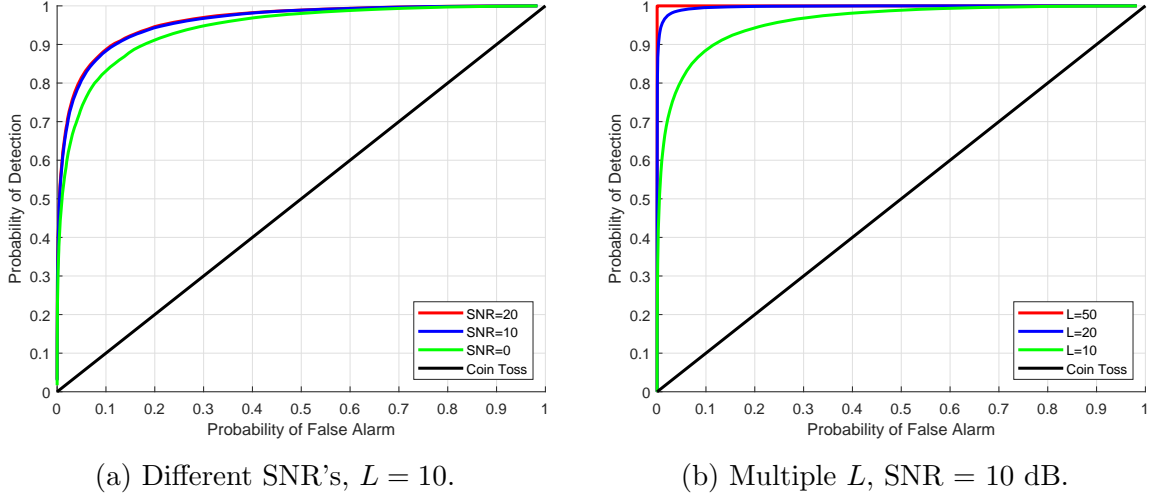


Figure 5.2: RoC Curves for 4 by 2 MIMO under i.i.d. Fading.

impedance becomes that of ULA with 0.1 wavelength separation, i.e.,

$$\mathbf{Z}_{\mathbf{A},cgd} = \begin{bmatrix} 73.0945 - j2.2728 & -67.5019 - j1.6580 \\ -67.5019 - j1.6580 & 73.0945 - j2.2728 \end{bmatrix} \Omega . \quad (5.19)$$

Its covariance matrix  $\mathbf{\Sigma}_2$  in (5.13) is then calculated accordingly. The load impedance is fixed at  $\mathbf{Z}_L = 50\mathbf{I}_M \Omega$ , so we do not assume knowing  $\mathbf{Z}_{\mathbf{A}}$ .

The average post-detection SNR of a received symbol is defined from (5.4) as [16, Sec. VIII],

$$\rho \triangleq \frac{E \text{Tr} [\mathbf{H}\mathbf{x}\mathbf{x}^H \mathbf{H}^H]}{E \text{Tr} [\mathbf{n}_L \mathbf{n}_L^H]} = \frac{P\sigma_H^2}{\sigma_{n,s}^2} , \quad (5.20)$$

where  $\sigma_{n,s}^2$  is the noise variance at each port of the  $M$ -port receiver and  $\sigma_H^2$  is the mean of diagonal entries of  $\mathbf{\Sigma}_{\mathbf{h}}$  in (5.6) calculated with  $\mathbf{Z}_{\mathbf{A}}$  (5.18),

$$\sigma_H^2 \triangleq \frac{1}{M} \text{Tr}[\mathbf{\Sigma}_{\mathbf{h}}] . \quad (5.21)$$

Receiver operating characteristic (RoC) curves are often plotted for hypothesis testing. These curves show the trade-off between probability of detection, say  $P_d$ , and probability of false alarm, say  $P_{fa}$ , of a given detector. We first explore the impacts of SNR and number of packets  $L$  to the detector in (5.16) in Figs. 5.2a and 5.2b, respectively. With increased SNR, the detection probability  $P_d$  improves for a given false alarm rate  $P_{fa}$ . Also, diminishing return is observed as SNR increases beyond 10 dB in Fig. 5.2a. So with reasonable SNR, this detector is limited in performance by diversity. In Fig. 5.2b, we address the question how many packets are needed, say at SNR = 10 dB, to approach the perfect classification at the upper left vertex, where  $P_d = 1$  and  $P_{fa} = 0$ . One observes the trend is more diversity (larger  $L$ ) leads to narrower gap between the RoC and the perfect classification. With  $L = 50$ , it seems the detector in (5.16) is in the small vicinity of the upper left vertex. Next we consider more practical scenarios, where consecutive packets are generally correlated.

Clarke's model is assumed for temporal correlation as in previous chapters [9, 21, 86], and the normalized channel correlation matrix is,

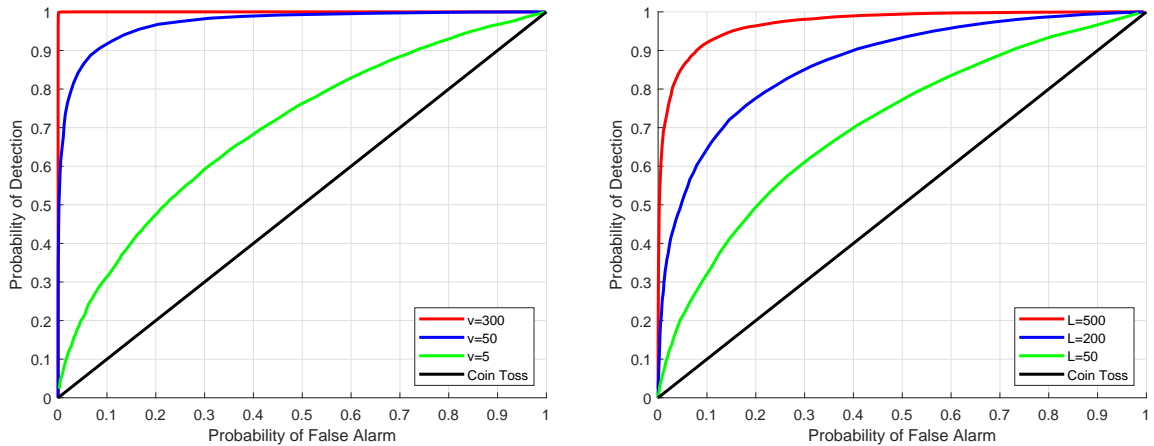
$$\mathbf{C}_h = \begin{bmatrix} R[0] & R[-1] & \cdots & R[-L+1] \\ R[1] & R[0] & \cdots & R[-L+2] \\ \vdots & \ddots & \ddots & \vdots \\ R[L-1] & R[L-2] & \cdots & R[0] \end{bmatrix}, \quad (5.22)$$

where  $R[l] = J_0(2\pi f_d T_s |l|)$ ,  $J_0(\cdot)$  is the zeroth-order Bessel function of the first kind,  $T_s = 1 \text{ ms}$  is the sampling interval, and  $l$  is the sample difference. The fading frequency (maximum Doppler frequency) is  $f_d \triangleq v/\lambda$ , where  $v$  is the velocity of the fast moving scatterer and  $\lambda$  the wave-length of the carrier frequency.

Fig. 5.3a demonstrates the impact of fading conditions on this detection task. Faster fading results in better trade-off between  $P_d$  and  $P_{fa}$ . In particular, the RoC curve for fast fading (red line) is similar to that of the i.i.d. fading (also red line) shown in Fig. 5.2b, in that they reach the upper left corner. For the medium fading, i.e.,  $v = 50$  km/h, reasonable detection versus false alarm trade-off can be achieved, e.g.,  $P_d > 0.9$  and  $P_{fa} < 0.1$ . However, this trade-off is less favorable when the fading is very slow, e.g.,  $v = 5$  km/h. Under this slow fading condition, the RoC curve (green line) is closer to the chance line, flipping a fair coin, than the other fading conditioned plotted. For example, if  $P_{fa} = 0.1$  is required, the detection probability is only slightly above 30%. On the other hand, a high detection rate, e.g.,  $P_d = 0.9$ , entails a high false alarm probability over 70%. In order to improve this improves, in Fig. 5.3b, we combine more packets than  $L = 50$ . We observe that added diversity indeed produces a better detection and false alarm trade-off, and  $P_d > 0.9$  and  $P_{fa} < 0.1$  is achieved with  $L = 500$  packets. With our system settings given at the beginning of this section, this means, in the worse-case, the detector in (5.16) results in practically useful detection and false alarm trade-off using observations within 1 second. If the fading is identical or faster than the medium fading ( $v = 50$  km/h), then this same trade-off can be obtained using observations within 0.1 seconds.

## 5.5 Conclusion

In this chapter, we formulated the problem of sensing variations in antenna impedance over time using hypothesis testing. In particular, we apply well known likelihood ratio tests for equality of covariance matrices. For example, Bartlett's modified likelihood ratio test is demonstrated to achieve a practically useful trade-off between probability of



(a) Different Fading, SNR = 10 dB,  $L = 50$ . (b) Slow Fading, Multiple  $L$ , SNR = 10 dB.

Figure 5.3: RoC Curves for 4 by 2 MIMO under Correlated Rayleigh Fading.

detection and probability of false alarm in general. Specifically, if the Rayleigh fading is not extremely slow, accurate detection with low false alarm rate can be obtained within a fraction of a second, assuming a set of system parameters consistent to LTE. If the fading is slow, then combination of more packets is needed for similar detection vs false alarm trade-off, which leads to a longer delay.

Future directions may include more accurate covariance estimation techniques when observations are highly correlated. Furthermore, other types of tracking algorithms may be developed. For example, we could assume a dynamic model for the time-varying antenna impedance and then apply appropriate filtering techniques, e.g., the extended Kalman filter.

# Chapter 6

## Conclusion

In this dissertation, we consider the problem of antenna impedance estimation at MIMO receivers. We first develop a hybrid estimation framework for joint estimation of channel information and antenna impedance at single-input, single-output (SISO) receiver in Rayleigh fading channels. Based on observation of training sequences via synchronously switched load at the receiver, we derive the joint maximum a posteriori and maximum-likelihood (MAP/ML) estimators for channel and impedance over multiple packets. However, this joint ML estimator is found inconsistent. We modify it to make it consistent, which improves channel estimation also.

Next, we consider a multiple-input, single-out (MISO) system, where multiple antennas are deployed at the transmitter. Here we decouple impedance estimation from channel estimation and formulate this problem in a classical estimation setting. Then, we derive the maximum-likelihood (ML) estimator for the antenna impedance over multiple packets. We also investigate the benefits on impedance estimation due to the added spatial diversity at the transmitter.

Further, we generalize this to the multiple-input, multiple-out (MIMO) problem,

where multiple antennas exist at both the transmitter and receiver. We derive the ML estimator for the antenna impedance matrix under i.i.d. fading channels. Additionally, a simple method of moments (MM) estimator is derived under correlated fading conditions. We compare the performance of these estimators to a fundamental lower bound in Cramér-Rao bound (CRB). More importantly, we find this antenna impedance estimation algorithm is fast and is able to compensate mismatch in milliseconds, via numerical examples. The ergodic channel capacity is studied, which demonstrates a sizable performance boost for a receiver using our impedance estimation algorithm, compared to the originally mismatched receiver.

## REFERENCES

- [1] 3GPP, “Technical specification group radio access network; (E-UTRA) and (E-UTRAN); physical channels and modulation; (release 8),” Tech. Rep., 2008.
- [2] 3GPP, “LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception; TS 36.101 version 14.5.0 Release 14,” Tech. Rep., 2017.
- [3] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed. Washington DC: National Bureau of Standards, Applied Mathematics Series No. 55, Dec. 1972.
- [4] T. Adali, P. J. Schreier and L. L. Scharf, “Complex-Valued Signal Processing: The Proper Way to Deal With Impropropriety,” *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5101–5125 Nov. 2011.
- [5] H. Arslan and G. E. Bottomley, “Channel estimation in narrowband wireless communication systems,” *Wirel. Commun. Mob. Comput.*, vol. 1, no. 2, pp. 201–219, 2001.
- [6] D. Astély and B. Ottersten, “The effects of local scattering on direction of arrival estimation with MUSIC,” *IEEE Trans. Signal Process.*, vol. 47, no. 12, pp. 3220–3234, 1999.
- [7] S. M. Ali, M. Buckley, J. Deforge, J. Warden and A. Danak, “Dynamic Measurement of Complex Impedance in Real-Time for Smart Handset Applications,” *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 9, pp. 3453–3460, Aug. 2013.
- [8] S. M. Ali, A. Mobasher and P. Lusina, “User Effects on MIMO Performance: From an Antenna to a Link Perspective,” *Int. J. Antennas Propag.*, vol. 2011, Article ID 918315, pp. 1–13, 2011. doi:10.1155/2011/918315
- [9] K. E. Baddour and N. C. Beaulieu, “Autoregressive modeling for fading channel simulation,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1650–1662, Jul. 2005.
- [10] M. H. Bakr, S. M. Ali, J. Warden and A. Sheynman, “Dynamic real-time calibration for antenna matching in the transmitting and receiving modes,” *Int. J. RF Microw. Comput. Eng.*, vol. 22, no. 1, pp. 59–67, Jan. 2012.
- [11] C. A. Balanis, *Antenna Theory: Analysis and Design*. 3rd edition. Wiley, 2005.

- [12] S. Bar and J. Tabrikian, “Bayesian Estimation in the Presence of Deterministic Nuisance Parameters—Part I: Performance Bounds,” *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6632–6646, Dec. 2015.
- [13] M. S. Bartlett, “A note on the multiplying factor for various  $\chi^2$  approximations in factor analysis”, *J. R. Stat. Soc.*, vol. 16, no. 2, pp. 296–298, 1954.
- [14] R. J. Baxley, B. T. Walkenhorst and G. Acosta-Marum, “Complex Gaussian ratio distribution with applications for error rate calculation in fading channels with imperfect CSI,” *IEEE Glob. Telecommun. Conf.*, 2010, pp. 1–5.
- [15] A. van Bezooijen, M. A. de Jongh, F. van Straten, R. Mahmoudi and A. H. M. van Roermund, “Adaptive Impedance-Matching Techniques for Controlling L Networks,” *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 2, pp. 495–505, Jun. 2010.
- [16] M. Biguesh and A. B. Gershman, “Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals,” *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.
- [17] G. E. P. Box, “A General Distribution Theory for a Class of Likelihood Criteria,” *Biometrika*, vol. 36, no. 3, pp. 317–346, 1949.
- [18] D. H. Brandwood, “A complex gradient operator and its application in adaptive array theory,” *IEE Proc. H Microwaves, Opt. Antennas*, vol. 130, no. 1, pp. 11–16 Dec. 1983.
- [19] J. W. Brewer, “Kronecker Products and Matrix Calculus in System Theory,” *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 772–781, 1978.
- [20] D. C. Chu, “Polyphase Codes with Good Periodic Correlation Properties,” *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 531–532, Jul. 1972.
- [21] R. H. Clarke, “A Statistical Theory of Mobile-Radio Reception,” *Bell Syst. Tech. J.*, vol. 47, no. 6, pp. 957–1000, Jul. 1968.
- [22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2nd, Wiley, 2006.
- [23] A. Das and B. D. Rao, “SNR and Noise Variance Estimation for MIMO Systems,” *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3929–3941, Aug. 2012.
- [24] X. Dai, W. Zhang, J. Xu, J. E. Mitchell, and Y. Yang, “Kalman interpolation filter for channel estimation of LTE downlink in high-mobility environments,” *EURASIP J. Wirel. Commun. Netw.*, vol. 2012, no. 1, p. 232, Dec. 2012.

- [25] C. Desoer, "The maximum power transfer theorem for n-ports," *IEEE Trans. Circuit Theory*, vol. 20, no. 3, pp. 328–330, 1973.
- [26] C. P. Domizioli and B. L. Hughes, "Noise correlation in compact diversity receivers," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1426–1436, May 2010.
- [27] C. P. Domizioli and B. L. Hughes, "Front-end design for compact MIMO receivers: A communication theory perspective," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2938–2949, Oct. 2012.
- [28] M. J. Gans, "Channel capacity between antenna Arrays - Part I: sky noise dominates," *IEEE Trans. Commun.*, vol. 54, no. 9, pp. 1586–1592, Sep. 2006.
- [29] M. J. Gans, "Channel capacity between antenna arrays - Part II: Amplifier noise dominates," *IEEE Trans. Commun.*, vol. 54, no. 11, pp. 1983–1992, Nov. 2006.
- [30] A. Hjørungnes and D. Gesbert, "Complex-Valued Matrix Differentiation: Techniques and Key Results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.
- [31] J. Gillard, "Asymptotic variance-covariance matrices for the linear structural model," *Stat. Methodol.*, vol. 8, no. 3, pp. 291–303, May 2011.
- [32] J. Gjengset, J. Xiong, G. McPhillips and K. Jamieson, "Phaser: Enabling Phased Array Signal Processing on Commodity WiFi Access Points," *20th annual int. conf. Mobile computing and networking, Mobicom*, 2014, vol. 19, no. 2, pp. 153-164.
- [33] Q. Gu, J. R. De Luis, A. S. Morris and J. Hilbert, "An Analytical Algorithm for Pi-Network Impedance Tuners," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 12, pp. 2894–2905, Jul. 2011.
- [34] Q. Gu and A. S. Morris, "A New Method for Matching Network Adaptive Control," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 1, pp. 587–595, Dec. 2013.
- [35] H. Haber, "Diagonalization of a general  $2 \times 2$  hermitian matrix", *Physics 216 Lecture Note*, University of California, Santa Cruz, pp. 1–4, 2012.
- [36] Y. Hassan and A. Wittneben, "Joint Spatial Channel and Coupling Impedance Matrices Estimation in Compact MIMO Systems : The Use of Adaptive Loads," *IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 29–33, 2015.
- [37] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, 2003.

- [38] R. He *et al.*, “High-Speed Railway Communications: From GSM-R to LTE-R,” *IEEE Veh. Technol. Mag.*, vol. 11, no. 3, pp. 49–58, Sep. 2016.
- [39] P. Hoeher, S. Kaiser, and P. Robertson, “Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering,” *IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, no. 3, pp. 1845–1848, 1997.
- [40] K. Hood, B. A. J. Nix, and T. C. Iles. “Asymptotic Information and Variance-Covariance Matrices for the Linear Structural Model”, *Journal of the Royal Statistical Society. Series D (The Statistician)*, vol. 48, no. 4, pp. 477–493, 1999.
- [41] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Second ed. Cambridge, UK: Cambridge University Press, 2012.
- [42] M. M. Hyder, R. H. Khan and K. Mahata, “An enhanced random access mechanism for Smart Grid M2M communications in WiMAX networks,” *IEEE Int. Conf. Smart Grid Commun.*, 2014, pp. 1–6.
- [43] M. T. Iurlac and J. A. Nossek,, “Toward a Circuit Theory of Communication,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 57, no. 7, pp. 1663–1683, Jul. 2010.
- [44] T. Jiang and Y. Qi, “Likelihood Ratio Tests for High-Dimensional Normal Distributions,” *Scand. J. Stat.*, vol. 42, no. 4, pp. 988–1009, Dec. 2015.
- [45] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, New Jersey: Prentice Hall, 1993.
- [46] S. Kay and Y. C. Eldar, “Rethinking biased estimation,” *IEEE Signal Process. Mag.*, vol. 25, no. 3, pp. 133–136, 2008.
- [47] K. Kreutz-Delgado, “The Complex Gradient Operator and the  $\mathbb{C}\mathbb{R}$ -Calculus,” arXiv Preprint, pp. 1–74, arXiv:0906.4835, Jun. 2009.
- [48] L. Kundu, “Information-Theoretic Limits on MIMO Antennas,” Ph.D. Dissertation, North Carolina State University, 2016.
- [49] C. D. Lai, G. R. Wood, and C. G. Qiao, “The mean of the inverse of a punctured normal distribution and its application,” *Biometrical J.*, vol. 46, no. 4, pp. 420–429, 2004.
- [50] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

- [51] B. K. Lau, J. B. Andersen, G. Kristensson and A. F. Molisch, “Impact of Matching Network on Bandwidth of Compact Antenna Arrays,” *IEEE Trans. Antennas Propag.*, vol. 54, no. 11, pp. 3225–3238, Nov. 2006.
- [52] W. C. Y. Lee, “Estimate of channel capacity in Rayleigh fading environment,” *IEEE Trans. Vehicular Techn.*, vol. 39, no. 3, pp. 187–189, Aug. 1990.
- [53] W. Li, “Fundamental Limits of Estimation Using Arbitrary Compact Arrays,” Ph.D. Dissertation, North Carolina State University, 2018.
- [54] Y. Liu, Z. Tan, H. Hu, L. J. Cimini, and G. Y. Li, “Channel Estimation for OFDM,” *IEEE Commun. Surv. Tutorials*, vol. 16, no. 4, pp. 1891–1908, 2014.
- [55] D. Maiwald and D. Kraus, “Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices,” *IEE Proc. - Radar, Sonar Navig.*, vol. 147, no. 4, p. 162–168, 2000.
- [56] R. Miller and C. Chang, “A modified Cramér-Rao bound and its applications,” *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 398–400, May 1978.
- [57] R. Mohammadkhani and J. S. Thompson, “Adaptive Uncoupled Termination for Coupled Arrays in MIMO Systems,” *IEEE Trans. Antennas Propag.*, vol. 61, no. 8, pp. 4284–4295, May 2013.
- [58] S. Nadarajah and H. S. Kwong, “A note on On the ratio of independent complex Gaussian random variables’,” *Multidim. Syst. Sign. Process.*, pp. 1-5, Nov. 2017, <https://doi.org/10.1007/s11045-017-0537-1>
- [59] E. S. Nadimi, M. H. Ramezani and V. Blanes-Vidal, “On the ratio of independent complex Gaussian random variables,” *Multidim. Syst. Sign. Process.*, pp. 1-9, Aug. 2017, <https://doi.org/10.1007/s11045-017-0519-3>.
- [60] Y. Noam and H. Messer, “Notes on the Tightness of the Hybrid Cramér-Rao Lower Bound,” *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2074–2084, Jun. 2009.
- [61] N. ODonoughue and J. M. F. Moura, “On the product of independent complex Gaussians,” *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1050–1063, Nov. 2012.
- [62] E. Pereira and J. Vitória, “Deflation for block eigenvalues of block partitioned matrices with an application to matrix polynomials of commuting matrices,” *Comput. Math. with Appl.*, vol. 42, no. 8-9, pp. 1177-1188, 2001.
- [63] M. D. Perlman, “Unbiasedness of the Likelihood Ratio Tests for Equality of Several Covariance Matrices and Equality of Several Multivariate Normal Populations,” *Ann. Stat.*, vol. 8, no. 2, pp. 247–263, Mar. 1980.

- [64] D. M. Pozar, *Microwave Engineering*. 4th ed. New York, NY: Wiley, 2012.
- [65] D. Qiao *et al.*, “Antenna Impedance Mismatch Measurement and Correction for Adaptive CDMA Transceivers,” *IEEE MTT-S Int. Microw. Symp. Dig.*, 2005, pp. 783–786.
- [66] J. Rahola, “Power waves and conjugate matching,” *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 55, no. 1, pp. 92–96, Jan. 2008.
- [67] C. Ren, J. Galy, E. Chaumette, P. Larzabal, and A. Renaux, “A Ziv-Zakai; type bound for hybrid parameter estimation,” *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, no. 1, pp. 4663–4667.
- [68] I. Reuven and H. Messer, “A Barankin-type lower bound on the estimation error of a hybrid parameter vector,” *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 1084–1093, May 1997.
- [69] Y. Rockah and P. Schultheiss, “Array shape calibration using sources in unknown locations—Part I: Far-field sources,” *IEEE Trans. Acoust.*, vol. 35, no. 3, pp. 286–299, Mar. 1987.
- [70] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data*. Cambridge University Press, New York, 2010.
- [71] H. Shin and J. H. Lee, “Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole,” *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2636–2647, Oct. 2003.
- [72] Y. Sun, J. Moritz and X. Zhu, “Adaptive impedance matching and antenna tuning for green software-defined and cognitive radio,” *IEEE 54th Int. Midwest Symp. Circuits and Syst.*, 2011, pp. 1–4.
- [73] E. Telatar, “Capacity of Multi-antenna Gaussian Channels,” *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.
- [74] I. Vasilev, J. Lindstrand, V. Plicanic, H. Sjoland and B. K. Lau, “Experimental Investigation of Adaptive Impedance Matching for a MIMO Terminal With CMOS-SOI Tuners,” *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 5, pp. 1622–1633, Apr. 2016.
- [75] I. Vasilev, V. Plicanic and B. K. Lau, “Impact of Antenna Design on MIMO Performance for Compact Terminals With Adaptive Impedance Matching,” *IEEE Trans. Antennas Propag.*, vol. 64, no. 4, pp. 1454–1465, Jan. 2016.

- [76] A. van den Bos, "Complex gradient and Hessian," *IEE Proc. Vision, Image, Signal Process.*, vol. 141, no. 6, pp. 380–382 Dec. 1994.
- [77] H. van Trees, K. Bell and Z. Zhi, *Detection, Estimation and Modulation Theory Part I - Detection, Estimation, and Filtering Theory*. New York: Wiley, 2013.
- [78] M. Vu, "MISO Capacity with Per-Antenna Power Constraint," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1268–1274, May 2011.
- [79] J. W. Wallace and M. A. Jensen, "Mutual Coupling in MIMO Wireless Systems: A Rigorous Network Theory Analysis," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp 1317–1325, Jul. 2004
- [80] S. Wu and B. L. Hughes, "Training-based joint channel and impedance estimation," *IEEE 2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, Princeton University, NJ, Mar. 2018.
- [81] S. Wu and B. L. Hughes, "A Hybrid Approach to Joint Estimation of Channel and Antenna Impedance," *IEEE 2018 52nd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 2018.
- [82] S. Wu and B. L. Hughes, "PCA-Based Antenna Impedance Estimation in Rayleigh Fading Channels," in manuscript, version 31, Nov. 30, 2018.
- [83] G. Yan and H. Fan, "A Newton-like algorithm for complex variables with applications in blind equalization," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 533–556, Feb. 2000.
- [84] C. Yang and H. Shao, "WiFi-based indoor positioning," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 150–157, Mar. 2015.
- [85] M. B. Yelten and K. G. Gard, "Theoretical analysis and characterization of the tunable matching networks in low noise amplifiers," *Eur. Conf. Circuit Theory and Des.*, 2009, pp. 890–893.
- [86] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920–928, Jun. 2003.
- [87] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental trade-off in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

## APPENDICES

# Appendix A

## Proof of Theorem 3 in Chapter 2

We now prove (2.39) in Theorem 3 of Chapter 2. Note the conditional mean of  $X_1$  given  $X_2$  is given by [45, Sec. 15.8]

$$E[X_1|X_2] = \mu_1 + \frac{\rho\sigma_1}{\sigma_2}(X_2 - \mu_2) .$$

It follows that the mean of the ratio can be written as

$$\begin{aligned} E\left[\frac{X_1}{X_2}\right] &= E\left[\frac{E[X_1|X_2]}{X_2}\right] = \frac{\rho\sigma_1}{\sigma_2} + \left(\mu_1 - \frac{\rho\sigma_1}{\sigma_2}\mu_2\right) E\left[\frac{1}{X_2}\right] \\ &= \frac{\mu_1}{\mu_2} + \left(\frac{\rho\sigma_1}{\sigma_2} - \frac{\mu_1}{\mu_2}\right) \left(1 - E\left[\frac{\mu_2}{X_2}\right]\right) . \end{aligned} \quad (\text{A.1})$$

It only remains to evaluate  $E[\mu_2/X_2]$ . The pdf of  $X_2$  is given in polar coordinates by

$$\begin{aligned} f(r, \theta) &= \frac{r}{\pi\sigma_2^2} \exp\left(-\frac{|re^{j\theta} - ve^{j\phi}|^2}{\sigma_2^2}\right) \\ &= \frac{r}{\pi\sigma_2^2} \exp\left(-\frac{r^2 + v^2 - 2vr \cos(\theta - \phi)}{\sigma_2^2}\right) , \end{aligned} \quad (\text{A.2})$$

where  $\mu_2 = ve^{j\phi}$ . It follows

$$\begin{aligned}
E \left[ \frac{\mu_2}{X_2} \right] &\triangleq \int_0^{2\pi} \int_0^\infty \frac{ve^{j\phi}}{re^{j\theta}} f(r, \theta) dr d\theta \\
&= \frac{v}{\sigma_2^2} \int_0^\infty \exp\left(\frac{-r^2 - v^2}{\sigma_2^2}\right) \left[ \frac{1}{\pi} \int_0^{2\pi} e^{\frac{2vr}{\sigma_2^2} \cos(\theta-\phi)} e^{-j(\theta-\phi)} d\theta \right] dr \\
&\stackrel{(a)}{=} \frac{2v}{\sigma_2^2} \int_0^\infty \exp\left(\frac{-r^2 - v^2}{\sigma_2^2}\right) I_1\left(\frac{2vr}{\sigma_2^2}\right) dr \\
&= e^{-v^2/\sigma_2^2} \frac{2v}{\sigma_2} \int_0^\infty e^{-x^2} I_1\left(\frac{2v}{\sigma_2}x\right) dx \stackrel{(b)}{=} 1 - e^{-v^2/\sigma_2^2}, \tag{A.3}
\end{aligned}$$

where  $I_1(\cdot)$  is the first-order modified Bessel function of the first kind and the change of variable  $x = r/\sigma_2$  is applied. Equation (a) follows from

$$\begin{aligned}
\frac{1}{\pi} \int_0^{2\pi} e^{z \cos(\theta-\phi)} e^{-j(\theta-\phi)} d\theta &= \frac{1}{\pi} \int_{-\pi}^{\pi} e^{z \cos \psi} e^{-j\psi} d\psi \\
&= \frac{1}{\pi} \int_{-\pi}^{\pi} e^{z \cos \psi} \cos \psi d\psi - j \frac{1}{\pi} \int_{-\pi}^{\pi} e^{z \cos \psi} \sin \psi d\psi \\
&= \frac{2}{\pi} \int_0^{\pi} e^{z \cos \psi} \cos \psi d\psi = 2I_1(z), \tag{A.4}
\end{aligned}$$

where the first step follows from periodicity of the integrand and the change of variable  $\psi = \theta - \phi$ , the third step by observing that one of the two preceding integrands is even and the other is odd, and the last step from [3, eq. 9.6.19]. Equation (b) follows by observing

$$\begin{aligned}
\frac{2v}{\sigma_2} \int_0^\infty e^{-x^2} I_1\left(\frac{2v}{\sigma_2}x\right) dx &= \frac{2v}{\sigma_2} \int_0^\infty e^{-x^2} \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2} \frac{2v}{\sigma_2} x\right)^{2k+1}}{k!(k+1)!} dx \\
&= 2 \sum_{k=0}^{\infty} \frac{(v/\sigma_2)^{2k+2}}{k!(k+1)!} \int_0^\infty e^{-x^2} x^{2k+1} dx \\
&= 2 \sum_{k=0}^{\infty} \frac{(v/\sigma_2)^{2k+2}}{k!(k+1)!} \frac{k!}{2} = e^{v^2/\sigma_2^2} - 1, \tag{A.5}
\end{aligned}$$

where the first step follows from the series [3, eq. 9.6.10]

$$I_1(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{2k+1}}{k!(k+1)!}.$$

The expression in (2.39) now follows by substituting (A.3) into (A.1) and observing  $v = |\mu_2|$ .

We define

$$X \triangleq (X_1 - \mu_1) - \frac{\rho\sigma_1}{\sigma_2}(X_2 - \mu_2). \quad (\text{A.6})$$

It follows that  $X$  is complex Gaussian,  $X \sim \mathcal{CN}(0, (1 - |\rho|^2)\sigma_1^2)$ , and independent to  $X_2$ , due to the fact uncorrelated complex Gaussian RV's are also independent [45, pg.558]. The first absolute moment of general complex Gaussian ratio, with arbitrary means and correlation, is derived to be finite,

$$\begin{aligned} E \left[ \left| \frac{X_1}{X_2} \right| \right] &= E \left[ \left| \frac{X + \mu_1 + \rho\sigma_1(X_2 - \mu_2)/\sigma_2}{X_2} \right| \right] \\ &\leq \left| \frac{\rho\sigma_1}{\sigma_2} \right| + E \left[ \left| X + \mu_1 - \frac{\rho\sigma_1\mu_2}{\sigma_2} \right| \right] E \left[ \frac{1}{|X_2|} \right] \\ &\leq \frac{|\rho|\sigma_1}{\sigma_2} + \frac{\sqrt{\pi(1 - |\rho|^2)}\sigma_1}{\sigma_2} \exp\left(-\frac{v^2}{2\sigma_2^2}\right) I_0\left(\frac{v^2}{2\sigma_2^2}\right) < \infty, \quad (\text{A.7}) \end{aligned}$$

where the last inequality follows from the finiteness of  $I_0(\cdot)$ , and the formula of  $E[1/|X_2|]$  for arbitrary complex Gaussian RV,  $X_2 \sim \mathcal{CN}(ve^{j\phi}, \sigma_2^2)$ , is from [80, eq. 37]. Similarly,

we can prove the second moment is unbounded,

$$\begin{aligned}
E \left[ \left| \frac{X_1}{X_2} \right|^2 \right] &= E \left[ \left| \frac{\rho\sigma_1}{\sigma_2} + \frac{X + \mu_1 - \frac{\rho\sigma_1\mu_2}{\sigma_2}}{X_2} \right|^2 \right] \\
&= \left| \frac{\rho\sigma_1}{\sigma_2} \right|^2 + E \left[ \left| X + \mu_1 - \frac{\rho\sigma_1\mu_2}{\sigma_2} \right|^2 \right] E \left[ \frac{1}{|X_2|^2} \right] \\
&\geq \sigma_1^2 (1 - |\rho|^2) E \left[ \frac{1}{|X_2|^2} \right] \rightarrow \infty, \quad \forall |\rho| \neq 1, \quad (\text{A.8})
\end{aligned}$$

where the unboundedness of  $E [1/|X_2|^2]$  is applied [80, eq. 40]. Finally, with Lyapunov inequality, we have for  $k > 2$ ,

$$\left( E \left[ \left| \frac{X_1}{X_2} \right|^k \right] \right)^{\frac{1}{k}} \geq \left( E \left[ \left| \frac{X_1}{X_2} \right|^2 \right] \right)^{\frac{1}{2}} \rightarrow \infty. \quad (\text{A.9})$$

This completes the proof. ◇

# Appendix B

## Details on Chapter 3

### B.1 Another ML Estimator for $F$

Consider another classical estimation framework, where no information on channel distribution is available. Thus, both the channel  $\mathbf{h}$  and  $F$  are modeled as deterministic unknowns.

We define the unknown parameters as

$$\zeta \triangleq \begin{bmatrix} \mathbf{h} \\ F \end{bmatrix}, \quad (\text{B.1})$$

where both  $\mathbf{h}$  and  $F$  are viewed as deterministic. The log-likelihood function of  $\mathbf{V}_s$  (3.84) is

$$\mathcal{L}(\zeta) \triangleq \sigma_n^2 \ln p(\mathbf{V}_s; \mathbf{h}, F) = -(\mathbf{y}_1 - \mathbf{h})^H(\mathbf{y}_1 - \mathbf{h}) - (\mathbf{y}_2 - F\mathbf{h})^H(\mathbf{y}_2 - F\mathbf{h}). \quad (\text{B.2})$$

In order to find the ML estimator for  $\zeta$ , it is necessary that the gradient vanishes,

$$\frac{\partial \mathcal{L}(\zeta)}{\partial \zeta^*} = \begin{bmatrix} (\mathbf{y}_1 - \mathbf{h}) + F^*(\mathbf{y}_2 - F\mathbf{h}) \\ \mathbf{h}^H(\mathbf{y}_2 - F\mathbf{h}) \end{bmatrix} = \mathbf{0}_{L+1} . \quad (\text{B.3})$$

Solving the first equation for  $\mathbf{h}$  and plugging into the second, we obtain a complex quadratic equation of  $F$ , i.e.,  $(\mathbf{y}_1^H + F\mathbf{y}_2^H)(\mathbf{y}_2 - F\mathbf{y}_1) = 0$ , which, using (3.87), simplifies into

$$S_{12}F^2 + (S_{11} - S_{22})F - S_{21} = 0 . \quad (\text{B.4})$$

This is the same complex quadratic equation and leads to the same consistent estimator in (3.29).

## B.2 Derivation of Complex Gradient and Complex Hessian

Given that the  $k$ -th correlation matrix is

$$\mathbf{C}_k(\boldsymbol{\theta}) = \begin{bmatrix} \lambda_k + \sigma_n^2 & \lambda_k F^* \\ \lambda_k F & \lambda_k |F|^2 + \sigma_n^2 \end{bmatrix}, \quad \lambda_k \triangleq d_k \sigma_h^2 ,$$

we derive the following identities related to the complex gradient of  $\mathbf{C}_k$  w.r.t.  $\boldsymbol{\theta}$ .

$$\begin{aligned}
\frac{\partial \mathbf{C}_k}{\partial F^*} &= \lambda_k \begin{bmatrix} 0 & 1 \\ 0 & F \end{bmatrix} = \lambda_k \begin{bmatrix} 1 \\ F \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} = \lambda_k \mathbf{e}_1 \begin{bmatrix} 0 & \sqrt{1+|F|^2} \end{bmatrix}, \\
\mathbf{C}_k^{-1} \frac{\partial \mathbf{C}_k}{\partial F^*} &= (\mu_1^{-1} \mathbf{e}_1 \mathbf{e}_1^H + \mu_2^{-1} \mathbf{e}_2 \mathbf{e}_2^H) \frac{\partial \mathbf{C}_k}{\partial F^*} = \frac{\lambda_k}{\mu_1} \mathbf{e}_1 \begin{bmatrix} 0 & \sqrt{1+|F|^2} \end{bmatrix}, \\
\mathbf{C}_k^{-1} \frac{\partial \mathbf{C}_k}{\partial F^*} \mathbf{C}_k^{-1} &= \frac{\lambda_k}{\mu_1} \mathbf{e}_1 \begin{bmatrix} 0 & \sqrt{1+|F|^2} \end{bmatrix} (\mu_1^{-1} \mathbf{e}_1 \mathbf{e}_1^H + \mu_2^{-1} \mathbf{e}_2 \mathbf{e}_2^H) \\
&= \frac{\lambda_k F}{\mu_1^2} \mathbf{e}_1 \mathbf{e}_1^H + \frac{\lambda_k}{\mu_1 \mu_2} \mathbf{e}_1 \mathbf{e}_2^H = \frac{1}{\sigma_n^2} \frac{\lambda_k}{\mu_1^2} \begin{bmatrix} 1 \\ F \end{bmatrix} \begin{bmatrix} -F \lambda_k & \sigma_n^2 + \lambda_k \end{bmatrix}, \quad (\text{B.5})
\end{aligned}$$

and

$$\frac{\partial \mathbf{C}_k}{\partial F} = \lambda_k \begin{bmatrix} 0 & 0 \\ 1 & F^* \end{bmatrix} = \lambda_k \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & F^* \end{bmatrix}, \quad (\text{B.6})$$

and, assuming  $\lambda_k = d_k \sigma_h^2$ ,

$$\begin{aligned}
\frac{\partial \mathbf{C}_k}{\partial \sigma_h^2} &= d_k \begin{bmatrix} 1 & F^* \\ F & |F|^2 \end{bmatrix} = d_k (1 + |F|^2) \mathbf{e}_1 \mathbf{e}_1^H, \\
\mathbf{C}_k^{-1} \frac{\partial \mathbf{C}_k}{\partial \sigma_h^2} &= (\mu_1^{-1} \mathbf{e}_1 \mathbf{e}_1^H + \mu_2^{-1} \mathbf{e}_2 \mathbf{e}_2^H) \frac{\partial \mathbf{C}_k}{\partial \sigma_h^2} = \frac{d_k (1 + |F|^2)}{\mu_1} \mathbf{e}_1 \mathbf{e}_1^H, \\
\mathbf{C}_k^{-1} \frac{\partial \mathbf{C}_k}{\partial \sigma_h^2} \mathbf{C}_k^{-1} &= \frac{d_k (1 + |F|^2)}{\mu_1^2} \mathbf{e}_1 \mathbf{e}_1^H. \quad (\text{B.7})
\end{aligned}$$

# Appendix C

## Details on Chapter 4

### C.1 Details on Miller-Chang Bound

We start from the vectorized sufficient statistics given in (3.41),

$$\mathbf{v} \triangleq \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \text{vec } \mathbf{Y}_1 \\ \text{vec } \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{h} \\ \text{vec } \mathbf{F}\mathbf{H} \end{bmatrix} + \mathbf{n}, \quad (\text{C.1})$$

where  $\mathbf{h} \triangleq \text{vec } \mathbf{H} \sim \mathcal{CN}(\mathbf{0}_{MN}, \mathbf{I}_N \otimes \boldsymbol{\Sigma}_{\mathbf{h}})$  (2.11) is the vectorized channel information, and  $\mathbf{n}$  is a zero-mean, complex Gaussian random vectors independent to  $\mathbf{h}$ , i.e.,  $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{2MN})$ .

Define a log-likelihood function (LLF) of the conditional PDF,

$$\begin{aligned} \mathcal{L}(\boldsymbol{\xi}) &\triangleq \ln p(\mathbf{v}|\mathbf{h}; \mathbf{f}) \\ &= -\frac{1}{\sigma_n^2} (\mathbf{y}_1 - \mathbf{h})^H (\mathbf{y}_1 - \mathbf{h}) - \frac{1}{\sigma_n^2} (\mathbf{y}_2 - \text{vec } \mathbf{F}\mathbf{H})^H (\mathbf{y}_2 - \text{vec } \mathbf{F}\mathbf{H}), \quad (\text{C.2}) \end{aligned}$$

where  $\mathbf{f} \triangleq \text{vec } \mathbf{F}$  and the unknown parameters are

$$\boldsymbol{\xi} \triangleq \begin{bmatrix} \mathbf{h} \\ \mathbf{f} \end{bmatrix}. \quad (\text{C.3})$$

Note two identifies exist for vectorizing  $\mathbf{FH}$ , i.e.,

$$\text{vec } \mathbf{FH} = (\mathbf{I}_N \otimes \mathbf{F}) \mathbf{h} = (\mathbf{H}^T \otimes \mathbf{I}_M) \mathbf{f}. \quad (\text{C.4})$$

In order to calculate the Miller-Chang bound (MCB), we need the Hessian of  $\mathcal{L}(\boldsymbol{\xi})$ . But first, the complex gradient w.r.t.  $\boldsymbol{\xi}$  is calculated as,

$$\begin{aligned} \frac{\partial \mathcal{L}(\boldsymbol{\xi})}{\partial \mathbf{h}^*} &= \frac{1}{\sigma_n^2} (\mathbf{y}_1 - \mathbf{h}) + \frac{1}{\sigma_n^2} (\mathbf{I}_N \otimes \mathbf{F}^H) (\mathbf{y}_2 - \text{vec } \mathbf{FH}), \\ \frac{\partial \mathcal{L}(\boldsymbol{\xi})}{\partial \mathbf{f}^*} &= \frac{1}{\sigma_n^2} (\mathbf{H}^* \otimes \mathbf{I}_M) (\mathbf{y}_2 - \text{vec } \mathbf{FH}). \end{aligned} \quad (\text{C.5})$$

Then the complex Hessian of  $\mathcal{L}(\boldsymbol{\xi})$  is

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}^* \partial \boldsymbol{\xi}^T} \triangleq \begin{bmatrix} \frac{\partial^2 \mathcal{L}(\mathbf{h}, \mathbf{f})}{\partial \mathbf{h}^* \partial \mathbf{h}^T} & \frac{\partial^2 \mathcal{L}(\mathbf{h}, \mathbf{f})}{\partial \mathbf{h}^* \partial \mathbf{f}^T} \\ \frac{\partial^2 \mathcal{L}(\mathbf{h}, \mathbf{f})}{\partial \mathbf{f}^* \partial \mathbf{h}^T} & \frac{\partial^2 \mathcal{L}(\mathbf{h}, \mathbf{f})}{\partial \mathbf{f}^* \partial \mathbf{f}^T} \end{bmatrix} = -\frac{1}{\sigma_n^2} \begin{bmatrix} \mathbf{I}_N \otimes (\mathbf{F}^H \mathbf{F} + \mathbf{I}_M) & \mathbf{H}^T \otimes \mathbf{F}^H \\ \mathbf{H}^* \otimes \mathbf{F} & \mathbf{H}^* \mathbf{H}^T \otimes \mathbf{I}_M \end{bmatrix}. \quad (\text{C.6})$$

The Miller-Chang bound is defined as [56, Sec. III, eq. 30],

$$\begin{aligned} \mathcal{M}(\boldsymbol{\xi}) &= -E_{\mathbf{h}} \left[ \left\{ \frac{\partial^2 \mathcal{L}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}^* \partial \boldsymbol{\xi}^T} \right\}^{-1} \right] \\ &= \sigma_n^2 E_{\mathbf{h}} \left[ \begin{bmatrix} (\mathbf{I}_N - \mathbf{P}) \otimes \mathbf{A}_1^{-1} + \mathbf{P} \otimes \mathbf{I}_M & -\mathbf{H}^T (\mathbf{H}^* \mathbf{H}^T)^{-1} \otimes \mathbf{F} \\ -(\mathbf{H}^* \mathbf{H}^T)^{-1} \mathbf{H}^* \otimes \mathbf{F}^H & (\mathbf{H}^* \mathbf{H}^T)^{-1} \otimes \mathbf{A}_2 \end{bmatrix} \right] \end{aligned} \quad (\text{C.7})$$

where the expectation is over  $\mathbf{h}$  (2.11), the second equality uses the matrix inverse lemma [70, pg. 275, eq. A1.42],  $\mathbf{P} \triangleq \mathbf{H}^T (\mathbf{H}^* \mathbf{H}^T)^{-1} \mathbf{H}^*$  is a complex projection matrix [45, pg. 231], and we have previously defined  $\mathbf{A}_1 = \mathbf{I}_M + \mathbf{F}^H \mathbf{F}$  and  $\mathbf{A}_2 = \mathbf{I}_M + \mathbf{F} \mathbf{F}^H$  in (4.31). The lower bound pertaining estimating  $\mathbf{f}$  is the trace of the lower right block matrix, i.e.,

$$\mathcal{M}(\mathbf{F}) = \sigma_n^2 \text{Tr} \left( E \left[ (\mathbf{H}^* \mathbf{H}^T)^{-1} \otimes \mathbf{A}_2 \right] \right) = \sigma_n^2 \text{Tr} \left( E \left[ (\mathbf{H}^* \mathbf{H}^T)^{-1} \right] \right) \text{Tr}[\mathbf{A}_2], \quad (\text{C.8})$$

where we have used  $\text{Tr}[\mathbf{C} \otimes \mathbf{D}] = \text{Tr}[\mathbf{C}] \text{Tr}[\mathbf{D}]$  [19, Tab. II].

## C.2 Another Classical ML Estimator

**Lemma 7** *The system by setting the gradients in (C.5) to zero leads to an ML estimator for  $\mathbf{F}$ , which satisfies the following necessary condition, i.e.,*

$$\mathbf{Y}_2 \mathbf{Y}_1^H - \mathbf{F} \mathbf{Y}_1 \mathbf{Y}_1^H + \mathbf{Y}_2 \mathbf{Y}_2^H \mathbf{F} - \mathbf{F} \mathbf{Y}_1 \mathbf{Y}_2^H \mathbf{F} = \mathbf{0}_{M \times M}. \quad (\text{C.9})$$

**Proof** We set the gradients (C.5) of the log-likelihood function  $\mathcal{L}(\boldsymbol{\xi})$  to zero. The first equation of (C.5) results in a matrix equation,

$$\mathbf{H} = \mathbf{A}_1^{-1} (\mathbf{Y}_1 + \mathbf{F}^H \mathbf{Y}_2), \quad (\text{C.10})$$

where  $\mathbf{A}_1 = \mathbf{I}_M + \mathbf{F}^H \mathbf{F}$  (4.31) has been defined previously. Next letting the second equation equal zero, we have

$$(\mathbf{Y}_2 - \mathbf{F} \mathbf{H}) \mathbf{H}^H = \mathbf{0}_{M \times N}. \quad (\text{C.11})$$

The proof completes after plugging the first matrix equation into the second and some straightforward simplification.

Among all the roots of this matrix equation (C.9), the one, along with its corresponding  $\mathbf{H}$  estimates, that maximizes the log-likelihood function given in (C.2) is the ML estimator for  $\mathbf{F}$ .

If we assign the sample covariance defined in (4.21) symbolic submatrices as

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix}, \quad (\text{C.12})$$

where  $\mathbf{S}_{ij} \in \mathbb{C}^{M \times M}$ , then the above matrix equation becomes,

$$\mathbf{S}_{21} - \mathbf{F}\mathbf{S}_{11} + \mathbf{S}_{22}\mathbf{F} - \mathbf{F}\mathbf{S}_{12}\mathbf{F} = \mathbf{0}_{M \times M}. \quad (\text{C.13})$$

It can be shown this is the same necessary condition that  $\hat{F}_{ML}$  in (4.40) satisfies, i.e., it is a function of a eigen-vector matrix that block diagonalizes  $\mathbf{S}$ . It remains unclear, though we conjecture it is true, if solutions of the above equation lead to a unique estimator as  $\hat{\mathbf{F}}_{ML}$  in (4.40).