11-10-2016

Channel and Noise Variance Estimation for Future 5G Cellular Networks

Jorge Iscar Vergara
Florida International University, jisca001@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

CHANNEL AND NOISE VARIANCE ESTIMATION FOR FUTURE 5G CELLULAR NETWORKS

A thesis submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

ELECTRICAL ENGINEERING

by

Jorge Iscar Vergara

2016

To: Interim Dean Ranu Jung
College of Engineering and Computing

This thesis, written by Jorge Iscar Vergara, and entitled Channel and Noise Variance Estimation for Future 5G Cellular Networks, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

A. Selcuk Uluagac

Ahmed Ibrahim

Hai Deng

İsmail Güvenç, Major Professor

Date of Defense: November 10, 2016

The thesis of Jorge Iscar Vergara is approved.

Interim Dean Ranu Jung
College of Engineering and Computing

Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2016

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor Professor İsmail Güvenç, whose support has helped me to complete this work. Furthermore, I would like to thank Nadisanka Rupasinghe and Sener Dikmese, who have also contributed to the production of this thesis.

This research was supported in part by the National Science Foundation under the award number CNS-1453678 and the Finnish Cultural Foundation.

ABSTRACT OF THE THESIS

CHANNEL AND NOISE VARIANCE ESTIMATION FOR FUTURE 5G CELLULAR NETWORKS

by

Jorge Iscar Vergara

Florida International University, 2016

Miami, Florida

Professor İsmail Güvenç, Major Professor

Future fifth generation (5G) cellular networks have to cope with the expected ten-fold increase in mobile data traffic between 2015 and 2021. To achieve this goal, new technologies are being considered, including massive multiple-input multiple-output (MIMO) systems and millimeter-wave (mmWave) communications. Massive MIMO involves the use of large antenna array sizes at the base station, while mmWave communications employ frequencies between 30 and 300 GHz. In this thesis we study the impact of these technologies on the performance of channel estimators.

Our results show that the characteristics of the propagation channel at mmWave frequencies improve the channel estimation performance in comparison with current, low frequency-based, cellular networks. Furthermore, we demonstrate the existence of an optimal angular spread of the multipath clusters, which can be used to maximize the capacity of mmWave networks. We also propose efficient noise variance estimators, which can be employed as an input to existing channel estimators.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

1G	first generation
2G	second generation
4G	fourth generation
5G	fifth generation
mmWave	millimeter-wave
MIMO	multiple-input multiple-output
BS	base station
SNR	signal to noise ratio
SINR	signal to interference plus noise ratio
MMSE	minimum mean square error
LS	least squares
AS	angular spread
ML	maximum likelihood
MM	method of moments
CRLB	Cramer-Rao lower bound
TDD	time-division duplexing
AOA	angle of arrival
DA	data aided
NDA	non-data aided
PDF	probability density function
EM	expectation-maximization
LOS	line of sight
RZF	regularized zero-forcing

CHAPTER 1

INTRODUCTION

1.1 Motivation

Mobile communication networks have evolved over the years in order to cope with the increasing demand of the users for better quality of service (coverage, Internet data rates, number of users served simultaneously, etc.). Since the first generation (1G) cellular standards were launched in the early 1980s, a new generation has been introduced approximately every 10 years, until the current fourth generation (4G) systems.

Several technologies have been developed and implemented over the past few decades to achieve the ever increasing users demands. The following are some of the most important improvements.

1. Analog to digital transition. The second generation (2G) standards brought digital communications to cellular networks. This technology made feasible the first cellular data services (text messages), and incremented the number of users that can be served simultaneously [1].

2. More efficient data transmission schemes. The communication between a user and the base station (BS) utilizes a part of the electromagnetic spectrum, which is a limited and regularized resource. Furthermore, it has been shown that the larger the amount of spectrum is used, the higher the data rates. The limited amount of this natural resource made imperative to develop more efficient data transmission schemes to reduce the portion of the spectrum employed by a given user while keeping, or even increasing, the quality of service.

3. Multiple antenna systems. The use of more than one antenna at both the BS and user terminal has helped to improve the quality of service as well. Depending

on the disposition and number of antennas, this technology can increase the coverage, data rates, and/or number of users served simultaneously.

Following the ten years interval between cellular standards, the fifth generation (5G) is expected to be released by 2020. The data in [2] shows that the mobile data traffic by that year is expected to be ten times that in 2015. In order to cope with this demand, we could continue exploiting the aforementioned technologies, however it seems that this approach will not achieve the required quality of service. Regarding the use of more efficient data transmission schemes, although new technologies are being investigated [3], this strategy is not expected to play a key role in future 5G cellular networks [4]. As we mentioned before, the data rates can also be risen by increasing the amount of electromagnetic spectrum, however the availability of this resource has become scarce in the range of the spectrum where cellular communications usually take place (microwaves frequencies) [4]. Finally, as far as multiple antenna systems are concerned, the number of antennas that can be deployed is limited due to space constraints since current working frequencies of cellular networks require relatively large antenna sizes.

The inability to continue exploiting the conventional technologies to cope with the expected mobile data traffic by 2020 has driven the development of new technologies. The work in [4] summarizes five of these technologies. This thesis focuses on two of them: millimeter-wave (mmWave) and massive multiple-input multiple-output (MIMO). The use of mmWave frequencies for cellular communications, as opposed to microwaves, offers a vast amount of available spectrum that can be used to increase the data rates. On the other hand, massive MIMO technology implies the use of a very large number of antennas at the BS so that the data rates and number of users served simultaneously may be arbitrarily high. Although we had previously discarded an increment in the number of antennas due to space constraint, this will

not be a restriction with mmWave frequencies since they require smaller antenna sizes.

The implementation of these two technologies requires research effort on several wireless communications areas. Specifically, the research of this thesis focuses on channel estimation techniques. In statistical signal processing, estimation theory studies different techniques to extract some desirable parameter from a noisy random signal. As far as the term channel is concerned, this parameter can be thought of as a signature that unambiguously characterizes a user. As a consequence, the knowledge of this parameter will allow the BS to send different data to the different users, as opposed to broadcasting where all users receive the same data. Therefore, the estimation of this parameter is of crucial importance in cellular networks since the users connected to a BS receive data independently.

1.2 Contribution of the Thesis

The work done in this thesis has resulted in the following three main contributions:

1. Derived maximum likelihood (ML) and method of moments (MM) noise variance estimators for massive MIMO systems. The corresponding Cramer-Rao lower bound (CRLB) is determined as well.
2. Proved the superior performance of minimum mean square error (MMSE) channel estimators in mmWave networks under pilot contamination, compared to current microwave cellular deployments.
3. Demonstrated the existence of an optimal, maximum capacity achieving, angular spread (AS) of the multipath clusters in mmWave systems under pilot contamination.

1.3 Organization of the Thesis

The prove of the three main contributions listed above is presented in Chapters 2, 3 and 4, respectively. Within each chapter, we first introduce the specific research problem. Then, after presenting the mathematical background, we introduce our findings. Finally, closing remarks are provided at the end of the chapter.

1.4 Notation

Throughout this work, vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. The identity matrix of size M is denoted by $\mathbf{I}_M$. Furthermore, $\mathbf{0}_M$ and $\mathbf{0}_{M\times N}$ represent a $M \times 1$ vector and a $M \times N$ matrix, respectively, whose elements are all zeros, while $\otimes$ represents the Kronecker product, $(.)^*$ the conjugate, $(.)^T$ the transpose, and $(.)^H$ the conjugate transpose (Hermitian operator). The trace and determinant of a matrix are denoted by $\mathrm{tr}\,[.]$ and $\det[.]$, respectively, and $\mathrm{diag}\,(\mathbf{x})$ represents the diagonal matrix whose diagonal elements are given by $\mathbf{x}$. The expected value, covariance and variance are represented by $\mathrm{E}\,[.]$, $\mathrm{Cov}\,[.]$, and $\mathrm{Var}\,[.]$, respectively, while $\mathcal{CN}$, $\mathcal{N}$, $\mathcal{U}$, $\mathcal{DU}$ denote the complex Gaussian, Gaussian, uniform and discrete uniform distributions.

CHAPTER 2

NOISE VARIANCE ESTIMATORS FOR MASSIVE MIMO SYSTEMS

2.1 Introduction

Future 5G cellular networks have to cope with the expected ten-fold increase in mobile data traffic between 2015 and 2021 [2]. In order to achieve this goal, new technologies are being considered, including massive MIMO systems and mmWave communications [4]. Massive MIMO involves the use of BSs with large antenna array sizes compared with the number of users [5]. The valuable result of this technology is that, under the extreme scenario of an infinite number of antennas at the BS, the capacity both in the uplink and downlink is only limited by pilot contamination. That is, the effects of noise, channel estimation error and interference vanish[1].

The pilot contamination impairment is depicted in Fig. 2.1. In time-division duplexing (TDD) cellular networks, each BS estimates the uplink channel of the users within its cell using known pilot sequences. In order to avoid intra-cell interference during the channel estimation stage, orthogonal time-frequency resources are used among the users. In TDD systems channel reciprocity applies, and therefore the BS uses these estimates to perform precoding and deliver high data rates to the users in the downlink [6]. In order to increase the efficiency of the network, aggressive frequency reuse factors are employed and the same time-frequency resources are shared among the BSs. This results in each BS receiving the pilot sequence not only from its desired user, but also from one user in every neighboring cell. This is known as the pilot contamination effect and causes a degradation in the channel estimation performance, which in turn compromises the resulting capacity [7].

[1]Here, we refer to channel estimation error and interference not related to pilot contamination itself.

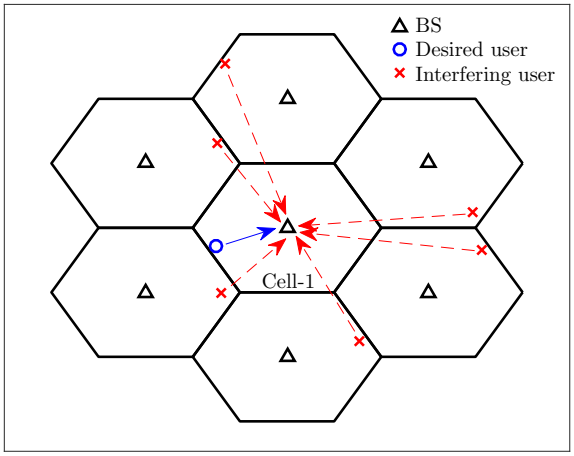


Figure 2.1: The pilot contamination effect in cellular networks. When the same time-frequency resources are used among cells, the BS estimates the channel of the desired user incorrectly, learning a combination of the target channel and those from the interfering users. This results in a degradation of the capacity.

Having the pilot contamination as the only remaining impairment in massive MIMO systems, exhaustive work has been done to overcome it; see the survey in [8] and the references therein. Of special interest is the work in [9], which shows that the pilot contamination effect depends on the mean angle of arrival (AOA) and AS of the multipath clusters of the desired and interfering users. Specifically, the authors demonstrate that the pilot contamination effect is completely eliminated when considering and infinite number of antennas at the BS, and the cluster of the desired user does not overlap with those of the interfering users in the angular domain. In practical scenarios with limited number of antennas and overlapping condition, the channel estimation error is seen to increase with the AS of the clusters. Furthermore, the authors in [9] propose a MMSE channel estimator and a pilot assignment strategy to improve the channel estimation performance under pilot contamination. The proposed estimator requires the noise variance at the receive antennas, which is assumed known. However, this assumption may be challenged under certain circumstances

(see Section 2.7), and, therefore, it has to be estimated. This is the objective of our work, the estimation of the noise variance at the receive antennas of massive MIMO systems, which can then be employed to estimate the channel as in [9].

2.1.1 Related work

The work that comes closest to ours is [10]. In that work, the authors employ a MMSE noise variance estimator, which depends on the channel estimate. Since the channel estimate depends, in turn, on the noise variance, an iterative algorithm is needed. In our work, however, the noise variance is estimated directly from the received samples and iterative processes are avoided.

In [11], the noise variance estimate also depends on the channel estimate. In order to avoid iterative algorithms, the authors substitute the MMSE channel estimate by the least squares (LS) estimate. As we will observe in our simulation results, the low performance of the LS channel estimator [9] compromises the noise variance estimation.

In [12], the authors estimate the noise variance using the sample covariance matrix, which requires the average of several received samples. This results in a delay between noise variance estimates, which compromises the performance of the estimator in scenarios where the noise variance is a non-stationary parameter (see Section 2.7). On the other hand, in our work, the noise variance is estimated continuously from the last received samples.

Finally, the noise variance estimator proposed in [13] is not suitable for massive MIMO systems since the output of the estimator may lead to the computation of the inverse of the channel covariance matrix during the channel estimation stage. However, the assumption of channel covariance matrix invertibility is challenged in massive MIMO scenarios due to rank deficiency [9].

2.1.2 Our contribution

We include the following contributions in this chapter.

- We propose non-iterative ML estimators of the noise variance in MIMO systems under three different scenarios: data aided (DA) model (known pilot sequence), non-data aided (NDA) model (unknown pilot sequence), and mixed DA and NDA model. Furthermore, for each scenario, two noise models are considered: the same or different noise variance at the receive antennas. Simulations results show that these estimators are *efficient* in certain scenarios since they attain the CRLB. Although neither the channel nor its estimate are an input parameter to the estimator, its covariance matrix needs to be known.
- We also propose MM estimators of the noise variance. Although the efficiency of these estimators is limited to very low signal to noise ratio (SNR) values, their computational complexity is reduced when compared to the ML estimators. The channel covariance matrix is also required for the MM estimators.
- We derive the CRLB for the noise variance estimation under the three different scenarios and noise models. The CRLB for scenarios where the channel is known has been studied in [14]. As a consequence, our proposed expressions are a generalization of those derived in [14] to account for scenarios where the channel is unknown.

The output of the proposed noise variance estimators can then be used to estimate the channel as in [9]. Although we are investigating massive MIMO scenarios, the proposed CRLBs and estimators for the noise variance are also valid for current antenna array sizes.

2.1.3 Organization of the chapter

The rest of the chapter is organized as follows. Section 2.2 describes the system model. In Section 2.3 we derive the CRLB for the noise variance estimation under the different pilot and noise models. For each of these scenarios, in Section 2.4 we develop ML and MM estimators for the noise variance. Furthermore, approximations for the CRLB and ML estimators under specific scenarios are derived in Section 2.5. Simulation results for the proposed CRLB and noise variance estimators are shown in Section 2.6. In Section 2.7 we discuss the availability of the channel covariance matrix, which is used in the noise variance estimators, as well as other alternative strategies to obtain the noise variance and the advantages of ours. Finally, in Section 2.8 we provide closing remarks.

2.2 System Model

We consider a cellular system with N_{L} cells, each of them managed by a BS equipped with M antennas. Furthermore, K single-antenna users populate each cell. We assume TDD and constant channels during the channel estimation stage in both the time (block fading) and frequency (flat fading) domains. Besides, we consider narrowband signals, which satisfies the assumption of flat fading channels. This scenario is depicted in Fig. 2.2.

During the uplink channel estimation stage, the K users in each cell transmit a pilot sequence within orthogonal time-frequency resources. In cellular networks with aggressive frequency reuse factors, the transmission resources are shared among the N_{L} cells, which results in each BS receiving $N_{\mathrm{L}} - 1$ interfering pilot sequences, causing the pilot contamination effect. Therefore, the received signal $\mathbf{Y}$ at the BS

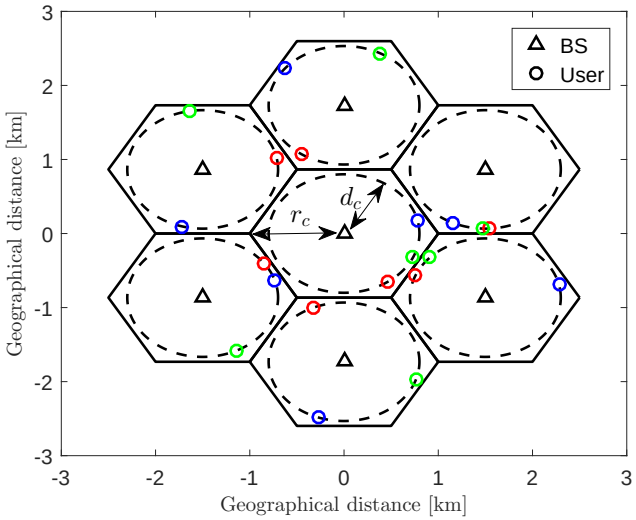


Figure 2.2: System model. Example for $N_{\mathrm{L}} = 7$ and $K = 3$. The users represented by the same color are assigned the same time-frequency resources, which causes the effect of pilot contamination.

corresponding to one of the K users is given by

$$\mathbf{Y} = \sum_{l=1}^{N_{\mathrm{L}}} \mathbf{h}_l \mathbf{s}^T + \mathbf{N} \ , \tag{2.1}$$

where $\mathbf{h}_l$ is the uplink channel from the user in the l-th cell to the BS, $\mathbf{s}$ is the pilot sequence, and $\mathbf{N}$ is the noise at the receive antennas. The dimensions of the variables are as follows: $\mathbf{h}$ is a $M \times 1$ vector, $\mathbf{s} = [s_1 \cdot \cdot \cdot \ s_\tau]^T$ is a $\tau \times 1$ vector, and $\mathbf{Y}$ and $\mathbf{N}$ are $M \times \tau$ matrices.

The column vector form of the matrix $\mathbf{Y}$ is given by

$$\mathbf{y} = \mathbf{S} \sum_{l=1}^{N_{\mathrm{L}}} \mathbf{h}_l + \mathbf{n} \ , \tag{2.2}$$

where $\mathbf{y}$ and $\mathbf{n}$ are obtained by stacking all columns of $\mathbf{Y}$ and $\mathbf{N}$, respectively, into one column vector, and $\mathbf{S} = \mathbf{s} \otimes \mathbf{I}_M$ is the training matrix of size $M\tau \times M$. We consider that each of the τ symbols of $\mathbf{s}$ has unit power and therefore $\mathbf{S}^H\mathbf{S} = \tau\mathbf{I}_M$. Finally, both the channels $\mathbf{h}_l$ and the noise $\mathbf{n}$ are considered random variables distributed as $\mathbf{h}_l \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{R}_l)$ and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}_{M\tau}, \mathbf{\Sigma})$. The noise covariance matrix $\mathbf{\Sigma}$ is a diagonal matrix and therefore the noise samples are spatially and temporally uncorrelated.

From this system model, the MMSE estimate of the channel of the desired user is given by [9]

$$\widehat{\mathbf{h}_1} = \mathbf{R}_1 \left(\boldsymbol{\Sigma} + \tau \sum_{l=1}^{N_\mathrm{L}} \mathbf{R}_l \right)^{-1} \mathbf{S}^H \mathbf{y} \ , \tag{2.3}$$

where we consider the estimation of the desired channel in cell 1 without loss of generality. The objective of our work is the estimation of $\boldsymbol{\Sigma}$ in (2.3) from the received signal $\mathbf{y}$ for the scenarios where $\boldsymbol{\Sigma}$ is an unknown parameter.

In order to estimate the noise covariance matrix we simplify the signal model in (2.2). As a result, the received signal at the BS becomes

$$\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{n} \ , \tag{2.4}$$

where $\mathbf{h} = \sum_{l=1}^{N_\mathrm{L}} \mathbf{h}_l$ results from summing all the N_L channels $\mathbf{h}_l$, such that $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{R})$. Since the channels are independent, the resulting covariance matrix $\mathbf{R}$ is given by

$$\mathbf{R} = \sum_{l=1}^{N_\mathrm{L}} \mathbf{R}_l \ . \tag{2.5}$$

2.3 CRLB for Noise Variance Estimation

The CRLB expresses the best achievable variance (lowest error) of any unbiased estimator. We will use this lower bound to measure the performance of the estimators proposed in the next section.

Extending the work in [14] to scenarios where the channel is unknown, we derive the CRLB under different pilot sequence models: DA model (the pilot sequence is known), NDA model (the pilot sequence is unknown), and mixed DA and NDA model (the pilot sequence is only partially known). Furthermore, for each of these scenarios, we consider two different noise models: the noise variance σ^2 at the BS antennas is the same, that is $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{M\tau}$, and different, which results in $\boldsymbol{\Sigma} = \mathbf{I}_\tau \otimes \mathrm{diag}\,(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = [\sigma_1^2 \cdot \cdot \cdot \ \sigma_M^2]$, and σ_m^2 is the noise variance at the m-th antenna.

The pilot symbols that form the sequence $\mathbf{s}$ belong to a finite constellation of N equiprobable and independent unit power symbols (i.e., N-PSK modulation). Additionally, $\mathbf{S}$, $\mathbf{h}$, and $\mathbf{n}$ are considered independent random variables.

2.3.1 CRLB for DA model with equal noise variance

In the DA model the pilot sequence is known and therefore $\mathbf{S}$ can be considered a deterministic variable. As a consequence, the received signal $\mathbf{y}$ in (2.4) is distributed as $\mathbf{y} \sim \mathcal{CN}\left(\boldsymbol{\mu}_{\mathbf{y}}, \mathbf{C}_{\mathbf{y}}\right)$, where the mean $\boldsymbol{\mu}_{\mathbf{y}}$ and covariance matrix $\mathbf{C}_{\mathbf{y}}$ are given by

$$\boldsymbol{\mu}_{\mathbf{y}} = \mathrm{E}\left[\mathbf{Sh} + \mathbf{n}\right] = \mathbf{S}\mathrm{E}\left[\mathbf{h}\right] + \mathrm{E}\left[\mathbf{n}\right] = \mathbf{0}_{M\tau} \ , \tag{2.6}$$

$$\mathbf{C}_{\mathbf{y}} = \mathrm{Cov}\left[\mathbf{Sh} + \mathbf{n}\right] = \mathrm{Cov}\left[\mathbf{Sh}\right] + \mathrm{Cov}\left[\mathbf{n}\right] = \mathbf{S}\mathbf{R}\mathbf{S}^H + \sigma^2\mathbf{I}_{M\tau} \ , \tag{2.7}$$

where the expected value is taken with respect to $\mathbf{y}$.

The CRLB for the estimation of σ^2 is defined as [15, ch. 3]

$$\mathrm{Var}\left[\widehat{\sigma^2}\right] \geq I\left(\sigma^2\right)^{-1} \ , \tag{2.8}$$

where the Fisher information $I\left(\sigma^2\right)$ for the case of Gaussian random variables is given by [15, ch. 15]

$$I\left(\sigma^2\right) = \mathrm{tr}\left[\left(\mathbf{C}_{\mathbf{y}}^{-1}\frac{\partial \mathbf{C}_{\mathbf{y}}}{\partial \sigma^2}\right)^2\right] + 2\mathrm{Re}\left[\frac{\partial \boldsymbol{\mu}_{\mathbf{y}}^H}{\partial \sigma^2}\mathbf{C}_{\mathbf{y}}^{-1}\frac{\partial \boldsymbol{\mu}_{\mathbf{y}}}{\partial \sigma^2}\right] \ , \tag{2.9}$$

where $\mathrm{Re}\left[x\right]$ denotes the real part of x.

Theorem 2.3.1 Consider the DA model with equal noise at the receive antennas. Then, the CRLB for the estimation of the noise variance is given by

$$\mathrm{Var}\left[\widehat{\sigma^2}\right] \geq I\left(\sigma^2\right)^{-1} = \frac{\sigma^4}{M\left(\tau - 1\right) + \sum\limits_{m=1}^{M}\frac{1}{\left(1 + \frac{\tau\lambda_m}{\sigma^2}\right)^2}} \ , \tag{2.10}$$

where λ_m are the eigenvalues of $\mathbf{R}$.

Proof. See Appendix A.1. □

The CRLB for the noise variance estimation above considers a random and unknown channel. As a consequence, this is a generalization of the expression derived in [14], which considers a known channel. We can, then, obtain that same expression by removing the channel uncertainty from (2.10). This is done by setting all the elements of the channel covariance matrix $\mathbf{R}$ to zero, that is, $\mathbf{R} = \mathbf{0}_{M\times M}$, which results in $\lambda_m = 0, \forall m$. Therefore, when the channel is known, the CRLB reduces to [14, eq. (18)]

$$\text{Var}\left[\widehat{\sigma^2}\right] \geq \frac{\sigma^4}{M\tau} \ . \tag{2.11}$$

We will analyze how the different variables in (2.10) affect the CRLB when discussing the simulation results in Section 2.6.

2.3.2 CRLB for NDA model with equal noise variance

The pilot sequence is always known at the receiver in typical wireless systems. In this work, an unknown pilot sequence entails scenarios where the noise variance is estimated from the user data, which is unknown. As we will observe in the simulation results, the estimation error increases in this scenario. On the other hand, this approach is not meant to be used alone, but in conjunction with a known pilot sequence. In this mixed strategy, to estimate the noise variance we will employ the received signal from the known pilot sequence, which is used to estimate the channel as in (2.3) as well, in conjunction with the received signal from the unknown user data. This approach, which is analyzed in the next subsection, offers a better performance than that of the DA model, as we will observe in the simulation results.

For clarification, we introduce the variable κ to represent the number of unknown symbols, while τ will still be representing the number of known pilot symbols. As a consequence, in this subsection, we only consider the variable κ.

When the received signal comes from an unknown pilot sequence, the training matrix $\mathbf{S}$ is no longer a deterministic variable, and therefore the probability density function (PDF) of the received signal $p(\mathbf{y})$ is not Gaussian. As a consequence, the expression for the Fisher information in (2.9) cannot be used to compute the CRLB, and we have to rely on the general expression, which is given by

$$I\left(\sigma^2\right) = -\mathrm{E}\left[\frac{\partial^2 \ln p\left(\mathbf{y};\sigma^2\right)}{\partial\left(\sigma^2\right)^2}\right] . \tag{2.12}$$

Theorem 2.3.2 Consider the NDA model with equal noise at the receive antennas. Then, the second derivative of $\ln p\left(\mathbf{y};\sigma^2\right)$ with respect to σ^2 is defined as

$$\begin{aligned}\frac{\partial^2 \ln p\left(\mathbf{y};\sigma^2\right)}{\partial\left(\sigma^2\right)^2} =& \frac{M\kappa}{\sigma^4} - \sum_{m=1}^{M} \frac{\frac{\left(\lambda_m\kappa\right)^3}{\sigma^4} + \frac{3\left(\lambda_m\kappa\right)^2}{\sigma^2} + 2\lambda_m\kappa}{\left(\sigma^2+\lambda_m\kappa\right)^3} \\ & - 2\frac{\mathbf{y}^H\mathbf{y}}{\sigma^6} + \frac{\partial^2 \ln\left(\sum_{s_1}\cdots\sum_{s_\kappa} T\right)}{\partial\left(\sigma^2\right)^2} ,\end{aligned} \tag{2.13}$$

where

$$\frac{\partial^2 \ln\left(\sum_{s_1}\cdots\sum_{s_\kappa} T\right)}{\partial\left(\sigma^2\right)^2} = \frac{\left(\sum_{s_1}\cdots\sum_{s_\kappa}(TD)\right)'}{\sum_{s_1}\cdots\sum_{s_\kappa} T} - \frac{\left(\sum_{s_1}\cdots\sum_{s_\kappa}(TD)\right)^2}{\left(\sum_{s_1}\cdots\sum_{s_\kappa} T\right)^2} , \tag{2.14}$$

with

$$\left(\sum_{s_1}\cdots\sum_{s_\kappa}(TD)\right)' = \frac{\partial\left(\sum_{s_1}\cdots\sum_{s_\kappa}(TD)\right)}{\partial\,\sigma^2} = \sum_{s_1}\cdots\sum_{s_\kappa}\left(TD^2+TD'\right) . \tag{2.15}$$

In (2.14) and (2.15), T, D and D' are given by (A.19), (A.22) and (A.23), respectively.

Proof. See Appendix A.2. □

The Fisher information in (2.12), and hence the CRLB in (2.8), can be obtained by evaluating the expected value of the result in (2.13) with respect to $\mathbf{y}$. Due to

the result in (2.14), the expected value cannot be obtained analytically as in the DA model, and other techniques have to be explored. In this work, we will rely on Monte Carlo simulations to obtain the CRLB. Other techniques such as the Gauss-Hermite quadrature are investigated in [14].

2.3.3 CRLB for mixed model with equal noise variance

As we advanced in the previous subsection, we can improve the performance in the estimation of the noise variance by considering not only the received signal from the known pilot sequence, but also the signal from the unknown user data. Let us redefine the signal model in (2.4) to account for this mixed DA and NDA model. If we consider τ known and κ unknown symbols, the pilot sequence $\mathbf{s}$ is now defined as $\mathbf{s} = [s_1 \cdot \cdot \cdot \; s_\tau \; s_{\tau+1} \cdot \cdot \cdot \; s_{\tau+\kappa}]^T$. As a consequence, the new dimension of the received signal $\mathbf{y}$, and hence of $\mathbf{n}$, is $M(\tau+\kappa) \times 1$. Furthermore, $\mathbf{S}$ is now a $M(\tau+\kappa) \times M$ matrix such that $\mathbf{S}^H\mathbf{S} = (\tau+\kappa)\,\mathbf{I}_M$.

In order to obtain the CRLB for the estimation of the noise variance σ^2 we proceed as in the NDA model. The objective is to find the Fisher information in (2.12) to calculate the CRLB as in (2.8).

Theorem 2.3.3 Consider the mixed DA and NDA model with equal noise at the receive antennas. Then, the second derivative of $\ln p(\mathbf{y};\sigma^2)$ with respect to σ^2 is defined as

$$\begin{aligned}\frac{\partial^2 \ln p(\mathbf{y};\sigma^2)}{\partial(\sigma^2)^2} =& \frac{M(\tau+\kappa)}{\sigma^4} - \sum_{m=1}^{M} \frac{\big(\lambda_m(\tau+\kappa)\big)^3}{\sigma^4} + \frac{3\big(\lambda_m(\tau+\kappa)\big)^2}{\sigma^2} + 2\lambda_m(\tau+\kappa) \\ & \qquad \big(\sigma^2+\lambda_m(\tau+\kappa)\big)^3 \\ & - 2\frac{\mathbf{y}^H\mathbf{y}}{\sigma^6} + \frac{\partial^2 \ln\left(\sum_{s_{\tau+1}} \cdot \cdot \cdot \sum_{s_{\tau+\kappa}} T\right)}{\partial(\sigma^2)^2}\,,\end{aligned} \tag{2.16}$$

where

$$\frac{\partial^2 \ln\left(\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} T\right)}{\partial\left(\sigma^2\right)^2} = \frac{\left(\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} (TD)\right)'}{\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} T} - \frac{\left(\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} (TD)\right)^2}{\left(\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} T\right)^2} , \tag{2.17}$$

with

$$\left(\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} (TD)\right)' = \sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} \left(TD^2 + TD'\right) . \tag{2.18}$$

In (2.17) and (2.18), T, D and D' are given by (A.28), (A.29) and (A.30), respectively.

Proof. See Appendix A.3. □

As in the NDA model, the analytical evaluation of the expected value of the second derivative of $\ln p\left(\mathbf{y};\sigma^2\right)$ in (2.16) is not possible. Therefore, we use Monte Carlo simulations to compute it and obtain the CRLB.

2.3.4 CRLB for DA model with different noise variance

In the following subsections we consider that the noise variance is different at the M antennas. Therefore, our goal is the estimation of σ_m^2, the noise variance at the m-th receive antenna. We consider that the M noise variances are uncorrelated and therefore each of them can only be estimated from the τ received samples at its corresponding antenna. This was not the case when the noise variance was the same across all the antennas, and hence the estimation could be calculated from $M\tau$ samples.

This new scenario requires a modified signal model. From the original signal model in (2.1), the received signal at the m-th antenna is

$$\mathbf{y}_m = \sum_{l=1}^{N_{\mathrm{L}}} h_{m,l}\mathbf{s} + \mathbf{n}_m = h_m\mathbf{s} + \mathbf{n}_m , \tag{2.19}$$

where h_m and $\mathbf{n}_m$ are the total channel and noise samples, respectively, at the m-th antenna. The variables $\mathbf{y}_m$, $\mathbf{s}$, and $\mathbf{n}_m$ are $\tau \times 1$ vectors. Furthermore, h_m and $\mathbf{n}_m$ are distributed as $h_m \sim \mathcal{CN}(0, \mathbf{R}_{mm})$ and $\mathbf{n}_m \sim \mathcal{CN}(\mathbf{0}_\tau, \sigma_m^2 \mathbf{I}_\tau)$, respectively, with $\mathbf{R}_{mm}$ being the mm-th element of the covariance matrix $\mathbf{R}$. Finally, as in the previous scenario, h_m, $\mathbf{s}$, and $\mathbf{n}_m$ are considered independent random variables.

In the DA model the pilot sequence $\mathbf{s}$ is a deterministic variable. Therefore, the received signal $\mathbf{y}_m$ is distributed as $\mathbf{y}_m \sim \mathcal{CN}\left(\boldsymbol{\mu}_{\mathbf{y}_m}, \mathbf{C}_{\mathbf{y}_m}\right)$ where the mean $\boldsymbol{\mu}_{\mathbf{y}_m}$ and the covariance matrix $\mathbf{C}_{\mathbf{y}_m}$ are expressed as

$$\boldsymbol{\mu}_{\mathbf{y}_m} = \mathrm{E}\left[h_m \mathbf{s} + \mathbf{n}_m\right] = \mathrm{E}\left[h_m\right]\mathbf{s} + \mathrm{E}\left[\mathbf{n}_m\right] = \mathbf{0}_\tau \ , \tag{2.20}$$

$$\mathbf{C}_{\mathbf{y}_m} = \mathrm{E}\left[\left(\mathbf{y}_m - \boldsymbol{\mu}_{\mathbf{y}_m}\right)\left(\mathbf{y}_m - \boldsymbol{\mu}_{\mathbf{y}_m}\right)^H\right] = \mathrm{E}\left[\mathbf{y}_m \mathbf{y}_m^H\right] = \mathbf{R}_{mm}\mathbf{s}\mathbf{s}^H + \sigma_m^2 \mathbf{I}_\tau \ , \tag{2.21}$$

where the expected value is taken with respect to $\mathbf{y}_m$.

Theorem 2.3.4 Consider the DA model with different noise at the receive antennas. Then, the CRLB for the estimation of the noise variance is given by

$$\mathrm{Var}\left[\widehat{\sigma_m^2}\right] \geq \frac{\sigma_m^4}{\tau - 1 + \dfrac{1}{\left(1 + \dfrac{\mathbf{R}_{mm}\tau}{\sigma_m^2}\right)^2}} \ . \tag{2.22}$$

Proof. See Appendix A.4. □

Let us compare the expression for the CRLB in (2.22) with that in (2.10), which correspond to the scenario where the noise variance is the same at all the receive antennas. We observe that the expression in (2.10) becomes (2.22) when considering that the number of antennas is one, that is, $M = 1$, and replacing λ_m by $\mathbf{R}_{mm}$. This result was expected since now we are estimating the noise variance from the samples received at one, instead of M, antennas.

2.3.5 CRLB for NDA model with different noise variance

When the pilot symbols are unknown, the received signal $\mathbf{y}_m$ is no longer a complex Gaussian random variable and we need to rely on the formula in (2.12) to compute the Fisher information of σ_m^2.

Theorem 2.3.5 Consider the NDA model with different noise at the receive antennas. Then, the second derivative of $\ln p\left(\mathbf{y}_m; \sigma_m^2\right)$ with respect to σ_m^2 is defined as

$$\begin{aligned}\frac{\partial^2 \ln p\left(\mathbf{y}_m; \sigma_m^2\right)}{\partial\left(\sigma_m^2\right)^2} = &-\frac{\frac{\left(\mathbf{R}_{mm}\kappa\right)^3}{\sigma_m^4} + \frac{3\left(\mathbf{R}_{mm}\kappa\right)^2}{\sigma_m^2} + 2\mathbf{R}_{mm}\kappa}{\left(\sigma_m^2 + \mathbf{R}_{mm}\kappa\right)^3} \\ &+ \frac{\kappa}{\sigma_m^4} - \frac{2\mathbf{y}_m^H\mathbf{y}_m}{\sigma_m^6} + \frac{\partial^2 \ln\left(\sum_{s_1}\cdots\sum_{s_\kappa} T\right)}{\partial\left(\sigma_m^2\right)^2},\end{aligned} \tag{2.23}$$

where

$$\frac{\partial^2 \ln\left(\sum_{s_1}\cdots\sum_{s_\kappa} T\right)}{\partial\left(\sigma_m^2\right)^2} = \frac{\left(\sum_{s_1}\cdots\sum_{s_\kappa}(TD)\right)'}{\sum_{s_1}\cdots\sum_{s_\kappa} T} - \frac{\left(\sum_{s_1}\cdots\sum_{s_\kappa}(TD)\right)^2}{\left(\sum_{s_1}\cdots\sum_{s_\kappa} T\right)^2}, \tag{2.24}$$

with

$$\left(\sum_{s_1}\cdots\sum_{s_\kappa}(TD)\right)' = \sum_{s_1}\cdots\sum_{s_\kappa}\left(TD^2 + TD'\right). \tag{2.25}$$

In (2.24) and (2.25), T, D and D' are given by (A.39), (A.42) and (A.43), respectively.

Proof. See Appendix A.5. □

The Fisher information, and hence the CRLB on the estimation of σ_m^2, can be obtained by evaluating the expected value of the second derivative in (2.23). An analytical evaluation is again not possible due to the expression in (2.24). As an alternative, we employ Monte Carlo simulations.

2.3.6 CRLB for mixed model with different noise variance

We can improve the estimation of σ_m^2 by combining the DA and NDA pilot models. In order to obtain the CRLB we proceed as in the first noise model, that is, equal noise variance at the receive antennas.

Theorem 2.3.6 Consider the mixed DA and NDA model with different noise at the receive antennas. Then, the second derivative of $\ln p\left(\mathbf{y}_m;\sigma_m^2\right)$ with respect to σ_m^2 is defined as

$$\begin{aligned}\frac{\partial^2 \ln p\left(\mathbf{y}_m;\sigma_m^2\right)}{\partial\left(\sigma_m^2\right)^2} = &-\frac{\frac{\left(\mathbf{R}_{mm}\left(\tau+\kappa\right)\right)^3}{\sigma_m^4}+\frac{3\left(\mathbf{R}_{mm}\left(\tau+\kappa\right)\right)^2}{\sigma_m^2}+2\mathbf{R}_{mm}\left(\tau+\kappa\right)}{\left(\sigma_m^2+\mathbf{R}_{mm}\left(\tau+\kappa\right)\right)^3} \\ &-\frac{2\mathbf{y}_m^H\mathbf{y}_m}{\sigma_m^6}+\frac{\partial^2 \ln\left(\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} T\right)}{\partial\left(\sigma_m^2\right)^2}+\frac{(\tau+\kappa)}{\sigma_m^4}\,,\end{aligned} \tag{2.26}$$

where

$$\frac{\partial^2 \ln\left(\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} T\right)}{\partial\left(\sigma_m^2\right)^2} = \frac{\left(\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} (TD)\right)'}{\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} T} - \frac{\left(\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} (TD)\right)^2}{\left(\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} T\right)^2}\,, \tag{2.27}$$

with

$$\left(\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} (TD)\right)' = \sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} \left(TD^2+TD'\right)\,. \tag{2.28}$$

In (2.27) and (2.28), T, D and D' are given by (A.45), (A.46) and (A.47), respectively.

Proof. See Appendix A.6. □

Monte Carlo simulations are used to obtain the expected value of (2.26). This value is then used to evaluate the Fisher information, and hence the CRLB on the estimation of σ_m^2.

2.4 Noise Variance Estimators

In the previous section we derived the CRLB for the estimation of the noise variance under different pilot and noise models. The objective in the section is to find the corresponding estimator.

The different strategies available to obtain an estimator for a given parameter are summarized in [15, ch. 14]. Among those strategies, the ones that are suitable for our signal models in (2.4) and (2.19) are: CRLB evaluation, Rao-Blackwell-Lehmann-Scheffe (RBLS) theorem, ML estimator, and MM estimator. The first approach results in the minimum variance unbiased estimator (MVUE), that is, in an estimator that offers the lowest error, for any value of the noise variance, among all unbiased estimators. Furthermore, it is an *efficient* estimator since it attains the CRLB. However, this strategy cannot be pursued in our context, for any of the pilot and noise models, since the PDF of the received signal does not satisfy the following expression:

$$\frac{\partial \ln p\left(\mathbf{y};\sigma^2\right)}{\partial\, \sigma^2} = I\left(\sigma^2\right)\left(g\left(\mathbf{y}\right) - \sigma^2\right) \ , \tag{2.29}$$

where $g\left(\mathbf{y}\right)$ would be the efficient estimator.

The second approach, the RBLS theorem, results in the MVUE as well. Nevertheless, we cannot employ this strategy either in this work since the PDF of the received signal cannot be expressed as

$$p\left(\mathbf{y};\sigma^2\right) = g\left(T\left(\mathbf{y}\right),\sigma^2\right) h\left(\mathbf{y}\right) \ , \tag{2.30}$$

where g is a function only of $T\left(\mathbf{y}\right)$ and σ^2, and h a function that depends only on the received samples $\mathbf{y}$. On the other hand, $T\left(\mathbf{y}\right)$ is a sufficient statistic.

Hence, since the first two strategies cannot be used, we investigate the ML and MM estimators. The ML estimator is efficient for large data records and certain conditions on the PDF. Our simulation results show that the number of samples

needed for the estimator to be efficient is actually very low. On the other hand, the MM estimator is, in general, not optimal. However, our results indicate that it is efficient for very low SNR values.

2.4.1 Estimators for DA model with equal noise variance

Let us recall the expression for the received signal when the M antennas have the same noise variance:

$$\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{n} \ . \tag{2.31}$$

As a consequence, the received signal is distributed as $\mathbf{y} \sim \mathcal{CN}\left(\boldsymbol{\mu}_{\mathbf{y}}, \mathbf{C}_{\mathbf{y}}\right)$, where the mean $\boldsymbol{\mu}_{\mathbf{y}}$ and covariance matrix $\mathbf{C}_{\mathbf{y}}$ are given by (2.6) and (2.7), respectively. Besides, the inverse and determinant of $\mathbf{C}_{\mathbf{y}}$ are as in (A.6) and (A.16), respectively (in (A.16), κ is replaced by τ). Therefore, we have

$$p\left(\mathbf{y}; \sigma^2\right) = \frac{\exp\left[-\mathbf{y}^H \left(\frac{\mathbf{I}_{M\tau}}{\sigma^2} - \frac{\mathbf{S}\mathbf{S}^H}{\sigma^2\tau} + \frac{\mathbf{S}\mathbf{W}\mathbf{B}\mathbf{W}^H\mathbf{S}^H}{\tau}\right)\mathbf{y}\right]}{\pi^{M\tau}\left(\sigma^2\right)^{M\tau} \prod_{m=1}^{M}\left(1 + \frac{\lambda_m \tau}{\sigma^2}\right)} \ . \tag{2.32}$$

ML estimator

The ML estimate of the noise variance is given by the value of σ^2 that maximizes the PDF in (2.32). Since maximizing $p\left(\mathbf{y}; \sigma^2\right)$ is the same as maximizing the natural logarithm of $p\left(\mathbf{y}; \sigma^2\right)$, we have

$$\widehat{\sigma^2} = \arg\max_{\sigma^2} \ln p\left(\mathbf{y}; \sigma^2\right) \ , \tag{2.33}$$

where

$$\begin{aligned} \ln p\left(\mathbf{y}; \sigma^2\right) = &-\frac{\mathbf{y}^H\mathbf{y}}{\sigma^2} + \frac{\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}}{\sigma^2\tau} - \sum_{m=1}^{M} \frac{|\mathbf{u}_m|^2}{\tau\left(\sigma^2 + \tau\lambda_m\right)} \\ &- M\tau \ln\left(\pi\right) - M\tau \ln\left(\sigma^2\right) - \sum_{m=1}^{M} \ln\left(1 + \frac{\lambda_m\tau}{\sigma^2}\right) \ . \end{aligned} \tag{2.34}$$

The expression above cannot be maximized analytically and therefore a closed form expression for $\widehat{\sigma^2}$ does not exist. Hence, we need to rely on numerical techniques to obtain the estimate. In this work we study the Newton-Raphson method, which iteratively estimates the noise variance as[2]

$$\widehat{\sigma^2}_{k+1} = \widehat{\sigma^2}_k - \left. \frac{\dfrac{\partial \ln p\left(\mathbf{y};\sigma^2\right)}{\partial \sigma^2}}{\dfrac{\partial^2 \ln p\left(\mathbf{y};\sigma^2\right)}{\partial \left(\sigma^2\right)^2}} \right|_{\sigma^2=\widehat{\sigma^2}_k} , \tag{2.35}$$

where the first and second derivatives are given by

$$\frac{\partial \ln p\left(\mathbf{y};\sigma^2\right)}{\partial \sigma^2} = -\frac{M\tau}{\sigma^2} + \sum_{m=1}^{M} \frac{\left(\lambda_m\tau\right)^2}{\sigma^2} + \lambda_m\tau + \frac{|\mathbf{u}_m|^2}{\tau} + \frac{\mathbf{y}^H\mathbf{y}}{\sigma^4} - \frac{\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}}{\tau\sigma^4} , \tag{2.36}$$

$$\begin{aligned} \frac{\partial^2 \ln p\left(\mathbf{y};\sigma^2\right)}{\partial \left(\sigma^2\right)^2} &= \frac{M\tau}{\sigma^4} - \frac{2\mathbf{y}^H\mathbf{y}}{\sigma^6} + \frac{2\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}}{\tau\sigma^6} \\ &\quad - \sum_{m=1}^{M} \frac{\left(\lambda_m\tau\right)^3}{\sigma^4} + \frac{3\left(\lambda_m\tau\right)^2}{\sigma^2} + 2\lambda_m\tau + \frac{2|\mathbf{u}_m|^2}{\tau} \\ &\quad \frac{}{\left(\sigma^2+\lambda_m\tau\right)^3} . \end{aligned} \tag{2.37}$$

In general, the convergence in the Newton-Raphson method is not guaranteed. However, our simulation results show that for the DA model with equal noise variance this method converges and always output an estimate for the noise variance.

MM estimator

The ML technique requires an iterative process to estimate the noise variance. Although it results in an efficient estimator, there may be applications where its time consuming characteristic makes it unfeasible. This motivates the development of the MM estimator, which will offer a closed form expression, and hence the reduction of the computation time. The drawback of the MM estimator is that it is only efficient for very low SNR environments.

[2]Our contributions of this work claimed that we were proposing non-iterative ML estimators. However, in that context, an iterative approach referred to the jointly iterative estimation of the channel and noise variance.

Theorem 2.4.1 Consider the DA model with equal noise at the receive antennas. Then, the MM estimator of the noise variance is given by

$$\widehat{\sigma^2} = \frac{\mathbf{y}^H \mathbf{y}}{M\tau} - \frac{\mathrm{tr}\,[\mathbf{R}]}{M} \ . \tag{2.38}$$

Proof. See Appendix A.7. □

In (2.38), the first summand computes the mean total power per received sample, while the second summand accounts for the mean channel power. Hence, the noise power is estimated by subtracting the mean channel power from the mean total power. This agrees with the signal model in (2.31) and recalling that the pilot symbols have unit power.

A similar noise variance estimator was derived in [13, eq. (15)] by solving a convex optimization problem. In that work, the final estimate is given by $\max(0, \widehat{\sigma^2})$, where $\widehat{\sigma^2}$ is as in (2.38), and, therefore, it may result in $\widehat{\sigma^2} = 0$. However, a noise variance estimate of zero would require the computation of the inverse of the covariance matrix $\mathbf{R}$ in (2.3) during the posterior channel estimation stage. From [9], this operation may be challenged when the number of receive antennas is large due to rank deficiency. As a consequence, the noise variance estimator in [13] is not suitable for massive MIMO scenarios.

2.4.2 Estimators for NDA model with equal noise variance

In the NDA model the received signal is given by (2.31), the same expression that we used for the DA model. However, since the pilot sequence is now unknown, the PDF of $\mathbf{y}$ is expressed as in (A.17). Let us recall that for the NDA model, we are replacing τ by κ to denote the length of the pilot sequence, that is, the number of unknown symbols.

ML estimator

As in the DA model, we cannot obtain an analytical expression for the ML estimate and we have to explore numerical techniques. Our simulation results showed that the Newton-Raphson method used for the DA model fails to converge in the NDA model and therefore other techniques were investigated. As in [14], we will finally implement the expectation-maximization (EM) method.

The idea behind the EM technique is as follows. Let us define $\mathbf{x}$ as the *complete* data set formed by the *incomplete* data $\mathbf{y}$ and the unknown training matrix $\mathbf{S}$. Then, instead of maximizing $\ln p(\mathbf{y};\sigma^2)$, we want to maximize $\ln p(\mathbf{x};\sigma^2)$. Since the data set $\mathbf{x}$ is not available, we instead try to maximize its conditional expected value given the received signal $\mathbf{y}$. This maximization results in an estimate of the noise variance. However, we do not use this result as the final estimate, but instead we employ it to improve the computation of the conditional expected value. This will, hopefully, result in a better estimate of the noise variance. This iterative process is repeated until convergence.

Theorem 2.4.2 Consider the NDA model with equal noise at the receive antennas. Then, the ML estimator of the noise variance is computed by means of the EM method, where E_k, the expression to be maximized during the k-th iteration, is given by

$$\begin{aligned}\mathrm{E}_k = &-\frac{\mathbf{y}^H\mathbf{y}}{\sigma^2} + \frac{\mathbf{y}^H\mathrm{E}\left[\mathbf{S}\mathbf{S}^H\right]\mathbf{y}}{\sigma^2\kappa} - \sum_{m=1}^{M}\frac{\mathrm{E}\left[|\mathbf{u}_m|^2\right]}{\kappa\left(\sigma^2+\kappa\lambda_m\right)} \\ &- M\kappa\ln(\pi) - M\kappa\ln\left(\sigma^2\right) - \sum_{m=1}^{M}\ln\left(1+\frac{\lambda_m\kappa}{\sigma^2}\right) ,\end{aligned} \tag{2.39}$$

with

$$\mathrm{E}[x] = \frac{\sum_{s_1}\cdots\sum_{s_\kappa} x\, p\left(\mathbf{y}|\mathbf{S};\widehat{\sigma^2}_k\right)}{\sum_{s_1}\cdots\sum_{s_\kappa} p\left(\mathbf{y}|\mathbf{S};\widehat{\sigma^2}_k\right)} , \tag{2.40}$$

where $p\left(\mathbf{y}|\mathbf{S};\widehat{\sigma^2}_k\right)$ is as in (2.32) with τ and σ^2 replaced by κ and $\widehat{\sigma^2}_k$, respectively. In order to maximize E_k we use the Newton-Raphson technique with the first and second derivatives given by (2.36) and (2.37), respectively, and replacing τ by κ, and $\mathbf{S}\mathbf{S}^H$ and $|\mathbf{u}_m|^2$ by their corresponding expected values computed with (2.40).

Proof. See Appendix A.8. □

We observe that the ML estimate of the noise variance for the NDA model is computed with a double iterative process: within the k-th iteration of the EM method, $\widehat{\sigma^2}_k$ is obtained by means of the Newton-Raphson technique.

MM estimator

Following the same steps as for the DA model and considering that $\mathbf{S}$, $\mathbf{h}$ and $\mathbf{n}$ are independent random variables, the MM estimator for the NDA model is also given by (2.38), that is

$$\widehat{\sigma^2} = \frac{\mathbf{y}^H\mathbf{y}}{M\kappa} - \frac{\mathrm{tr}\left[\mathbf{R}\right]}{M} \ . \tag{2.41}$$

2.4.3 Estimators for mixed model with equal noise variance

The PDF of the received signal for the mixed DA and NDA model is derived in (A.27). Let us recall that with this model we are estimating the noise variance σ^2 from τ known and κ unknown pilot symbols. In practice, the known symbols correspond to the pilot sequence used to estimate the channel as in (2.3), while the unknown symbols refer to the user data.

ML estimator

The ML estimate of the noise variance for this model is obtained by the same double iterative process described in the NDA model, that is, Newton-Raphson method

within the EM algorithm. Following the same steps described for the NDA model, we obtain that E_k, the function to be maximized, is given by (2.39) with κ replaced by $\tau + \kappa$:

$$\begin{aligned}\mathrm{E}_k = &- \frac{\mathbf{y}^H\mathbf{y}}{\sigma^2} + \frac{\mathbf{y}^H\mathrm{E}\left[\mathbf{S}\mathbf{S}^H\right]\mathbf{y}}{\sigma^2\left(\tau+\kappa\right)} - M\left(\tau+\kappa\right)\ln\left(\pi\right) - \sum_{m=1}^{M}\frac{\mathrm{E}\left[|\mathbf{u}_m|^2\right]}{\left(\tau+\kappa\right)\left(\sigma^2+\left(\tau+\kappa\right)\lambda_m\right)} \\ &- M\left(\tau+\kappa\right)\ln\left(\sigma^2\right) - \sum_{m=1}^{M}\ln\left(1+\frac{\lambda_m\left(\tau+\kappa\right)}{\sigma^2}\right) ,\end{aligned} \tag{2.42}$$

where the expected value is taken with respect to $s_{\tau+1},\ldots,s_{\tau+\kappa}|\mathbf{y}$ evaluated in $\widehat{\sigma^2}_k$. Therefore, the expected values above can be obtained as

$$\mathrm{E}\left[x\right] = \frac{\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} x\, p\left(\mathbf{y}|s_{\tau+1},\ldots,s_{\tau+\kappa};\widehat{\sigma^2}_k\right)}{\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}} p\left(\mathbf{y}|s_{\tau+1},\ldots,s_{\tau+\kappa};\widehat{\sigma^2}_k\right)} , \tag{2.43}$$

where $p\left(\mathbf{y}|s_{\tau+1},\ldots,s_{\tau+\kappa};\widehat{\sigma^2}_k\right)$ is as in (2.32) with τ and σ^2 replaced by $\tau+\kappa$ and $\widehat{\sigma^2}_k$, respectively.

In order to maximize (2.42) and obtain the k-th estimate of the noise variance, we use the Newton-Raphson technique with the first and second derivatives given by (2.36) and (2.37), respectively, and replacing τ by $\tau+\kappa$, and $\mathbf{S}\mathbf{S}^H$ and $|\mathbf{u}_m|^2$ by their corresponding expected values computed with (2.43).

MM estimator

The MM estimator for this model is derived following the same procedure as for the DA model, and considering that $s_{\tau+1},\ldots,s_{\tau+\kappa}$, $\mathbf{h}$ and $\mathbf{n}$ are independent random variables. Hence, we have

$$\widehat{\sigma^2} = \frac{\mathbf{y}^H\mathbf{y}}{M\left(\tau+\kappa\right)} - \frac{\mathrm{tr}\left[\mathbf{R}\right]}{M} . \tag{2.44}$$

2.4.4 Estimators for DA model with different noise variance

When the noise variance is different at the M antennas, we need to perform M independent estimations. Each of these estimations will be obtained from the received signal at the corresponding m-th antenna, which is given by (2.19):

$$\mathbf{y}_m = h_m \mathbf{s} + \mathbf{n}_m \ . \tag{2.45}$$

As we mentioned when deriving the CRLB for this scenario, the received signal is distributed as $\mathbf{y}_m \sim \mathcal{CN}\left(\boldsymbol{\mu}_{\mathbf{y}_m}, \mathbf{C}_{\mathbf{y}_m}\right)$, where the mean $\boldsymbol{\mu}_{\mathbf{y}_m}$ and covariance matrix $\mathbf{C}_{\mathbf{y}_m}$ are given by (2.20) and (2.21), respectively. Furthermore, the determinant and inverse of $\mathbf{C}_{\mathbf{y}_m}$ were obtained in (A.37) and (A.32), respectively (in (A.37), κ is replaced by τ). As a consequence, the PDF of $\mathbf{y}_m$ may be written as

$$p\left(\mathbf{y}_m; \sigma_m^2\right) = \frac{\exp\left[-\mathbf{y}_m^H \left(\dfrac{\mathbf{I}_\tau}{\sigma_m^2} - \dfrac{\mathbf{s}\mathbf{s}^H}{\dfrac{\sigma_m^4}{\mathbf{R}_{mm}} + \sigma_m^2 \tau}\right) \mathbf{y}_m\right]}{\pi^\tau \left(\sigma_m^2\right)^\tau \left(1 + \dfrac{\mathbf{R}_{mm}\tau}{\sigma_m^2}\right)} \ . \tag{2.46}$$

ML estimator

The natural logarithm of the PDF in (2.46) can be expressed as

$$\begin{aligned} \ln p\left(\mathbf{y}_m; \sigma_m^2\right) = &-\tau \ln\left(\pi\right) - \tau \ln\left(\sigma_m^2\right) - \ln\left(1 + \frac{\mathbf{R}_{mm}\tau}{\sigma_m^2}\right) \\ &- \frac{\mathbf{y}_m^H \mathbf{y}_m}{\sigma_m^2} + \frac{\mathbf{y}_m^H \mathbf{s}\mathbf{s}^H \mathbf{y}_m}{\sigma_m^2 \tau} - \frac{\mathbf{y}_m^H \mathbf{s}\mathbf{s}^H \mathbf{y}_m}{\tau\left(\sigma_m^2 + \mathbf{R}_{mm}\tau\right)} \ . \end{aligned} \tag{2.47}$$

The estimate of the noise variance is given by the value of σ_m^2 that maximizes the expression above. Although this expression is simpler than its version for the model with equal noise variance in (2.34), we still cannot find a closed form expression for the ML estimator. Hence, the estimate is computed by means of the Newton-Raphson

method in (2.35), where the first and second derivatives are given by

$$\begin{aligned}\frac{\partial \ln p\left(\mathbf{y}_m;\sigma_m^2\right)}{\partial \sigma_m^2} =& \frac{\mathbf{y}_m^H\mathbf{y}_m}{\sigma_m^4} - \frac{\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\tau\sigma_m^4} + \frac{\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\tau\left(\sigma_m^2+\mathbf{R}_{mm}\tau\right)^2} \\ & - \frac{\tau}{\sigma_m^2} + \frac{\dfrac{\left(\mathbf{R}_{mm}\tau\right)^2}{\sigma_m^2} + \mathbf{R}_{mm}\tau}{\left(\sigma_m^2+\mathbf{R}_{mm}\tau\right)^2} \ , \end{aligned} \tag{2.48}$$

$$\begin{aligned}\frac{\partial^2 \ln p\left(\mathbf{y}_m;\sigma_m^2\right)}{\partial\left(\sigma_m^2\right)^2} =& -\frac{2\mathbf{y}_m^H\mathbf{y}_m}{\sigma_m^6} + \frac{2\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\tau\sigma_m^6} - \frac{2\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\tau\left(\sigma_m^2+\mathbf{R}_{mm}\tau\right)^3} \\ & + \frac{\tau}{\sigma_m^4} - \frac{\dfrac{\left(\mathbf{R}_{mm}\tau\right)^3}{\sigma_m^4} + \dfrac{3\left(\mathbf{R}_{mm}\tau\right)^2}{\sigma_m^2} + 2\mathbf{R}_{mm}\tau}{\left(\sigma_m^2+\mathbf{R}_{mm}\tau\right)^3} \ . \end{aligned} \tag{2.49}$$

Let us compare the expressions of the first and second derivatives for the cases of equal and different noise variance. We observe that the expressions for the model of equal noise variance, given by (2.36) and (2.37), reduce to the expressions obtained above for the case of different noise variance, when $M = 1$ and $\lambda_m = \mathbf{R}_{mm}$. Therefore, as expected, the ML estimator for the model with different noise variance is a special case of the estimator derived for the model with equal noise variance, when considering only one antenna.

MM estimator

For applications where the time consuming characteristic of the ML estimator is not acceptable, we propose a MM estimator, which results in a closed form expression for the noise variance estimate.

Theorem 2.4.3 Consider the DA model with different noise at the receive antennas. Then, the MM estimator of the noise variance is given by

$$\widehat{\sigma_m^2} = \frac{\mathbf{y}_m^H\mathbf{y}_m}{\tau} - \mathbf{R}_{mm} \ . \tag{2.50}$$

Proof. See Appendix A.9. □

As we observed with the ML estimator, the MM estimator above is a special case of the MM estimator derived for the model with equal noise variance in (2.38).

2.4.5 Estimators for NDA model with different noise variance

For the NDA model with different noise variance, the PDF of the received signal at the m-th antenna is given by (A.38).

ML estimator

As in the model with equal noise variance, the ML estimator of σ_m^2 cannot be obtained either in closed form expression or with the Newton-Raphson iterative method. Therefore, we investigate again the EM algorithm.

Theorem 2.4.4 Consider the NDA model with different noise at the receive antennas. Then, the ML estimator of the noise variance at the m-th antenna is computed by means of the EM method, where E_k, the expression to be maximized during the k-th iteration, is given by

$$\begin{aligned}\mathrm{E}_k = &-\kappa \ln(\pi) - \kappa \ln\left(\sigma_m^2\right) - \ln\left(1 + \frac{\mathbf{R}_{mm}\kappa}{\sigma_m^2}\right) \\ &- \frac{\mathbf{y}_m^H \mathbf{y}_m}{\sigma_m^2} + \frac{\mathbf{y}_m^H \mathrm{E}\left[\mathbf{s}\mathbf{s}^H\right] \mathbf{y}_m}{\sigma_m^2 \kappa} - \frac{\mathbf{y}_m^H \mathrm{E}\left[\mathbf{s}\mathbf{s}^H\right] \mathbf{y}_m}{\kappa\left(\sigma_m^2 + \mathbf{R}_{mm}\kappa\right)} ,\end{aligned} \tag{2.51}$$

with

$$\mathrm{E}\left[\mathbf{s}\mathbf{s}^H\right] = \frac{\sum_{s_1} \cdots \sum_{s_\kappa} \mathbf{s}\mathbf{s}^H p\left(\mathbf{y}_m | \mathbf{s}; \widehat{\sigma_{mk}^2}\right)}{\sum_{s_1} \cdots \sum_{s_\kappa} p\left(\mathbf{y}_m | \mathbf{s}; \widehat{\sigma_{mk}^2}\right)} , \tag{2.52}$$

where $p\left(\mathbf{y}_m | \mathbf{s}; \widehat{\sigma_{mk}^2}\right)$ is as in (2.46) with τ and σ_m^2 replaced by κ and $\widehat{\sigma_{mk}^2}$, respectively. The maximization of E_k is carried out by the Newton-Raphson method with the first and second derivatives given by (2.48) and (2.49), respectively, and replacing τ by κ, and $\mathbf{s}\mathbf{s}^H$ by its expected value computed with (2.52).

Proof. See Appendix A.10. □

MM estimator

Considering that $\mathbf{s}$, h_m and $\mathbf{n}_m$ are independent random variables, the MM estimator is given as in (2.50), with τ replaced by κ:

$$\widehat{\sigma_m^2} = \frac{\mathbf{y}_m^H \mathbf{y}_m}{\kappa} - \mathbf{R}_{mm} \ . \tag{2.53}$$

2.4.6 Estimators for mixed model with different noise variance

When the pilot sequence includes τ known and κ unknown symbols, the PDF of the received signal at the m-th antenna is given by (A.44).

ML estimator

The EM algorithm is also employed in this model to obtain the ML estimate of the noise variance. Following the same procedure as in the NDA model, we obtain

$$\begin{aligned} \mathrm{E}_k = &- (\tau+\kappa)\ln(\pi) - (\tau+\kappa)\ln\left(\sigma_m^2\right) - \ln\left(1 + \frac{\mathbf{R}_{mm}(\tau+\kappa)}{\sigma_m^2}\right) - \frac{\mathbf{y}_m^H \mathbf{y}_m}{\sigma_m^2} \\ &+ \frac{\mathbf{y}_m^H \mathrm{E}\left[\mathbf{s}\mathbf{s}^H\right] \mathbf{y}_m}{\sigma_m^2 (\tau+\kappa)} - \frac{\mathbf{y}_m^H \mathrm{E}\left[\mathbf{s}\mathbf{s}^H\right] \mathbf{y}_m}{(\tau+\kappa)\left(\sigma_m^2 + \mathbf{R}_{mm}(\tau+\kappa)\right)} \ , \end{aligned} \tag{2.54}$$

where the expected value is taken with respect to $s_{\tau+1}, \ldots, s_{\tau+\kappa}|\mathbf{y}_m$ evaluated in $\widehat{\sigma_{mk}^2}$. Therefore, we have

$$\mathrm{E}\left[\mathbf{s}\mathbf{s}^H\right] = \frac{\sum_{s_{\tau+1}} \cdots \sum_{s_{\tau+\kappa}} \mathbf{s}\mathbf{s}^H p\left(\mathbf{y}_m | s_{\tau+1}, \ldots, s_{\tau+\kappa}; \widehat{\sigma_{mk}^2}\right)}{\sum_{s_{\tau+1}} \cdots \sum_{s_{\tau+\kappa}} p\left(\mathbf{y}_m | s_{\tau+1}, \ldots, s_{\tau+\kappa}; \widehat{\sigma_{mk}^2}\right)} \ , \tag{2.55}$$

where $p\left(\mathbf{y}_m | s_{\tau+1}, \ldots, s_{\tau+\kappa}; \widehat{\sigma_{mk}^2}\right)$ is as in (2.46) with τ and σ_m^2 replaced by $\tau+\kappa$ and $\widehat{\sigma_{mk}^2}$, respectively.

The estimate of the noise variance during the k-th iteration of the EM algorithm is computed by maximizing (2.54) by means of the Newton-Raphson method. To

this purpose, the first and second derivatives of E_k are given by (2.48) and (2.49), respectively, and replacing τ by $\tau + \kappa$, and $\mathbf{s}\mathbf{s}^H$ by its expected value in (2.55).

MM estimator

Considering that $s_{\tau+1}, \ldots, s_{\tau+\kappa}$, h_m and $\mathbf{n}_m$ are independent random variables, the MM estimator is given as in (2.50), with τ replaced by $\tau + \kappa$:

$$\widehat{\sigma_m^2} = \frac{\mathbf{y}_m^H \mathbf{y}_m}{\tau + \kappa} - \mathbf{R}_{mm} \ . \tag{2.56}$$

2.5 Limitations and Alternative Expressions for the CRLB and ML Estimator

For the NDA and mixed pilot models, the numerical computation of the CRLB and ML estimator is not feasible for certain scenarios. Specifically, for the NDA and equal noise variance model, when the term $\mathbf{y}^H \mathbf{S}\mathbf{S}^H \mathbf{y}$ grows above a limit, the computation of T in (A.19) and the exponential term of $p(\mathbf{y}; \sigma^2)$ in (2.32) is not possible. This occurs because the resulting number exceeds the maximum value allowed by the IEEE Standard 754 for double precision. Although these numbers are intermediate results and will then be divided by a number of similar magnitude, they compromise the computation of the CRLB and ML estimators.

This limitation affects the scenarios where $\mathbf{y}^H \mathbf{S}\mathbf{S}^H \mathbf{y}$ result in a *large* number, which entails situations with large number of receive antennas (massive MIMO), pilot sequence length, and/or SNR. In order to solve this problem we propose an alternative expression for the CRLB and ML estimator, which is valid for the scenarios where this limitation applies. We proceed as in [14][3]. The computation of the CRLB requires

[3]The alternative expression for the ML estimator proposed in [14] for high SNR scenarios may be used in substitution of the EM algorithm. However, in our work (unknown channel), the alternative form is mandatory.

the second derivative in (2.14). For large values of $\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}$, the summatories in s_i can be replaced by the summand corresponding to the maximum $\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}$. As a result, the second derivative in (2.14) simplifies to

$$\frac{\partial^2 \ln\left(\sum_{s_1}\cdot\ \cdot\ \cdot\sum_{s_\kappa} T\right)}{\partial\left(\sigma^2\right)^2} = D'_{max}\ , \tag{2.57}$$

where D'_{max} is computed with the pilot matrix $\mathbf{S}$ that maximizes $\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}$.

A similar approach can be used to derive the alternative expression for the ML estimate. In this case, the expected value in (2.40) can be rewritten as

$$\mathrm{E}\left[x\right] = x_{max}\ , \tag{2.58}$$

where x_{max} is computed with the pilot matrix $\mathbf{S}$ that maximizes $\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}$. In order to compute E_k in (A.61) we need $\mathrm{E}\left[\mathbf{S}\mathbf{S}^H\right]$ and $\mathrm{E}\left[|\mathbf{u}_m|^2\right]$, which, from (2.58), are given by $\mathbf{S}\mathbf{S}^H$ and $|\mathbf{u}_m|^2$, respectively, and using the value of $\mathbf{S}$ that maximizes $\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}$. These expressions do not depend on $\widehat{\sigma^2}_k$ and, therefore, the alternative expression for the ML estimator only necessitates one iteration within the EM algorithm.

The alternative expressions for the other pilot and noise models can be derived accordingly.

2.6 Simulation Results

In this section we evaluate the performance of the proposed noise variance estimators. For this purpose, we use the normalized squared error as performance metric, which is defined as [9]

$$\mathrm{err} = \frac{\sum_{l=1}^{N_\mathrm{L}} \left\| \widehat{\boldsymbol{\theta}}_l - \boldsymbol{\theta}_l \right\|_F^2}{\sum_{l=1}^{N_\mathrm{L}} \left\| \boldsymbol{\theta}_l \right\|_F^2}\ , \tag{2.59}$$

where $\widehat{\boldsymbol{\theta}}_l = \left[\widehat{\sigma^2_{1,l}} \cdot \cdot \cdot \widehat{\sigma^2_{M,l}}\right]$ and $\boldsymbol{\theta}_l = \left[\sigma^2_{1,l} \cdot \cdot \cdot \sigma^2_{M,l}\right]$, with the index l referring to the l-th cell[4]. From (2.59), we can then compute the normalized mean squared error (MSE) through Monte Carlo simulations. Within each Monte Carlo iteration, a new realization of all the involved random variables (i.e., $\mathbf{s}$, $\mathbf{h}$, $\mathbf{n}$, and hence $\mathbf{y}$) is generated.

In order to ease the regeneration of our results, we next describe the simulation scenarios. As depicted in Fig. 2.2, we consider hexagonal cells, and within each of them, the users are distributed uniformly on the cell edge. On the other hand, during the channel estimation stage, the channel between the user in the l-th cell and the target BS is defined as in [9]:

$$\mathbf{h}_l = \frac{1}{\sqrt{N_{\mathrm{P}}}} \sum_{p=1}^{N_{\mathrm{P}}} \mathbf{a}\left(\phi_{p,l}\right) \alpha_{p,l} \ , \tag{2.60}$$

where N_{P} is the number of paths that define the channel, $\mathbf{a}\left(\phi_p\right)$ denotes the steering vector associated to the AOA ϕ_p of the p-th path, and α_p represents the p-th path attenuation. In order to keep the assumption of a flat fading channel, we consider that the N_{P} paths arrive at the BS within the symbol period of the received signal. Since we are considering narrowband signals, this consideration is assumable. The steering vector may be obtained as

$$\mathbf{a}\left(\phi_p\right) = \left[1 \; e^{-j2\pi \frac{D}{\lambda}\cos(\phi_p)} \cdot \cdot \cdot \; e^{-j2\pi \frac{D}{\lambda}(M-1)\cos(\phi_p)}\right]^T \ , \tag{2.61}$$

where D is the antenna spacing at the BS and λ the wavelength. Furthermore, the AOAs are independent Gaussian random variables distributed as $\phi_p \sim \mathcal{N}\left(\bar{\phi}, \sigma^2_{\mathrm{AOA}}\right)$, where $\bar{\phi}$ represents the angular position of the user. Finally, the attenuation of the p-th path α_p is also a random variable, which does not depend on the path index, and is distributed as $\alpha_p \sim \mathcal{CN}\left(0, \beta\right)$, where β is the distance-based path loss given

[4]Note that for the equal noise variance model we have $\sigma^2_{1,l} = \cdot \cdot \cdot = \sigma^2_{M,l} = \sigma^2_l$ and, therefore, $\widehat{\boldsymbol{\theta}}_l = \left[\widehat{\sigma^2_l} \cdot \cdot \cdot \widehat{\sigma^2_l}\right]$.

by

$$\beta = \frac{\epsilon}{d^{\gamma}} \,, \tag{2.62}$$

in which ϵ is a constant computed to meet the SNR requirement, d is the distance between the user and the target BS, and γ represents the path-loss exponent.

For the equal noise variance model, we set the true value of the noise variance to 1, that is, $\sigma^2 = 1$. On the other hand, for the different noise variance model, the noise levels at the M antennas are independent uniform random variables distributed as $\sigma_m^2 \sim \mathcal{U}\,[0.5, 1.5]$. Besides, regarding both the EM and Newton-Raphson methods, convergence is considered to be achieved when the difference between two consecutive outputs is less than 0.001.

For all simulation scenarios we consider binary PSK (BPSK) modulation ($N = 2$). The reason for this is the considerable amount of simulation time required for higher modulation orders due to the multiple summatories in s_i. Furthermore, in order to increase the number of effective Monte Carlo iterations, we consider the same noise variance at all the N_{L} BSs.

In Table 2.1 we summarize the parameters that are common to all the simulation scenarios, along with their corresponding values. For a better understanding of these parameters, we refer the reader to Fig. 2.2. The values of the remaining parameters are shown along the figures' captions.

2.6.1 Equal noise variance model

We start the performance analysis of the proposed estimators for the case of equal noise variance at the receive antennas. In Fig. 2.3 we plot the CRLB and the associated ML estimators for the three pilot models (i.e., DA, NDA, and mixed DA and NDA). The performance is measured in terms of MSE against the SNR. The number of pilot symbols for the DA and NDA models is 5 in both cases, while for

Table 2.1: General Simulation Parameters

Number of cells, N_{L}	2
Cell radius, r_c	1 km
Distance from BS to cell edge, d_c	800 m
Number of users, K	1
Carrier frequency	2 GHz
Number of paths, N_{P}	50
AOA standard deviation, σ_{AOA}	10 degrees
Path-loss exponent, γ	3
Antenna spacing, D	$\lambda/2$

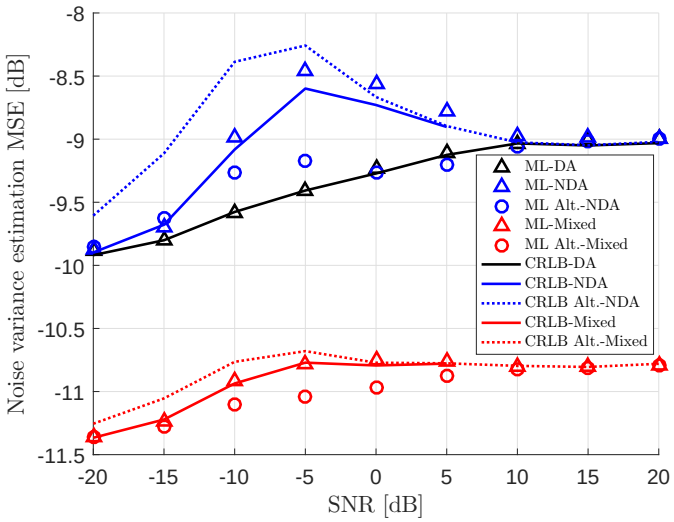


Figure 2.3: Noise variance estimation MSE versus SNR. Specific simulation parameters: $M = 2$, $\tau = 5$ and $\kappa = 5$ (for the mixed model $\kappa = 2$).

the mixed model we consider 7 symbols, from which 5 are known and 2 unknown. Finally, for the NDA and mixed models, we show both the original and alternative forms. Regarding the curves for the CRLB, we observe that the original expressions for the NDA and mixed models do not offer any result for SNR values higher than 5 dB due to the limitation mentioned in Section 2.5. For those values we have to consider the alternative form which seems to fit the original expression's trend accurately. On the other hand, as expected, we notice that for low SNR values the alternative forms

do not approximate the original expressions correctly. If we compare the CRLB for the three pilot models, we confirm that the mixed model offers the best performance due to the extra 2 unknown symbols in comparison to the DA model. On the other hand, as anticipated, for the same number of symbols, the DA model outperforms the NDA model. However, for very low and high SNR values, this difference banishes, which implies that, for those SNR values, the knowledge of the pilot symbols does not improve the noise variance estimation. Finally, we notice that, for the three models, the CRLB increases, on average, with the SNR, which implies that the higher the signal power, the more difficult to estimate the noise variance. For the DA model, this conclusion can also be extracted, analytically, from (2.10).

As far as the ML estimators are concerned, several conclusions may be extracted:

- For the DA model, the ML estimator is efficient for all SNR values.
- For the NDA model, the ML estimator in the original form is efficient for very low and high SNR values. On the other hand, we observe that the ML estimator for the mixed model is efficient for all the SNR range, as in the DA model. However, we expect that it loses efficiency as the number of unknown symbols increases, approaching the performance observed in the NDA model. Finally, we notice that the estimator outputs results for all the SNR values, unlike the CRLB expression, which implies that the latter is more susceptible to the limitation in Section 2.5.
- We observe that the alternative expression for the ML estimator, in both the NDA and mixed models, has the same performance as the original form for very low and high SNR values. For the rest of the range, we observe that the alternative ML estimator outperforms the CRLB. This, a priori contradiction, is justified when commenting on Fig. 2.5.

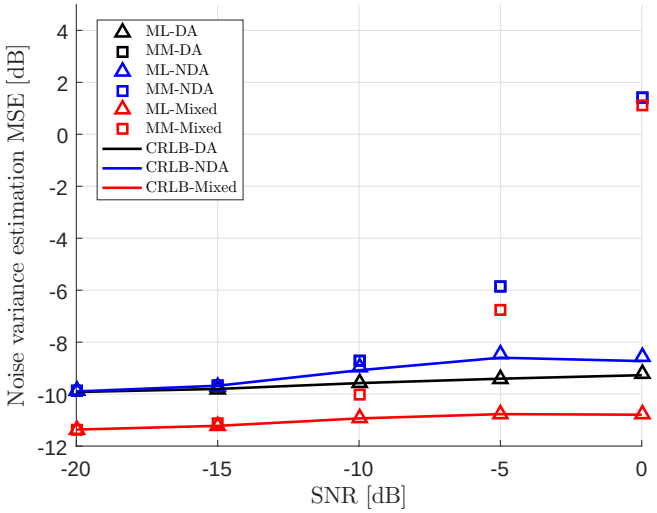


Figure 2.4: Noise variance estimation MSE versus SNR (low SNR range). Specific simulation parameters: $M = 2$, $\tau = 5$ and $\kappa = 5$ (for the mixed model $\kappa = 2$). Note that the performance of the MM estimators for the DA and NDA models is the same. This occurs because those estimators do not depend on the pilot symbols, but only on the number of available samples, which is the same in both cases.

Fig. 2.4 studies the performance of the MM estimators. We use the same scenario as in Fig. 2.3 but the investigation is limited to very low to medium SNR values. We observe that, for the three pilot models, the MM estimators behave efficiently and the same as the ML estimators for very low SNR values, while their performance is compromised as the SNR increases. The equivalent performance of the MM and ML estimators comes from the fact that both estimators simplify to the energy detector for low SNR values. For example, for the DA model, this can be proved by considering that $\sigma^2 \gg \lambda_m$ and $\sigma^2 \gg \mathrm{tr}\left[\mathbf{R}\right]/M$ in (2.36) and (A.54), respectively, for the ML and MM estimators. In both cases the noise variance estimator is given by

$$\widehat{\sigma^2} = \frac{\mathbf{y}^H \mathbf{y}}{M\tau} \ , \tag{2.63}$$

which is the energy detector.

In Fig. 2.5 we show the performance of the ML estimators for different numbers of antennas at the BS, along with their corresponding CRLB. Regarding the CRLB

curves, we first observe that, for the three pilot models, the estimation accuracy increases with M due to the greater number of available samples to estimate the noise variance. On other hand, we also notice that the gap between the curves for the DA and NDA models decreases with M. This indicates that knowing the pilot symbols does not benefit the noise variance estimation performance for large numbers of receive antennas. This was observed as well in Fig. 2.3, but in that case for very low and high SNR values. Finally, Fig. 2.5 shows that the alternative expression for the CRLB approximates the original form accurately for high M values. The explanation for this is that a higher M contributes to increase the term $\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}$ discussed in Section 2.5.

As far as the ML estimator in the original form is concerned, we observe that for the NDA model it gains efficiency as M increases. On the other hand, we notice that the alternative ML estimators improve the performance of the original forms, even further than the CRLB. The reason for this is that the CRLB sets a lower bound for any *unbiased* estimator, and the ML estimators in the alternative forms turned out to be *biased*, which means that, on average, the estimators do not attain the true value. In Fig. 2.6 we justify this explanation by showing the bias (average deviation from the true value in %, μ). Indeed, we observe that the greater the bias, the larger the gap between the alternative ML estimator and the CRLB.

The performance of the proposed ML estimators in massive MIMO scenarios is studied in Fig. 2.7, which shows the noise variance estimation MSE for different large numbers of receive antennas. Since, from Fig. 2.5, the estimation performance is the same in the DA and NDA pilot models for a large M, we only show the results for the DA model. From the figure, we observe that the original form of the ML estimator for the mixed model does not work for large antenna arrays due to the limitation in Section 2.5. Hence, for massive MIMO scenarios, the alternative expression is needed.

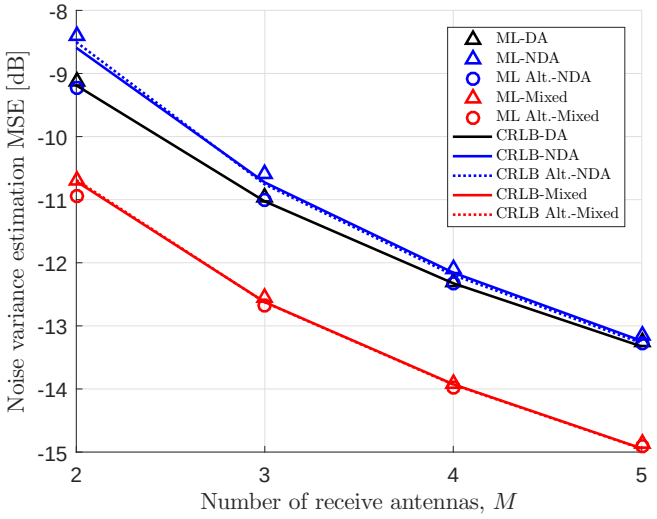


Figure 2.5: Noise variance estimation MSE versus number of receive antennas. Specific simulation parameters: SNR $= 0$ dB, $\tau = 5$ and $\kappa = 5$ (for the mixed model $\kappa = 2$).

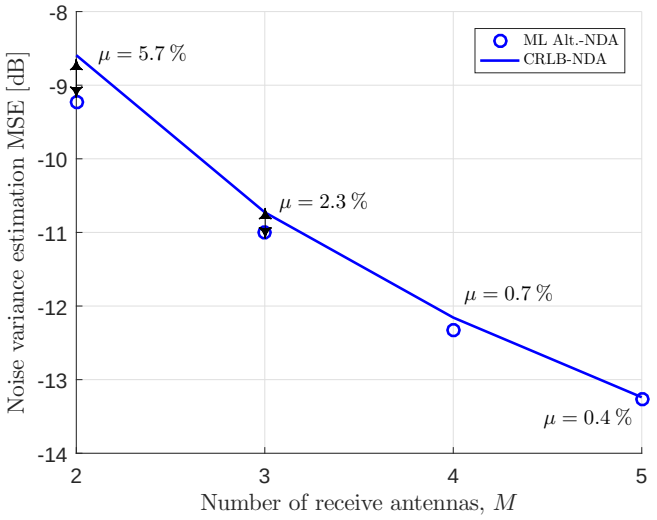


Figure 2.6: Biasness of the alternative ML estimator. The biasness of the alternative ML estimator results in a lower MSE than the CRLB. However, on average, this estimator does not attain the true value of the noise variance. We employ the same scenario as in Fig. 2.5.

For comparison, we also show the performance of the estimators proposed in [11] and [12, eq. (10)]. The LS estimator in [12] depends on $\mathbf{C}_{\mathbf{y}}$, the sample covariance matrix, which is assumed unknown in our work. Hence, for a fair comparison, we substitute it

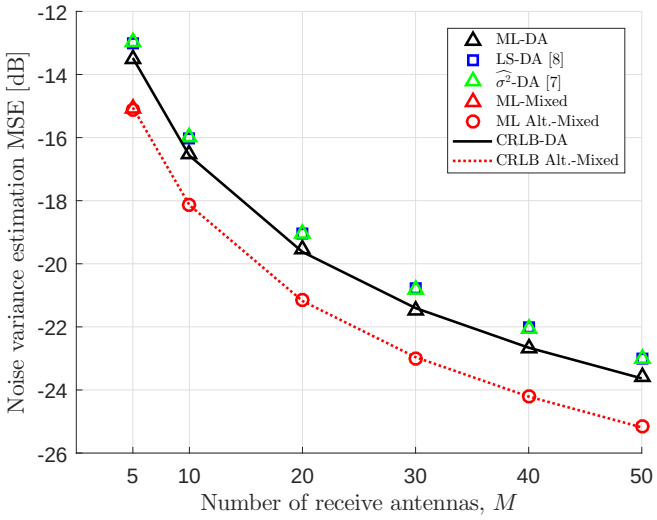


Figure 2.7: Noise variance estimation MSE in massive MIMO scenarios. Specific simulation parameters: SNR $= 0$ dB, $\tau = 5$ and $\kappa = 2$. For massive MIMO scenarios, the alternative expression of the ML estimator is needed due to the computational limitations of the original form.

by its instant value, that is, $\widetilde{\mathbf{C}_{\mathbf{y}}} = \mathbf{y}\mathbf{y}^H$. We observe that our ML estimator offers the best estimation performance, achieving efficiency for both the DA and mixed models.

Finally, in Fig. 2.8 we study how the noise variance estimation impacts the channel estimation performance. For this purpose, we generate several iterations of the channel in (2.60) and estimate it with the MMSE estimator in (2.3), with the noise covariance matrix $\mathbf{\Sigma}$ substituted by $\widehat{\mathbf{\Sigma}} = \text{diag}\left(\left[\widehat{\sigma^2} \cdots \widehat{\sigma^2}\right]\right)$. Then, the normalized channel squared error is computed as

$$\text{err} = 10\log_{10}\left(\frac{\sum_{l=1}^{N_{\text{L}}}\left\|\widehat{\mathbf{h}}_l - \mathbf{h}_l\right\|_F^2}{\sum_{l=1}^{N_{\text{L}}}\|\mathbf{h}_l\|_F^2}\right), \tag{2.64}$$

from which the corresponding MSE is obtained. Fig. 2.8 shows the channel estimation MSE versus the SNR for three different assumptions on the noise covariance matrix $\mathbf{\Sigma}$: the noise power is perfectly known, and it is estimated with the ML and MM

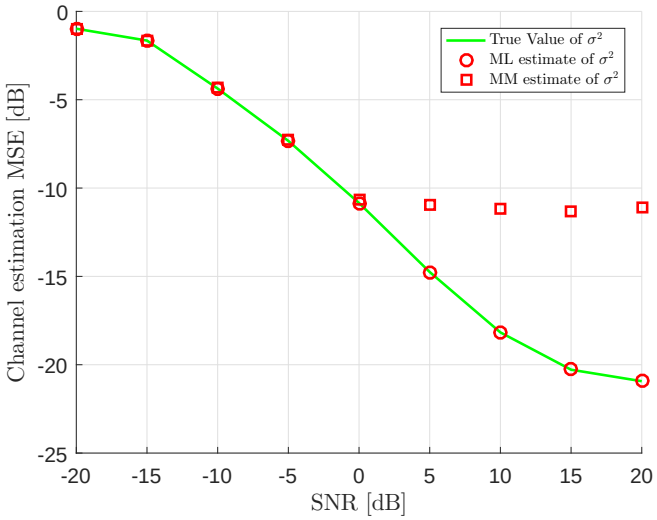


Figure 2.8: Channel estimation MSE versus SNR. Specific simulation parameters: $M = 20$, $\tau = 5$ and $\kappa = 2$.

estimators (mixed pilot model). Since we are considering a massive MIMO scenario ($M = 20$), we employ the alternative expression for the ML estimator.

We observe that the performance of the proposed noise variance ML estimator helps to accomplish a channel estimation accuracy very similar to that of the scenario where the noise power is perfectly known. On the other hand, we notice that the MM estimator performs well for very low SNR values, while compromises the accuracy of the channel estimator as the SNR increases. This was expected from the results observed in Fig. 2.4.

2.6.2 Different noise variance model

When the noise variance is different at the receive antennas, the estimation can only be performed from the samples received at the corresponding antenna. The result of this is that the estimation error does not decrease with the number of antennas, as we observe in Fig 2.9. The behavior of the CRLB and noise variance estimators for

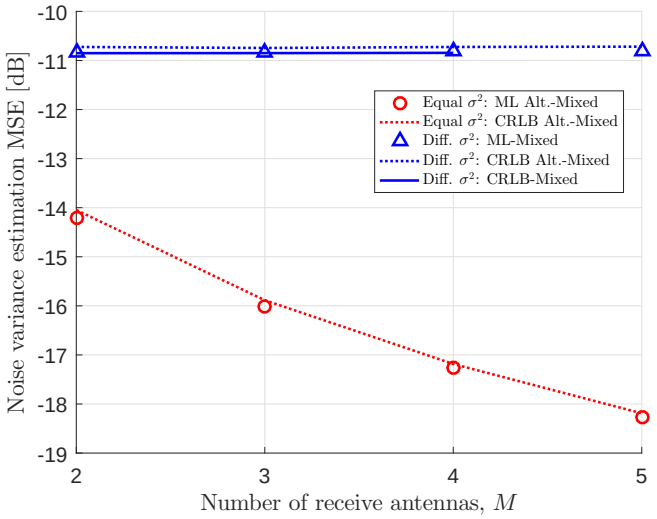


Figure 2.9: Noise variance estimation MSE with different noise at the receive antennas. When the noise variance is different at the receive antennas, the estimation performance does not improve with larger antenna arrays. Specific simulation parameters: SNR $= 0$ dB, $\tau = 10$ and $\kappa = 4$.

different SNR values is the same than that for the equal noise variance model, and hence the reader is referred to Figs. 2.3 and 2.4.

2.7 Discussions

The noise power at the receiver is due to several factors such as thermal noise, receiver nonlinearity, imperfect front end filters, quantization noise, etc [16]. While the contribution of some of these factors may not change over time, there are others that make the noise variance a non-stationary variable. In order to obtain this parameter, an alternative strategy to statistical estimation, like the one here proposed, is to directly measure the noise power through a calibration process. The advantage of our proposal over the calibration approach is that it is robust against the non-stationarity of the noise power, since its estimate is updated continuously.

The proposed expressions for the CRLB as well as the noise variance estimators depend on the channel covariance matrix $\mathbf{R}$, which is supposed known. However, in practice, it has to be estimated, and will be used not only during the noise variance estimation stage, but also to compute the channel estimate as in (2.3). Since $\mathbf{R} = \mathrm{E}\left[\mathbf{h}\mathbf{h}^H\right]$, the ideal channel covariance matrix estimator would be given by

$$\widehat{\mathbf{R}} = \frac{1}{\mathrm{N}_s} \sum_{i=1}^{\mathrm{N}_s} \mathbf{h}_i \mathbf{h}_i^H \ , \tag{2.65}$$

where N_s denotes the number of channel measurements employed to estimate $\mathbf{R}$. Furthermore, the channel can be measured either in time [17], [18] or frequency [19]. However, this strategy cannot be used in our work since the channel is unknown.

This problem can be solved by substituting $\mathbf{h}_i$ in (2.65) by a "preliminary" channel estimate. For this purpose, we could use the MMSE channel estimator in (2.3), but it depends on $\mathbf{R}$, and therefore we have to rely on less powerful methods such as a the LS estimator [11]. As proposed in [9], we may imagine future networks where the channel covariance matrix is learned through a specific training stage.

2.8 Conclusions

In this chapter we have developed ML and MM estimators for the noise variance at the receive antennas of the BS. These estimates can then be used as an input to the MMSE channel estimator. Furthermore, we derive the corresponding CRLB.

Our simulations results show that the ML noise variance estimator is optimal for massive MIMO scenarios since it gains efficiency as the number of receive antennas increases. On the other hand, the MM estimator is only efficient for very low SNR values. However, it is given as a closed form expression, which makes it suitable for applications where the computational requirement of the ML estimators are prohibitive.

CHAPTER 3

CHANNEL ESTIMATION PERFORMANCE IN MMWAVE SYSTEMS

3.1 Introduction

In Chapter 2 we introduced the pilot contamination impairment in cellular networks. In order to reduce its effects on the communications performance, an MMSE channel estimator was proposed in [9], which considers the noise variance at the receive antennas as a known parameter. However, in practical scenarios, this parameter has to be estimated. Our contribution was the derivation of ML and MM noise variance estimators, which can then be integrated to the channel estimator in [9].

In [9], the authors demonstrate that the pilot contamination effect is completely eliminated when considering and infinite number of antennas at the BS, and the cluster of the desired user does not overlap with those of the interfering users in the angular domain. From this result, the authors propose a pilot assignment strategy by which the pilots are assigned to the users in such manner that the non-overlapping condition is fulfilled.

In this chapter we study how the channel estimation performance is affected by the pilot contamination impairment in future 5G mmWave cellular networks. Our results indicate that, compared to current low frequency-based deployments, mmWave cellular networks offer a better channel estimation performance due to the characteristics of its propagation model.

3.2 System and Propagation Models

We use the same system model described in Section 2.2, that is, a cellular system with N_{L} cells, each of them managed by a BS equipped with M antennas. Furthermore, K single-antenna users populate each cell.

We are interested in the performance of the channel estimator in (2.3) under mmWave scenarios. To this end, we use the downlink propagation model for mmWave bands recently presented in [20]. In order to adapt the cited model to the channel estimation stage described in Section 2.2, we assume that the uplink-downlink reciprocity property can be applied, which is a common practice in current cellular deployments [21].

The nomenclature employed in [20] differs from that we used in Chapter 2. In the mmWave model in [20], a channel is characterized by N_{S} *spatial lobes*, which correspond to the multipath clusters in Chapter 2, and N_{C} *time clusters*. As a consequence, the N_{P} paths that define the channel as in (2.60) are grouped in both the time domain (time clusters) as well as in the angular domain (spatial lobes). Furthermore, this model allows more than one time cluster arriving within the same spatial lobe. Once the differences between the two models have been clarified, we next summarize how to obtain the necessary parameters to generate the channels as in (2.60).

1. Mean AOA of the spatial lobes: The mean angle of each arriving spatial lobe at the BS is uniformly distributed over $[0, 2\pi]$ in such a manner that they do not overlap each other. The mean angles are denoted by $\bar{\phi}_i$, for $i = 1, \ldots, N_{\mathrm{S}}$.

2. Number of paths in each time cluster and their AOA: There are N_{T} paths within each time cluster. The AOA of each path is randomly distributed as $\mathcal{N}(\bar{\phi}_i, \Delta^2)$, where $i \sim \mathcal{DU}(1, N_{\mathrm{S}})$. Then, the $N_{\mathrm{P}} = N_{\mathrm{C}} N_{\mathrm{T}}$ total paths are uniformly distributed among the arriving spatial lobes.

3. Path Loss: The mean path loss for a distance d is $PL[\text{dB}] = 20\log_{10}(4\pi/\lambda) + 10\gamma\log_{10}(4\pi d/\lambda)$. Hence, the distance-based path loss is given by[1]

$$\beta = \frac{\epsilon}{10^{PL[\text{dB}]/10}} \ , \tag{3.1}$$

in which ϵ is a constant computed to meet the uplink SNR requirement.

To reproduce the low frequency scenario we use the 3GPP model in [21] (microcell environment), which is also fully described by the previous parameters. This propagation model considers a sector antenna at the BS and therefore the mean AOA of the spatial lobes is uniformly distributed, with respect to the line of sight (LOS) direction, over $\left[-40^{\circ}, 40^{\circ}\right]$. Since we consider isotropic antennas in our work, the mean angles are finally distributed as uniform random variables over $[0, 2\pi]$. Therefore, as far as this parameter is concerned, the only difference between these propagation models is that in the low frequency scenario the spatial lobes can overlap each other. Furthermore, only one time cluster is allowed per spatial lobe ($N_{\text{C}} = N_{\text{S}}$).

To accurately represent complex cellular scenarios, the K users per cell are uniformly distributed over $[0, 2\pi]$ around their corresponding BS. However, note that the mean AOA of the spatial lobes does not depend on the position of the user in mmWave deployments.

3.3 Channel Estimation in mmWave Scenarios

Here, we compare the performance of the pilot assignment strategy proposed in [9] at low frequencies (1.9 GHz) and mmWave bands (28 GHz and 73 GHz), using the channel models discussed in Section 3.2. Specific simulation parameters for both

[1]The actual distance-based path loss would be $PL[\text{dB}]$, however, following others works as [9], we use the expression in (3.1), which includes the transmit power needed to achieve a given uplink SNR.

Table 3.1: Simulation Parameters

	Frequency Scenario		
Parameter	**1.9 GHz**	**28 GHz**	**73 GHz**
N_{S}	6	2	2
$N_{\mathrm{P}} = N_{\mathrm{C}} N_{\mathrm{T}}$	120	64	64
$\Delta[^{\circ}]$	5	9	7
γ	–	3.4	3.3

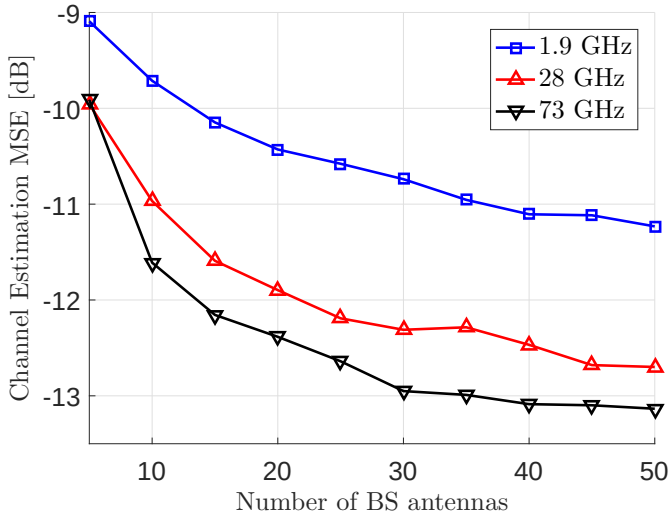
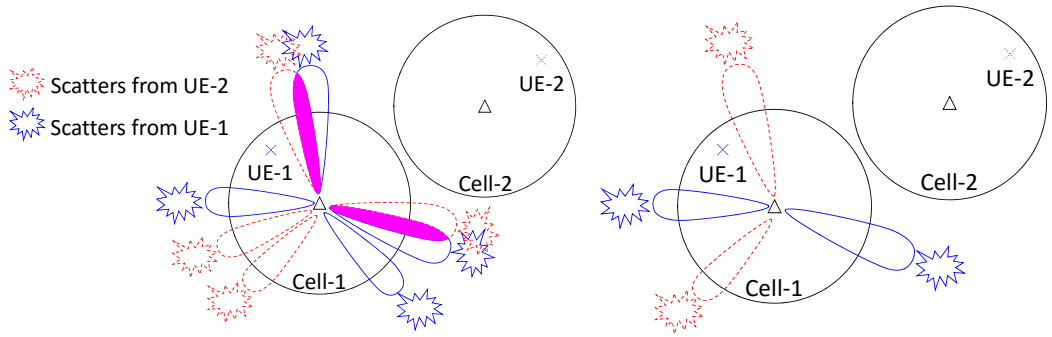


Figure 3.1: Channel estimation MSE in mmWave systems. Simulation parameters: 7 cells, SNR = 20 dB, $D = \lambda/2$ and $K = 10$.

models are summarized in Table 3.1. While N_{S}, N_{C} and N_{T} are fixed in the 3GPP model, in the mmWave model those values are random variables with specific means. In our work, we fix the value of each of those parameters to the nearest integer to its mean so that the same simulation procedure can be used in both cases.

Fig. 3.1 shows the channel estimation error for the three frequency bands in a 50 m radius cell network where all the users are located at a distance of 40 m from the BS. On the one hand, we can observe a better estimation accuracy for the mmWave scenario. This is due to the lower number of arriving spatial lobes at the BS which helps to achieve the non-overlapping condition. This behavior is illustrated



(a) Low frequency scenario. (b) MmWave scenario.

Figure 3.2: Comparison of propagation models in mmWave and current systems. The lower number of arriving spatial lobes at mmWave bands helps to accomplish the non-overlapping condition, and hence a lower pilot contamination. For simplification only 4 spatial lobes are considered in (a).

in Fig. 3.2. On the other hand, we notice that at 73 GHz the performance is better than at 28 GHz. This is explained by the greater standard deviation in the 28 GHz scenario, which compromises the non-overlapping condition and results in larger pilot contamination.

3.4 Discussions

Our results in the previous section show that the pilot assignment strategy proposed in [9] performs better in the mmWave bands when considering the same SNR at 1.9 GHz, 28 GHz, and 73 GHz. Due to the greater path loss at higher frequencies, this assumption would imply using more transmission power at the user terminal in the mmWave scenario making the results in Fig. 3.1 unfair. However, as mentioned in [22], the higher path loss attenuation can be compensated by the inherent use of beamforming in mmWave massive MIMO scenarios.

3.5 Conclusions

This chapter compares the performance of the joint channel estimation and pilot assignment strategy proposed in [9], under low frequency and mmWave massive MIMO scenarios for 5G networks. Our results show that the channel estimation error is reduced in the mmWave bands due to the lower number of arriving spatial lobes at the BS compared to the estimation error under low frequency scenarios.

CHAPTER 4

OPTIMAL ANGULAR SPREAD IN MMWAVE SYSTEMS

4.1 Introduction

In this chapter we continue the study of the pilot contamination impairment in mmWave cellular networks. Specifically, we evaluate how the AS of the multipath clusters impacts the capacity. Our results show that there exists an optimal AS that maximizes the ergodic achievable rate, and it will depend on the design parameters of the cellular network. From [20], the AS in mmWave systems depends on the frequency, and therefore, for a given cellular system, we can choose the working frequency that maximizes the capacity. To the best of our knowledge, it has not been reported any similar analysis. The works closest to ours are [23] and [24], where the authors study the impact of the AS on the capacity in single-user MIMO scenarios. Since the pilot contamination is not present when considering only one user, the capacity constantly increases with the AS without reaching a maximum.

4.2 System Model

We use the same system model described in Section 2.2, that is, a cellular system with N_{L} cells, each of them managed by a BS equipped with M antennas. Furthermore, K single-antenna users populate each cell. Let us recall that, from this system model, the MMSE estimate of the channel of the desired user is given by (2.3)

$$\widehat{\mathbf{h}}_1 = \mathbf{R}_1 \left(\sigma^2 \mathbf{I}_{M\tau} + \tau \sum_{l=1}^{N_{\mathrm{L}}} \mathbf{R}_l \right)^{-1} \mathbf{S}^H \mathbf{y} \ , \tag{4.1}$$

where we consider the estimation of the desired channel in cell 1 without loss of generality.

From [7, Theorem 1], the downlink ergodic capacity of user m in cell j is given by

$$C_{jm} = \log_2 (1 + \gamma_{jm}) \ , \tag{4.2}$$

where the signal to interference plus noise ratio (SINR) γ_{jm} can be obtained as

$$\gamma_{jm} = \frac{S_{jm}}{N + I_{jm}^{(1)} + I_{jm}^{(2)}} \ , \tag{4.3}$$

with

$$N = 1/\rho \ , \tag{4.4}$$

$$S_{jm} = \lambda_j \left| \mathrm{E}\left[\mathbf{h}_{jjm}^H \mathbf{w}_{jm} \right] \right|^2 \ , \tag{4.5}$$

$$I_{jm}^{(1)} = \lambda_j \mathrm{Var}\left[\mathbf{h}_{jjm}^H \mathbf{w}_{jm} \right] \ , \tag{4.6}$$

$$I_{jm}^{(2)} = \sum_{(l,k)\neq(j,m)} \lambda_l \mathrm{E}\left[\left| \mathbf{h}_{ljm}^H \mathbf{w}_{lk} \right|^2 \right] \ . \tag{4.7}$$

In (4.4), ρ denotes the downlink signal to noise ratio (SNR), while in (4.5)-(4.7), $\mathbf{h}_{ljm}$ represents the channel between the user m in cell j and the l-th BS, $\mathbf{W}_l = [\mathbf{w}_{l1} \cdot \cdot \cdot \ \mathbf{w}_{lK}]$ is the precoding matrix, and λ_l is defined as

$$\lambda_l = \frac{1}{\mathrm{E}\left[\frac{1}{K} \mathrm{tr}\left[\mathbf{W}_l \mathbf{W}_l^H \right] \right]} \ . \tag{4.8}$$

The choice of the precoder will determine the performance of the network. Two relevant linear precoders are the eigenbeamforming and regularized zero-forcing (RZF). We will consider the latter since it is seen to achieve a superior performance [25]. Then, assuming the RZF precoder, the matrix $\mathbf{W}_l$ is given by

$$\mathbf{W}_l = \left(\widehat{\mathbf{H}}_{ll} \widehat{\mathbf{H}}_{ll}^H + \frac{M}{\rho} \mathbf{I}_M \right)^{-1} \widehat{\mathbf{H}}_{ll} \ , \tag{4.9}$$

where $\widehat{\mathbf{H}}_{ll} = \left[\widehat{\mathbf{h}}_{ll1} \cdot \cdot \cdot \ \widehat{\mathbf{h}}_{llK} \right]$ can be obtained by estimating the K desired channels in all N_{L} cells as in (4.1).

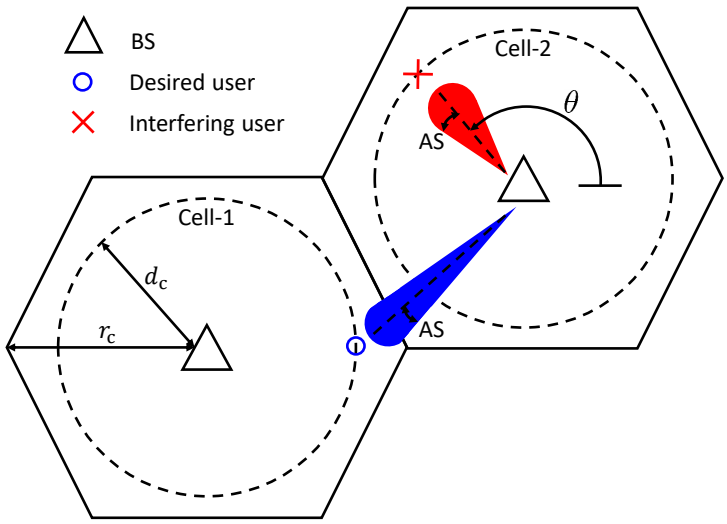


Figure 4.1: Simplified cellular network. The pilot contamination effect grows with the overlapping of the multipath clusters of the users.

4.3 Impact of the Angular Spread on the Capacity

In this section, we present our main contribution: the existence of an optimal AS of the AOAs, which maximizes the downlink capacity in the presence of pilot contamination. In order to ease the justification of this finding, in this section we assume a simplified cellular network as in Fig. 4.1, where the pilot contamination effect can be easily observed. In the next section, we will consider a practical, more complex cellular scenario. We consider 2 cells, each of them populated by one user, which are located at a distance d_c from their corresponding BS. Furthermore, the user in cell 1 (user 1) is positioned at 0° with respect to its BS, while the user in cell 2 (user 2) is located at an angle θ. We are interested in the ergodic capacity of user 1 under pilot contamination, that is, when user 2 is assigned the same time-frequency resource during the channel estimation stage. We assume a single multipath cluster scenario, where the AOA of the multipath components follows a uniform distribution, with the mean AOA $\bar{\phi}$ given by the line of sight of the user. Finally, the distance-based path loss is given by $\beta = \epsilon/d^{\gamma}$, in which ϵ is a constant computed to meet the uplink SNR requirement,

Table 4.1: Common Parameters for the Simplified Cellular Network

Cell radius, r_{c}	50 m
Distance from BS to cell edge, d_{c}	40 m
Carrier frequency	2 GHz
Number of paths, N_{P}	50
Path-loss exponent, γ	3
Antenna spacing, D	$\lambda/2$
Uplink SNR	0 dB
Pilot sequence length, τ	1
Downlink SNR, ρ	0 dB

d is the distance between the user and the target BS, and γ represents the path-loss exponent.

In Table 4.1, we summarize the parameters which are common to all network configurations that we will study in this section, unless otherwise specified. The value of the specific parameters are shown along the figures' captions.

Fig. 4.2 shows the capacity of user 1 as defined in (4.2) (C_{11}). In order to show more clearly how the AS impacts the capacity through the pilot contamination effect, we consider here an infinite uplink SNR ($\sigma^2 = 0$). Therefore, any error in channel estimation will be due to the pilot contamination impairment. Our results show that the capacity reaches a maximum value at AS $= 16^\circ$.

This behavior can be explained from the different components of the expression for the capacity in (4.2), namely S_{11} (signal power), $I_{11}^{(1)}$ and $I_{11}^{(2)}$ (interference powers) and N (noise power), which are given by (4.5), (4.6) and (4.7), respectively. In Fig. 4.2, we also show how these variables are affected by the AS (the noise power N is not depicted since it does not depend on the AS)[1]. We observe that $I_{11}^{(2)}$, the interference from the BS in cell 2 (BS 2), starts increasing at around AS $= 16^\circ$, which

[1]To appreciate the behavior of S_{11}, $I_{11}^{(1)}$ and $I_{11}^{(2)}$ more clearly, we scale S_{11} by a factor of 5.

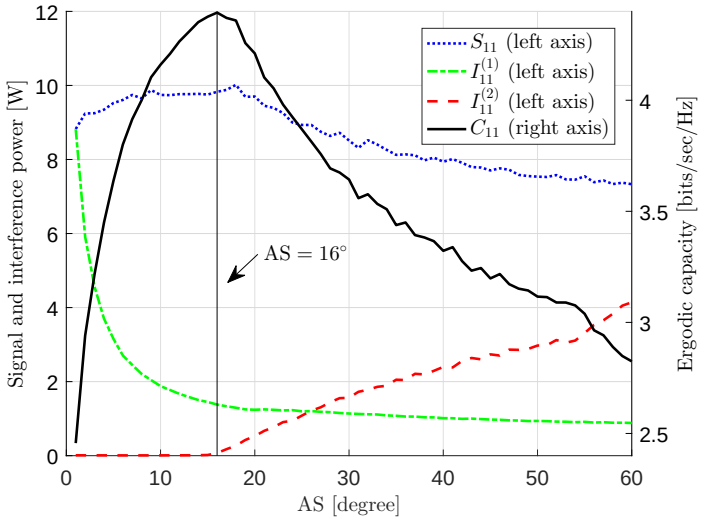


Figure 4.2: Ergodic capacity versus the AS (infinite uplink SNR). The capacity reaches a maximum value due to the behavior of the signal and interference powers with respect to the AS: the capacity first increases due to the raise in signal power and decrease in interference $I_{11}^{(1)}$. Then, it decreases as a consequence of the raise in interference $I_{11}^{(2)}$. For this example, the BS is equipped with $M = 50$ antennas and $\theta = 200°$.

matches the AS that maximizes the capacity. This is explained as follows: during the uplink channel estimation stage, the arriving clusters at BS 2 from both users start overlapping at AS $= 16°$, which causes the BS to estimate the channel from user 2 incorrectly, learning the channel from user 1 as well. As a consequence, BS 2 will transmit undesired data to user 1 in the downlink, which is seen as interference. The overlapping point of the clusters can be obtained geometrically from Fig. 4.1. We observe in Fig. 4.2 that $I_{11}^{(2)}$ actually starts increasing before the overlapping point. This is due to the limited number of antennas (M) considered at the BSs. As derived in [9], the pilot contamination completely vanishes only as $M \to \infty$.

On the other hand, we observe that the signal power S_{11} starts decreasing at around AS $= 20°$. In this case, this is due to the channel estimation error in the BS of cell 1 (BS 1). Now, the overlapping of the clusters causes this BS to learn not only the channel from user 1, the desired channel, but also that from user 2. As

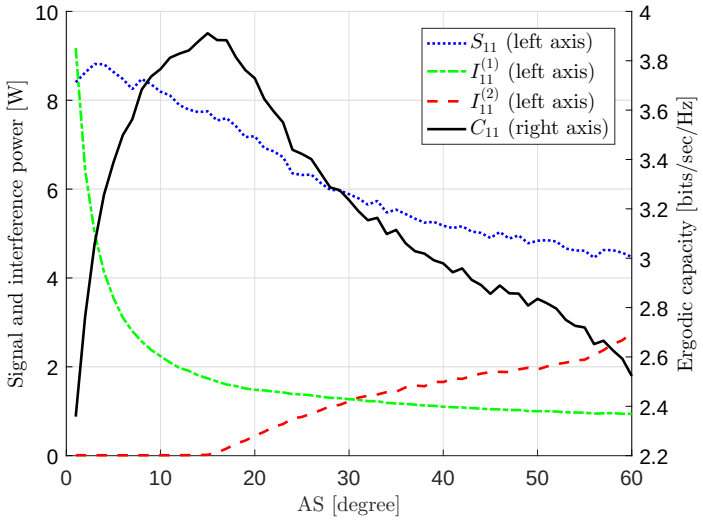


Figure 4.3: Ergodic capacity versus the AS (finite uplink SNR). When noise is considered during the uplink channel estimation stage, the signal power constantly decreases with the AS. Therefore, the raise in capacity is only due to the decrease in interference $I_{11}^{(1)}$.

a consequence, BS 1 divides the transmission power between the two users in the downlink, which is seen as a loss in the received signal power by user 1. Finally, we observe that the remaining interference $I_{11}^{(1)}$ constantly decreases with the AS. From the behavior of S_{11}, $I_{11}^{(1)}$ and $I_{11}^{(2)}$, the dependence of the capacity with respect to the AS can be explained as follows: the capacity first increases due to both the increase in signal power and decrease in interference $I_{11}^{(1)}$. Then, the capacity reaches a maximum value and starts decreasing due to the raise in interference $I_{11}^{(2)}$. We observe that the decrease in signal power at AS $= 20°$ does not cause any further significant reduction in the capacity. This occurs because the change in signal power is relatively low compared to that of the interference $I_{11}^{(2)}$.

In Fig. 4.3, we recreate the scenario in Fig. 4.2 but taking into consideration the effects of noise, that is, we assume a finite uplink SNR as in Table 4.1. The main difference between the two scenarios is the behavior of the signal power with respect to the AS. In the previous figure we observed that it started decreasing at the

point where the arriving clusters at BS 1 from both users overlapped. However, when the noise is taken into consideration, the signal power constantly decreases with the AS. Therefore, the uplink channel estimation error due to the noise causes BS 1 to transmit downlink data in undesired directions, reducing the received signal power by user 1. Furthermore, the higher the AS, the stronger the impact of the noise on the signal power.

So far, we have computed the capacity in (4.2) by means of Monte Carlo (MC) simulations since it involves the evaluation of expected values. For the rest of the paper, in order to avoid this time-consuming approach, we will compute the capacity with the closed form approximation derived in [25, Eq. (26)]. We did not employ this approach in the previous analysis since the approximations of the individual components S_{11}, $I_{11}^{(1)}$ and $I_{11}^{(2)}$ do not show the impact of the AS on the pilot contamination as clearly as the MC method. However, as shown in Fig. 4.4, the approximation of the capacity is very accurate. In Fig. 4.4, we recreate the same scenario but for different positions θ of user 2. We observe that as user 2 approaches user 1 (increasing θ), the maximum capacity is achieved at lower values of AS. The explanation for this is that the overlapping of the clusters occurs faster as the users get closer, which accelerates the appearance of the pilot contamination effect.

4.4 Optimal Angular Spread in mmWave Systems

In this section we show that the results obtained previously also apply to more complex scenarios, such as future mmWave cellular networks. To this end, we use the downlink propagation model for mmWave bands recently presented in [20] and employed in the Chapter 3.

In Table 4.2 we summarize the parameters that are common to all network configurations that we will study in this section. The value of the specific parameters are

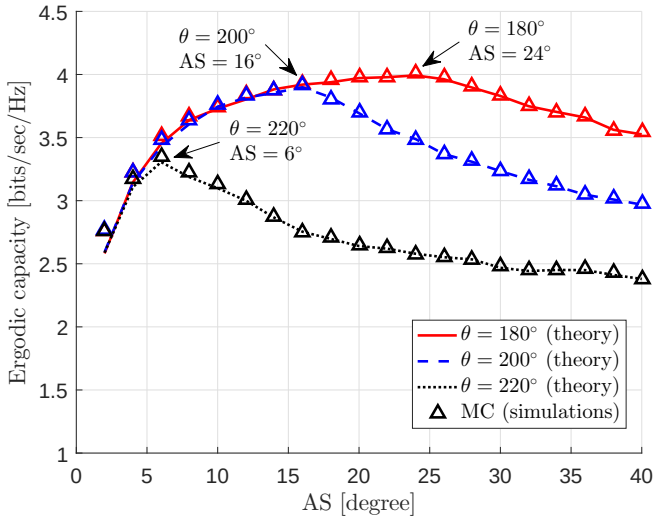


Figure 4.4: Ergodic capacity versus the AS for different user's positions. The proximity of the two users accelerates the appearance of the pilot contamination effect, which results in a lower optimal value for the AS. For this example, the BS is equipped with $M = 50$ antennas.

Table 4.2: Common Parameters for the mmWave Cellular Network

Cell radius, r_c	50 m
Distance from BS to cell edge, d_c	40 m
Carrier frequency	73 GHz
Number of spatial lobes, N_{S}	2
Number of time clusters, N_{C}	4
Number of paths per time cluster, N_{T}	16
Path-loss exponent, γ	3.3
Antenna spacing, D	$\lambda/2$
Uplink SNR	0 dB
Pilot sequence length, τ	1
Downlink SNR, ρ	0 dB

shown in the figures. To accurately represent complex cellular scenarios, the position θ of the K users per cell is uniformly distributed over $[0, 2\pi]$. Finally, a minimum of 100 network realizations have been performed for each configuration to account for the randomness of the mean AOA of the spatial lobes.

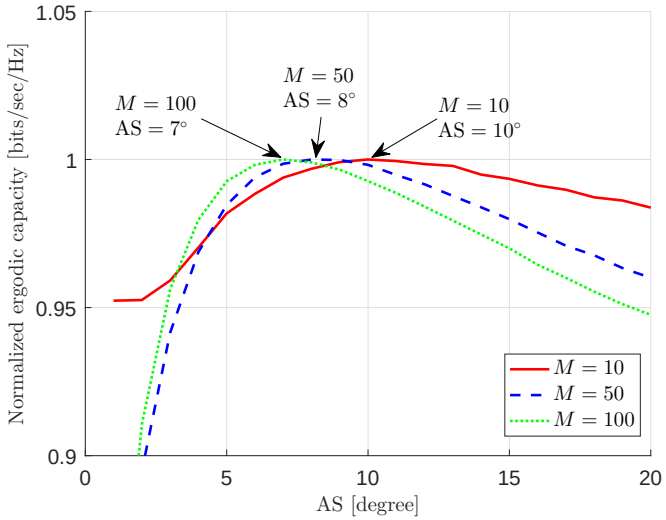


Figure 4.5: Capacity versus the AS in mmWave systems for different M values. The effects of pilot contamination in mmWave cellular networks cause the existence of an optimal AS of the spatial lobes, which depends on the number of antennas at the BS.

In Fig. 4.5 we consider two cells, each populated by one user. We are interested in the ergodic capacity of the user in cell 1 for different values of the AS. Furthermore, different values for the number of antennas M are considered. We observe that there exists an optimal value of the AS that maximizes the capacity. Furthermore, this value will depend on the number of antennas at the BS. In order to visualize the dependence of the optimal AS with the number of antennas more clearly, we show the normalized capacity, which is obtained by dividing the original capacity by its maximum value. Note that the capacity would be higher for increasing numbers of antennas at the BS.

Fig. 4.6 studies the capacity versus the AS for different number of users per cell K and number of cells N_{L}. We observe that with increasing numbers of both users per cell and cells, the optimal AS decreases. The explanation for this is that the *effective* AS increases with both K and N_{L}, which accelerates the appearance of the pilot contamination effects. As in Fig. 4.5, we show the normalized ergodic capacity.

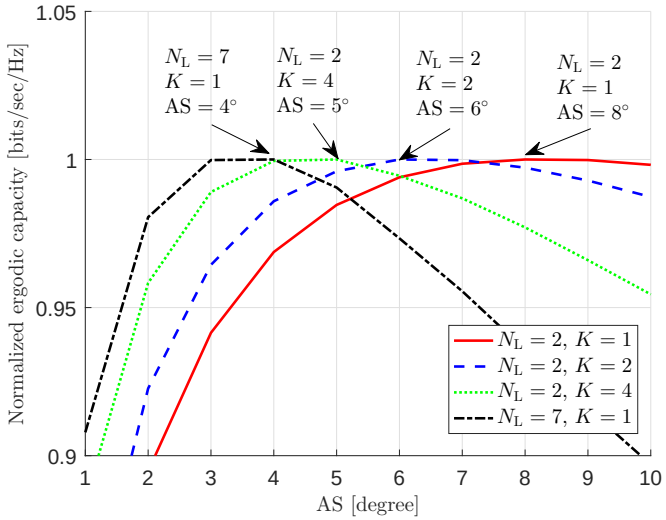


Figure 4.6: Capacity versus the AS in mmWave systems for different N_{L} and K values. Besides the number of antennas at the BS, the optimal AS also depends on the number of cells N_{L} and users per cell K. For this example, the BS is equipped with $M = 50$ antennas.

Note that the capacity would be lower for increasing numbers of cells and users per cell.

In this section we have shown that the optimal AS that maximizes the capacity in mmWave cellular networks depends on the specific configuration of the network, such as the number of antennas at the BS, number of users per cell, and total number of cells. Since, from [20], the AS depends on the working frequency, we can employ the frequency that maximizes the given configuration of the network. The work in [20] has reported studies for 28 and 73 GHz, with AS of 9° and 7°, respectively. In order to take advantage of our results, it would be necessary to study other frequencies as well.

4.5 Conclusions

In this chapter we have shown the existence of an optimal AS of the multipath clusters, which maximizes the capacity under pilot contamination. This result has been first demonstrated for a simplified cellular scenario, and then for a more complex, mmWave cellular network. Furthermore, for the latter scenario, the optimal value of the AS depends on the specific configuration of the network, such as the number of antennas installed at the BS, and the number of cells and users per cell.

Recent studies on mmWave propagation models have shown that the AS depends on the frequency. Therefore, we can use that result and the one here proposed to select the working frequency that maximizes the capacity of a given network configuration.

LIST OF PUBLICATIONS

- J. Iscar, N. Rupasinghe, I. Güvenç and S. Dikmese, "Channel and Noise Variance Estimation for 5G Wireless Networks under Pilot Contamination", submitted to *IEEE Wireless Commun. and Networking Conf. (WCNC)*, San Francisco, CA, March 2017.

- J. Iscar, N. Rupasinghe, I. Güvenç and S. Dikmese, "Non-iterative Joint Channel and Noise Variance Estimator for Massive MIMO Systems", in preparation for *IEEE J. Sel. Areas Commun.*

- J. Iscar, I. Güvenç and S. Dikmese, "Optimal Angular Spread of the Multipath Clusters in mmWave Systems under Pilot Contamination", in preparation for *IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017.

BIBLIOGRAPHY

[1] (2014, Jun.) The evolution of mobile technologies. Qualcomm. [Online]. Available: https://www.qualcomm.com/documents/evolution-mobile-technologies-1g-2g-3g-4g-lte

[2] (2016, Jun.) Ericsson mobility report: on the pulse of the networked society. [Online]. Available: https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf

[3] P. Banelli, S. Buzzi, G. Colavolpe, A. Modenini, F. Rusek, and A. Ugolini, "Modulation formats and waveforms for 5G networks: Who will be the heir of OFDM?: An overview of alternative modulation schemes for improved spectral efficiency," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 80–93, Nov. 2014.

[4] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[5] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[6] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge University Press, 2005.

[7] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[8] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO-5G system," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 905–923, 2016.

[9] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[10] V. Savaux, Y. Lout, M. Djoko-Kouam, M. Djoko-Kouam, and A. Skrzypczak, "An iterative and joint estimation of SNR and frequency selective channel for OFDM systems," in *European Wireless Conf.*, Poznan, Poland, Apr. 2012, pp. 1–7.

[11] N. Czink, G. Matz, D. Seethaler, and F. Hlawatsch, "Improved MMSE estimation of correlated MIMO channels using a structured correlation estimator," in *IEEE Workshop Signal Process. Advances Wireless Commun.*, New York, NY, Jun. 2005, pp. 595–599.

[12] F. A. Dietrich, T. Ivanov, and W. Utschick, "Estimation of channel and noise correlations for MIMO channel estimation," in *Proc. Int. ITG Workshop Smart Antennas*, Ulm, Germany, Mar. 2006.

[13] E. Dall'Anese, A. Assalini, and S. Pupolin, "Effect of channel covariance estimation error on the MIMO linear MMSE channel estimator," in *Int. Symp. Wireless Personal Multimedia Commun. (WPMC)*, Sendai, Japan, Sep. 2009.

[14] 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.

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice-Hall, 1993.

[16] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 4–17, Feb. 2008.

[17] Y.-C. Liang and F. P. S. Chin, "Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 2, pp. 222–232, Feb. 2001.

[18] M. Jordan, A. Dimofte, X. Gong, and G. Ascheid, "Conversion from uplink to downlink spatio-temporal correlation with cubic splines," in *Vehicular Technology Conf. (VTC) Spring*, Barcelona, Spain, Apr. 2009, pp. 1–5.

[19] G. Barriac and U. Madhow, "Space-time communication for OFDM with implicit channel feedback," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3111–3129, Dec. 2004.

[20] M. K. Samimi and T. S. Rappaport, "Statistical channel model with multi-frequency and arbitrary antenna beamwidth for millimeter-wave outdoor communications," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, San Diego, CA, Dec. 2015, pp. 1–7.

[21] "Spatial channel model for multiple input multiple output (MIMO) simulations," Tech. Rep. 3GPP 25.996 V13.0.0, Jan. 2016.

[22] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[23] C. Schneider, N. Iqbal, and R. S. Thom, "Impact of angular spread and number of multipath clusters on MIMO channels," in *IEEE Int. Symp. Personal, Indoor, and Mobile Radio Commun. (PIMRC)*, Hong Kong, China, Aug. 2015, pp. 512–516.

[24] M. A. Saeed, B. M. Ali, S. Khatun, and M. Ismail, "Impact of the angular spread and antenna spacing on the capacity of correlated MIMO fading channels," *Int. Arab J. Inform. Technology*, vol. 6, no. 1, pp. 60–66, Jan. 2009.

[25] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[26] X. Wautelet, C. Herzet, A. Dejonghe, J. Louveaux, and L. Vandendorpe, "Comparison of EM-based algorithms for MIMO channel estimation," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 216–226, Jan. 2007.

APPENDIX A

PROVES OF THEOREMS IN CHAPTER 2

A.1 Proof of Theorem 2.3.1:

By plugging (2.6) and (2.7) into (2.9), the Fisher information can be expressed as a function of $\mathbf{C_y}$ as

$$I\left(\sigma^2\right) = \mathrm{tr}\left[\left(\mathbf{C}_\mathbf{y}^{-1}\right)^2\right] \ . \tag{A.1}$$

The inverse of the covariance matrix $\mathbf{C_y}$ can be expressed in a simpler way by means of the Woodbury matrix identity[1]. As a result, we have

$$\mathbf{C}_\mathbf{y}^{-1} = \left(\mathbf{S}\mathbf{R}\mathbf{S}^H + \sigma^2\mathbf{I}_{M\tau}\right)^{-1} = \frac{1}{\sigma^2}\mathbf{I}_{M\tau} - \frac{1}{\sigma^4}\mathbf{S}\mathbf{A}\mathbf{S}^H \ , \tag{A.2}$$

where the matrix $\mathbf{A}$ is defined as

$$\mathbf{A} = \left(\mathbf{R}^{-1} + \frac{1}{\sigma^2}\mathbf{S}^H\mathbf{S}\right)^{-1} = \left(\mathbf{R}^{-1} + \frac{\tau}{\sigma^2}\mathbf{I}_M\right)^{-1} \ . \tag{A.3}$$

We notice that $\mathbf{A}$ depends on the inverse of $\mathbf{R}$, whose computation may be challenged when the number of antennas is large due to rank deficiency [9]. However, as we will observe, the final expression of the CRLB does not depend on this operation. Since $\mathbf{R}$ is a covariance matrix, it is positive semi-definite and therefore its eigen-decomposition is given by

$$\mathbf{R} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^H \ . \tag{A.4}$$

The matrices $\mathbf{W}$ and $\mathbf{\Lambda}$ contain the eigenvectors and eigenvalues of $\mathbf{R}$, respectively, such that $\mathbf{W}\mathbf{W}^H = \mathbf{I}_M$ and $\mathbf{\Lambda} = \mathrm{diag}\left([\lambda_1 \cdot \cdot \cdot \ \lambda_M]\right)$, with λ_m being the eigenvalues of $\mathbf{R}$. By plugging (A.4) into (A.3) and applying the Woodbury matrix identity again,

[1] $(\mathbf{X}+\mathbf{U}\mathbf{Z}\mathbf{V})^{-1} = \mathbf{X}^{-1} - \mathbf{X}^{-1}\mathbf{U}\left(\mathbf{Z}^{-1} + \mathbf{V}\mathbf{X}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}\mathbf{X}^{-1}$ for some matrices $\mathbf{X}$, $\mathbf{U}$, $\mathbf{Z}$, and $\mathbf{V}$.

the matrix $\mathbf{A}$ becomes

$$\mathbf{A} = \left(\left(\mathbf{W}\mathbf{\Lambda}\mathbf{W}^H\right)^{-1} + \frac{\tau}{\sigma^2}\mathbf{I}_M\right)^{-1} = \frac{\sigma^2}{\tau}\mathrm{I}_M - \frac{\sigma^4}{\tau}\mathrm{WBW}^H \ , \tag{A.5}$$

where $\mathbf{B} = \mathrm{diag}\left(\left[\left(\sigma^2 + \lambda_1\tau\right)^{-1} \cdot \cdot \cdot \left(\sigma^2 + \lambda_M\tau\right)^{-1}\right]\right)$.

The inverse of the covariance matrix can finally be obtained by combining (A.2) and (A.5). Then, we have

$$\mathbf{C}_{\mathbf{y}}^{-1} = \frac{1}{\sigma^2}\mathbf{I}_{M\tau} - \frac{1}{\sigma^2\tau}\mathbf{S}\mathbf{S}^H + \frac{1}{\tau}\mathbf{S}\mathbf{W}\mathbf{B}\mathbf{W}^H\mathbf{S}^H \ . \tag{A.6}$$

In order to compute the Fisher information we need to calculate $\left(\mathbf{C}_{\mathbf{y}}^{-1}\right)^2$, which is given by

$$\left(\mathbf{C}_{\mathbf{y}}^{-1}\right)^2 = \frac{1}{\sigma^4}\mathbf{I}_{M\tau} - \frac{1}{\sigma^4\tau}\mathbf{S}\mathbf{S}^H + \frac{1}{\tau}\mathbf{S}\mathbf{W}\mathbf{B}^2\mathbf{W}^H\mathbf{S}^H \ . \tag{A.7}$$

We can now compute the Fisher information by plugging (A.7) into (A.1). Then, we have

$$I\left(\sigma^2\right) = \mathrm{tr}\left[\frac{1}{\sigma^4}\mathbf{I}_{M\tau} - \frac{1}{\sigma^4\tau}\mathbf{S}\mathbf{S}^H + \frac{1}{\tau}\mathbf{S}\mathbf{W}\mathbf{B}^2\mathbf{W}^H\mathbf{S}^H\right] \ . \tag{A.8}$$

The expression above can be solved using the cyclic property of the trace[2]. As a result, we have

$$I\left(\sigma^2\right) = \frac{1}{\sigma^4}\mathrm{tr}\left[\mathbf{I}_{M\tau}\right] - \frac{1}{\sigma^4}\mathrm{tr}\left[\mathbf{I}_M\right] + \mathrm{tr}\left[\mathbf{B}^2\right] = \frac{M\tau}{\sigma^4} - \frac{M}{\sigma^4} + \frac{1}{\sigma^4}\sum_{m=1}^{M}\frac{1}{\left(1 + \frac{\tau\lambda_m}{\sigma^2}\right)^2} \ . \tag{A.9}$$

The CRLB can be obtained by plugging (A.9) into (2.8), which proves Theorem 2.3.1. □

A.2 Proof of Theorem 2.3.2:

The first step is to obtain $p\left(\mathbf{y};\sigma^2\right)$. For this purpose, we define the joint PDF of the received signal $\mathbf{y}$ and training matrix $\mathbf{S}$, that is, $p\left(\mathbf{y},\mathbf{S};\sigma^2\right)$. Now, $p\left(\mathbf{y};\sigma^2\right)$ can be

[2] $\mathrm{tr}\left[\mathbf{XUZV}\right] = \mathrm{tr}\left[\mathbf{UZVX}\right] = \mathrm{tr}\left[\mathbf{ZVXU}\right] = \mathrm{tr}\left[\mathbf{VXUZ}\right]$ for some matrices $\mathbf{X}$, $\mathbf{U}$, $\mathbf{Z}$, and $\mathbf{V}$.

obtained as the marginal PDF of $p(\mathbf{y}, \mathbf{S}; \sigma^2)$:

$$p\left(\mathbf{y}; \sigma^2\right) = \sum_{s_1} \cdot \cdot \cdot \sum_{s_\kappa} p\left(\mathbf{y}, \mathbf{S}; \sigma^2\right) \ , \tag{A.10}$$

where s_i are the κ symbols that form the pilot sequence, and the κ summations are over all the N equiprobable symbols of the constellation. The joint PDF can also be expressed as

$$p\left(\mathbf{y}, \mathbf{S}; \sigma^2\right) = p\left(\mathbf{y}|\mathbf{S}; \sigma^2\right) p\left(\mathbf{S}\right) = p\left(\mathbf{y}|\mathbf{S}; \sigma^2\right) \frac{1}{N^\kappa} \ . \tag{A.11}$$

By plugging (A.11) into (A.10), the PDF of $\mathbf{y}$ is finally given by

$$p\left(\mathbf{y}; \sigma^2\right) = \frac{1}{N^\kappa} \sum_{s_1} \cdot \cdot \cdot \sum_{s_\kappa} p\left(\mathbf{y}|\mathbf{S}; \sigma^2\right) \ . \tag{A.12}$$

In (A.12), the conditional PDF of $\mathbf{y}$ given $\mathbf{S}$ is expressed as

$$p\left(\mathbf{y}|\mathbf{S}; \sigma^2\right) = \frac{\exp\left[-\left(\mathbf{y} - \boldsymbol{\mu}_\mathbf{y}\right)^H \mathbf{C}_\mathbf{y}^{-1} \left(\mathbf{y} - \boldsymbol{\mu}_\mathbf{y}\right)\right]}{\pi^{M\kappa} \det\left[\mathbf{C}_\mathbf{y}\right]} \ . \tag{A.13}$$

By taking into considerations the result in (2.6), the expression above can be rewritten as

$$p\left(\mathbf{y}|\mathbf{S}; \sigma^2\right) = \frac{\exp\left[-\mathbf{y}^H \mathbf{C}_\mathbf{y}^{-1} \mathbf{y}\right]}{\pi^{M\kappa} \det\left[\mathbf{C}_\mathbf{y}\right]} \ , \tag{A.14}$$

where $\mathbf{C}_\mathbf{y}^{-1}$ is as in (A.6), with τ replaced by κ. The determinant of $\mathbf{C}_\mathbf{y}$ can be solved using the matrix determinant lemma[3]. Then, we have

$$\det\left[\mathbf{C}_\mathbf{y}\right] = \det\left[\mathbf{S}\mathbf{R}\mathbf{S}^H + \sigma^2 \mathbf{I}_{M\kappa}\right] = \det\left[\mathbf{R}^{-1} + \frac{\kappa}{\sigma^2}\mathbf{I}_M\right] \det\left[\mathbf{R}\right] \det\left[\sigma^2 \mathbf{I}_{M\kappa}\right] \ . \tag{A.15}$$

We can then apply the matrix determinant lemma to (A.15) again, in addition to the fact that $\mathbf{R} = \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}^H$. Hence,

$$\det\left[\mathbf{C}_\mathbf{y}\right] = \det\left[\boldsymbol{\Lambda} + \frac{\sigma^2}{\kappa}\mathbf{I}_M\right] \det\left[\frac{\kappa}{\sigma^2}\mathbf{I}_M\right] \det\left[\sigma^2 \mathbf{I}_{M\kappa}\right] = \left(\sigma^2\right)^{M\kappa} \prod_{m=1}^{M} \left(1 + \frac{\lambda_m \kappa}{\sigma^2}\right) \ . \tag{A.16}$$

[3] $\det\left[\mathbf{X} + \mathbf{U}\mathbf{Z}\mathbf{V}^H\right] = \det\left[\mathbf{Z}^{-1} + \mathbf{V}^H\mathbf{X}^{-1}\mathbf{U}\right] \det\left[\mathbf{Z}\right] \det\left[\mathbf{X}\right]$ for some matrices $\mathbf{X}$, $\mathbf{U}$, $\mathbf{Z}$, and $\mathbf{V}$.

The determinant of $\mathbf{C_y}$ above can be plugged into (A.14) and this, in turn, into (A.12). As a result, the PDF of the received signal $\mathbf{y}$ may be rewritten as

$$p\left(\mathbf{y};\sigma^2\right) = \frac{\exp\left[-\dfrac{\mathbf{y}^H\mathbf{y}}{\sigma^2}\right]}{N^{\kappa}\pi^{M\kappa}\left(\sigma^2\right)^{M\kappa}\prod\limits_{m=1}^{M}\left(1+\dfrac{\lambda_m\kappa}{\sigma^2}\right)} \times \sum_{s_1}\cdots\sum_{s_\kappa}\exp\left[\frac{\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}}{\sigma^2\kappa} - \frac{\mathbf{y}^H\mathbf{S}\mathbf{W}\mathbf{B}\mathbf{W}^H\mathbf{S}^H\mathbf{y}}{\kappa}\right] . \tag{A.17}$$

The next step to proof Theorem 2.3.2 is the evaluation of the natural logarithm of $p\left(\mathbf{y};\sigma^2\right)$ in (A.17):

$$\ln p\left(\mathbf{y};\sigma^2\right) = -\kappa\ln\left(N\right) - M\kappa\ln\left(\pi\right) - M\kappa\ln\left(\sigma^2\right) - \sum_{m=1}^{M}\ln\left(1+\frac{\lambda_m\kappa}{\sigma^2}\right) - \frac{\mathbf{y}^H\mathbf{y}}{\sigma^2} + \ln\left(\sum_{s_1}\cdots\sum_{s_\kappa} T\right) , \tag{A.18}$$

where T is defined as

$$T = \exp\left[\frac{\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}}{\sigma^2\kappa} - \frac{\mathbf{y}^H\mathbf{S}\mathbf{W}\mathbf{B}\mathbf{W}^H\mathbf{S}^H\mathbf{y}}{\kappa}\right] . \tag{A.19}$$

Finally, the first derivative of $\ln p\left(\mathbf{y};\sigma^2\right)$ with respect to σ^2 is given by

$$\frac{\partial \ln p\left(\mathbf{y};\sigma^2\right)}{\partial\,\sigma^2} = -\frac{M\kappa}{\sigma^2} + \sum_{m=1}^{M}\frac{\left(\lambda_m\kappa\right)^2}{\left(\sigma^2+\lambda_m\kappa\right)^2} + \frac{\mathbf{y}^H\mathbf{y}}{\sigma^4} + \frac{\partial \ln\left(\sum\limits_{s_1}\cdots\sum\limits_{s_\kappa} T\right)}{\partial\,\sigma^2} , \tag{A.20}$$

where

$$\frac{\partial \ln\left(\sum\limits_{s_1}\cdots\sum\limits_{s_\kappa} T\right)}{\partial\,\sigma^2} = \frac{\sum\limits_{s_1}\cdots\sum\limits_{s_\kappa}\left(TD\right)}{\sum\limits_{s_1}\cdots\sum\limits_{s_\kappa} T} . \tag{A.21}$$

In (A.21), D represents the first derivative of the exponent of T with respect to σ^2 and is given by

$$D = \frac{\partial\left(\dfrac{\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}}{\sigma^2\kappa} - \dfrac{\mathbf{y}^H\mathbf{S}\mathbf{W}\mathbf{B}\mathbf{W}^H\mathbf{S}^H\mathbf{y}}{\kappa}\right)}{\partial\,\sigma^2} = \sum_{m=1}^{M}\frac{\left|\mathbf{u}_m\right|^2}{\kappa\left(\sigma^2+\lambda_m\kappa\right)^2} - \frac{\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}}{\kappa}\frac{1}{\sigma^4} , \tag{A.22}$$

with $\mathbf{u}_m$ being the m-th element of the vector $\mathbf{u} = \mathbf{y}^H\mathbf{S}\mathbf{W}$. Furthermore, the first derivative of D with respect to σ^2 is expressed as

$$D' = \frac{\partial\, D}{\partial\, \sigma^2} = \sum_{m=1}^{M} \frac{-2|\mathbf{u}_m|^2}{\kappa\left(\sigma^2 + \lambda_m \kappa\right)^3} + \frac{2\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}}{\kappa}\frac{1}{\sigma^6}\ . \tag{A.23}$$

The second derivative of $\ln p\left(\mathbf{y};\sigma^2\right)$ with respect to σ^2 can then be obtained from (A.20), which proves Theorem 2.3.2. □

A.3 Proof of Theorem 2.3.3:

To obtain the PDF of $\mathbf{y}$, $p\left(\mathbf{y};\sigma^2\right)$, we define the joint PDF of $\mathbf{y}$ and the unknown symbols $s_{\tau+1},\ldots,s_{\tau+\kappa}$, that is, $p\left(\mathbf{y}, s_{\tau+1},\ldots,s_{\tau+\kappa};\sigma^2\right)$. As a result, the PDF of $\mathbf{y}$ is

$$p\left(\mathbf{y};\sigma^2\right) = \sum_{s_{\tau+1}} \cdots \sum_{s_{\tau+\kappa}} p\left(\mathbf{y}, s_{\tau+1},\ldots,s_{\tau+\kappa};\sigma^2\right)\ , \tag{A.24}$$

where the joint PDF can also be expressed as

$$\begin{aligned} p\left(\mathbf{y}, s_{\tau+1},\ldots,s_{\tau+\kappa};\sigma^2\right) =& p\left(\mathbf{y}|s_{\tau+1},\ldots,s_{\tau+\kappa};\sigma^2\right) p\left(s_{\tau+1},\ldots,s_{\tau+\kappa}\right) = \\ & p\left(\mathbf{y}|s_{\tau+1},\ldots,s_{\tau+\kappa};\sigma^2\right)\frac{1}{N^\kappa}\ . \end{aligned} \tag{A.25}$$

Then, by combining (A.24) with (A.25), the PDF of $\mathbf{y}$ may be rewritten as

$$p\left(\mathbf{y};\sigma^2\right) = \frac{1}{N^\kappa}\sum_{s_{\tau+1}} \cdots \sum_{s_{\tau+\kappa}} p\left(\mathbf{y}|s_{\tau+1},\ldots,s_{\tau+\kappa};\sigma^2\right)\ , \tag{A.26}$$

where the conditional PDF of $\mathbf{y}$ given the unknown symbols is given as in the NDA model by (A.14), but with κ replaced by $\tau+\kappa$ in (A.14) and (A.16), and τ replaced by $\tau+\kappa$ in (A.6). This assertion is possible since the only difference between the conditional PDF in the NDA and mixed models is the total number of symbols that form the sequence $\mathbf{s}$: κ and $\tau+\kappa$, respectively. As a consequence, the PDF of the

received signal in (A.26) may be rewritten as

$$p\left(\mathbf{y};\sigma^2\right)=\exp\left[-\frac{\mathbf{y}^H\mathbf{y}}{\sigma^2}\right]\frac{\sum_{s_{\tau+1}}\cdots\sum_{s_{\tau+\kappa}}\exp\left[\frac{\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}}{\sigma^2\left(\tau+\kappa\right)}-\frac{\mathbf{y}^H\mathbf{S}\mathbf{W}\mathbf{B}\mathbf{W}^H\mathbf{S}^H\mathbf{y}}{\left(\tau+\kappa\right)}\right]}{N^{\kappa}\pi^{M(\tau+\kappa)}\left(\sigma^2\right)^{M(\tau+\kappa)}\prod_{m=1}^{M}\left(1+\frac{\lambda_m\left(\tau+\kappa\right)}{\sigma^2}\right)}\ . \tag{A.27}$$

Once we have the PDF of $\mathbf{y}$ we can proceed as in the NDA model (Appendix A.2) to obtain the first derivative of $\ln p\left(\mathbf{y};\sigma^2\right)$ with respect to σ^2. Since the necessary steps have already been described we present the final results for T, D and D':

$$T=\exp\left[\frac{\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}}{\sigma^2\left(\tau+\kappa\right)}-\frac{\mathbf{y}^H\mathbf{S}\mathbf{W}\mathbf{B}\mathbf{W}^H\mathbf{S}^H\mathbf{y}}{\left(\tau+\kappa\right)}\right]\ , \tag{A.28}$$

$$D=\sum_{m=1}^{M}\frac{|\mathbf{u}_m|^2}{\left(\tau+\kappa\right)\left(\sigma^2+\lambda_m\left(\tau+\kappa\right)\right)^2}-\frac{\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}}{\left(\tau+\kappa\right)}\frac{1}{\sigma^4}\ , \tag{A.29}$$

$$D'=\sum_{m=1}^{M}\frac{-2|\mathbf{u}_m|^2}{\left(\tau+\kappa\right)\left(\sigma^2+\lambda_m\left(\tau+\kappa\right)\right)^3}+\frac{2\mathbf{y}^H\mathbf{S}\mathbf{S}^H\mathbf{y}}{\left(\tau+\kappa\right)}\frac{1}{\sigma^6}\ . \tag{A.30}$$

The second derivative of $\ln p\left(\mathbf{y};\sigma^2\right)$ with respect to σ^2 can then be obtained from (A.27) by computing the first derivative as in Appendix A.2. This proves Theorem 2.3.3. □

A.4 Proof of Theorem 2.3.4:

The first step is to obtain the Fisher information as in (2.9). Then, we have

$$I\left(\sigma_m^2\right)=\mathrm{tr}\left[\left(\mathbf{C}_{\mathbf{y}_m}^{-1}\right)^2\right]\ . \tag{A.31}$$

The inverse of the covariance matrix above can be evaluated with the Woodbury matrix identity. As a result, we have

$$\mathbf{C}_{\mathbf{y}_m}^{-1}=\frac{1}{\sigma_m^2}\mathbf{I}_{\tau}-\frac{\mathbf{s}\mathbf{s}^H}{\frac{\sigma_m^4}{\mathbf{R}_{mm}}+\sigma_m^2\tau}\ . \tag{A.32}$$

Therefore,

$$\left(\mathbf{C}_{\mathbf{y}_m}^{-1}\right)^2 = \frac{1}{\sigma_m^4}\mathbf{I}_\tau - \frac{\left(\frac{2\sigma_m^2}{\mathbf{R}_{mm}} + \tau\right)\mathbf{s}\mathbf{s}^H}{\left(\frac{\sigma_m^4}{\mathbf{R}_{mm}} + \sigma_m^2\tau\right)^2} \ . \tag{A.33}$$

By combining (A.33) and (A.31), and applying the cyclic property of the trace, the Fisher information of σ_m^2 becomes

$$I\left(\sigma_m^2\right) = \frac{1}{\sigma_m^4}\left(\tau - 1 + \left(1 + \frac{\mathbf{R}_{mm}\tau}{\sigma_m^2}\right)^{-2}\right) \ . \tag{A.34}$$

The CRLB can be obtained by plugging (A.34) into (2.8), which proves Theorem 2.3.4. □

A.5 Proof of Theorem 2.3.5:

We proceed as in the previous noise model. Let us define the joint PDF of received signal $\mathbf{y}_m$ and the unknown pilot sequence $\mathbf{s}$, that is, $p\left(\mathbf{y}_m, \mathbf{s}; \sigma_m^2\right)$. Hence,

$$\begin{aligned} p\left(\mathbf{y}_m; \sigma_m^2\right) &= \sum_{s_1} \cdot \ \cdot \ \cdot \sum_{s_\kappa} p\left(\mathbf{y}_m, \mathbf{s}; \sigma_m^2\right) = \sum_{s_1} \cdot \ \cdot \ \cdot \sum_{s_\kappa} p\left(\mathbf{y}_m | \mathbf{s}; \sigma_m^2\right) p\left(\mathbf{s}\right) \\ &= \frac{1}{N^\kappa} \sum_{s_1} \cdot \ \cdot \ \cdot \sum_{s_\kappa} p\left(\mathbf{y}_m | \mathbf{s}; \sigma_m^2\right) \ , \end{aligned} \tag{A.35}$$

where the conditional PDF of $\mathbf{y}_m$ given $\mathbf{s}$ is defined as

$$p\left(\mathbf{y}_m | \mathbf{s}; \sigma_m^2\right) = \frac{\exp\left[-\left(\mathbf{y}_m - \boldsymbol{\mu}_{\mathbf{y}_m}\right)^H \mathbf{C}_{\mathbf{y}_m}^{-1}\left(\mathbf{y}_m - \boldsymbol{\mu}_{\mathbf{y}_m}\right)\right]}{\pi^\kappa \det\left[\mathbf{C}_{\mathbf{y}_m}\right]} \ , \tag{A.36}$$

where the mean $\boldsymbol{\mu}_{\mathbf{y}_m}$ and the inverse of the signal covariance matrix are given by (2.20) and (A.32), respectively, with τ replaced by κ. On the other hand, the determinant of $\mathbf{C}_{\mathbf{y}_m}$ can be solved with the matrix determinant lemma:

$$\det\left[\mathbf{R}_{mm}\mathbf{s}\mathbf{s}^H + \sigma_m^2\mathbf{I}_\kappa\right] = \left(\sigma_m^2\right)^\kappa \left(1 + \frac{\mathbf{R}_{mm}\kappa}{\sigma_m^2}\right) \ . \tag{A.37}$$

As a result, the PDF of $\mathbf{y}_m$ in (A.35) may be rewritten as

$$p\left(\mathbf{y}_m;\sigma_m^2\right)=\frac{\exp\left[-\dfrac{\mathbf{y}_m^H\mathbf{y}_m}{\sigma_m^2}\right]\sum_{s_1}\cdot\cdot\cdot\sum_{s_\kappa}T}{N^\kappa\pi^\kappa\left(\sigma_m^2\right)^\kappa\left(1+\dfrac{\mathbf{R}_{mm\kappa}}{\sigma_m^2}\right)}\ , \tag{A.38}$$

where the variable T is given by

$$T=\exp\left[\frac{\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\dfrac{\sigma_m^4}{\mathbf{R}_{mm}}+\sigma_m^2\kappa}\right]\ . \tag{A.39}$$

The next step is the evaluation of the natural logarithm of $p\left(\mathbf{y}_m;\sigma_m^2\right)$ in (A.38):

$$\begin{aligned}\ln p\left(\mathbf{y}_m;\sigma_m^2\right)&=-\kappa\ln\left(N\right)-\kappa\ln\left(\pi\right)-\kappa\ln\left(\sigma_m^2\right)\\&-\ln\left(1+\frac{\mathbf{R}_{mm\kappa}}{\sigma_m^2}\right)-\frac{\mathbf{y}_m^H\mathbf{y}_m}{\sigma_m^2}+\ln\left(\sum_{s_1}\cdot\cdot\cdot\sum_{s_\kappa}T\right)\ .\end{aligned} \tag{A.40}$$

Therefore, the first derivative of $\ln p\left(\mathbf{y}_m;\sigma_m^2\right)$ with respect to σ_m^2 is given by

$$\frac{\partial\ln p\left(\mathbf{y}_m;\sigma_m^2\right)}{\partial\,\sigma_m^2}=-\frac{\kappa}{\sigma_m^2}+\frac{\dfrac{\left(\mathbf{R}_{mm\kappa}\right)^2}{\sigma_m^2}+\mathbf{R}_{mm\kappa}}{\left(\sigma_m^2+\mathbf{R}_{mm\kappa}\right)^2}+\frac{\mathbf{y}_m^H\mathbf{y}_m}{\sigma_m^4}+\frac{\partial\ln\left(\sum_{s_1}\cdot\cdot\cdot\sum_{s_\kappa}T\right)}{\partial\,\sigma_m^2}\ . \tag{A.41}$$

The second derivative can then be obtained from (A.41) as in (2.23), with D and D' given by

$$D=\frac{\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\kappa\left(\sigma_m^2+\mathbf{R}_{mm\kappa}\right)^2}-\frac{\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\kappa\sigma_m^4}\ , \tag{A.42}$$

$$D'=-\frac{2\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\kappa\left(\sigma_m^2+\mathbf{R}_{mm\kappa}\right)^3}+\frac{2\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\kappa\sigma_m^6}\ . \tag{A.43}$$

This proves Theorem 2.3.5. □

A.6 Proof of Theorem 2.3.6:

Let us consider τ known and κ unknown symbols. Then, by proceeding as in Appendix A.3, the PDF of the received signal $\mathbf{y}_m$ is given by

$$p\left(\mathbf{y}_m;\sigma_m^2\right)=\frac{\exp\left[-\dfrac{\mathbf{y}_m^H\mathbf{y}_m}{\sigma_m^2}\right]\sum\limits_{s_{\tau+1}}\cdots\sum\limits_{s_{\tau+\kappa}}T}{N^{\kappa}\pi^{(\tau+\kappa)}\left(\sigma_m^2\right)^{(\tau+\kappa)}\left(1+\dfrac{\mathbf{R}_{mm}\left(\tau+\kappa\right)}{\sigma_m^2}\right)}\ , \tag{A.44}$$

where the variable T is given by

$$T=\exp\left[\frac{\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\dfrac{\sigma_m^4}{\mathbf{R}_{mm}}+\sigma_m^2\left(\tau+\kappa\right)}\right]\ . \tag{A.45}$$

Once we have the PDF of the received signal, the second derivative of $\ln p\left(\mathbf{y}_m;\sigma_m^2\right)$ with respect to σ_m^2 can be obtained from (A.44) as in Appendix A.5, with D and D' given by

$$D=\frac{\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\left(\tau+\kappa\right)\left(\sigma_m^2+\mathbf{R}_{mm}\left(\tau+\kappa\right)\right)^2}-\frac{\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\left(\tau+\kappa\right)\sigma_m^4}\ , \tag{A.46}$$

$$D'=-\frac{2\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\left(\tau+\kappa\right)\left(\sigma_m^2+\mathbf{R}_{mm}\left(\tau+\kappa\right)\right)^3}+\frac{2\mathbf{y}_m^H\mathbf{s}\mathbf{s}^H\mathbf{y}_m}{\left(\tau+\kappa\right)\sigma_m^6}\ . \tag{A.47}$$

This proves Theorem 2.3.6. □

A.7 Proof of Theorem 2.4.1:

The MM works by evaluating the moments of the received signal $\mathbf{y}$. The objective is to find a moment that depends on the unknown parameter, in this case the noise variance. Here we show that this requirement is satisfied by the second moment, which is defined as

$$\boldsymbol{\mu}_2=\mathrm{E}\left[\mathbf{y}^H\mathbf{y}\right]\ , \tag{A.48}$$

where the expected value is taken with respect to the received signal $\mathbf{y}$. Plugging (2.31) into (A.48), we have

$$\boldsymbol{\mu}_2 = \mathrm{E}\left[\mathbf{h}^H\mathbf{S}^H\mathbf{S}\mathbf{h}\right] + \mathrm{E}\left[\mathbf{h}^H\mathbf{S}^H\mathbf{n}\right] + \mathrm{E}\left[\mathbf{n}^H\mathbf{S}\mathbf{h}\right] + \mathrm{E}\left[\mathbf{n}^H\mathbf{n}\right] \ . \tag{A.49}$$

As for the derivation of the CRLB, we consider that the channel $\mathbf{h}$ and noise $\mathbf{n}$ are independent random variables. On the other hand, since in the DA model the pilot sequence is known, the training matrix $\mathbf{S}$ is a deterministic variable. Hence,

$$\mathrm{E}\left[\mathbf{h}^H\mathbf{S}^H\mathbf{n}\right] = \mathrm{E}\left[\mathbf{h}^H\right]\mathbf{S}^H\mathrm{E}\left[\mathbf{n}\right] = 0 \ , \tag{A.50}$$

$$\mathrm{E}\left[\mathbf{n}^H\mathbf{S}\mathbf{h}\right] = \mathrm{E}\left[\mathbf{n}^H\right]\mathbf{S}\mathrm{E}\left[\mathbf{h}\right] = 0 \ . \tag{A.51}$$

The other two summands in (A.49) can be obtained as

$$\mathrm{E}\left[\mathbf{h}^H\mathbf{S}^H\mathbf{S}\mathbf{h}\right] = \tau\mathrm{E}\left[\mathbf{h}^H\mathbf{h}\right] = \tau\sum_{m=1}^{M}\mathrm{E}\left[h_m^*h_m\right] = \tau\sum_{m=1}^{M}\mathbf{R}_{mm} = \tau\mathrm{tr}\left[\mathbf{R}\right] \ , \tag{A.52}$$

$$\mathrm{E}\left[\mathbf{n}^H\mathbf{n}\right] = \sum_{i=1}^{M\tau}\mathrm{E}\left[\mathbf{n}_i^*\mathbf{n}_i\right] = \sum_{i=1}^{M\tau}\sigma^2 = M\tau\sigma^2 \ , \tag{A.53}$$

where h_m and $\mathbf{n}_i$ define the m-th and i-th element of $\mathbf{h}$ and $\mathbf{n}$, respectively.

Hence, by plugging (A.50)-(A.53) into (A.49), the second moment of $\mathbf{y}$ becomes

$$\boldsymbol{\mu}_2 = \tau\mathrm{tr}\left[\mathbf{R}\right] + M\tau\sigma^2 \ . \tag{A.54}$$

Therefore,

$$\sigma^2 = \frac{\boldsymbol{\mu}_2}{M\tau} - \frac{\mathrm{tr}\left[\mathbf{R}\right]}{M} \ . \tag{A.55}$$

The MM estimate of the noise variance is obtained by substituting in (A.55) the second moment of $\mathbf{y}$ by the given value of $\mathbf{y}^H\mathbf{y}$, that is, $\boldsymbol{\mu}_2 = \mathbf{y}^H\mathbf{y}$. This results in (2.38), which proves Theorem 2.4.1. □

A.8 Proof of Theorem 2.4.2:

In order to obtain the estimator, we proceed as in [26]. In the k-th iteration, the conditional expected value of $\ln p\left(\mathbf{x};\sigma^2\right)$ given the received signal $\mathbf{y}$ is

$$\begin{aligned}\mathrm{E}_k &= \int_{\mathbf{x}} \ln p\left(\mathbf{x};\sigma^2\right) p\left(\mathbf{x}|\mathbf{y};\widehat{\sigma^2}_k\right) d\mathbf{x} \\ &= \int_{\mathbf{y}} \left(\sum_{s_1} \cdots \sum_{s_\kappa} \ln p\left(\mathbf{x};\sigma^2\right) p\left(\mathbf{x}|\mathbf{y};\widehat{\sigma^2}_k\right) \right) d\mathbf{y} \ ,\end{aligned} \tag{A.56}$$

where

$$\ln p\left(\mathbf{x};\sigma^2\right) = \ln p\left(\mathbf{y},\mathbf{S};\sigma^2\right) = \ln \left(p\left(\mathbf{y}|\mathbf{S};\sigma^2\right) p\left(\mathbf{S}\right)\right) \ , \tag{A.57}$$

$$p\left(\mathbf{x}|\mathbf{y};\widehat{\sigma^2}_k\right) = p\left(\mathbf{y},\mathbf{S}|\mathbf{y};\widehat{\sigma^2}_k\right) = p\left(\mathbf{S}|\mathbf{y};\widehat{\sigma^2}_k\right). \tag{A.58}$$

Since the received signal $\mathbf{y}$ is known, the integral with respect to $\mathbf{y}$ in (A.56) can be removed. Considering this and the results in (A.57) and (A.58), E_k in (A.56) becomes

$$\mathrm{E}_k = \sum_{s_1} \cdots \sum_{s_\kappa} \ln \left(p\left(\mathbf{y}|\mathbf{S};\sigma^2\right) p\left(\mathbf{S}\right)\right) p\left(\mathbf{S}|\mathbf{y};\widehat{\sigma^2}_k\right) \ . \tag{A.59}$$

Furthermore, since E_k is going to be maximized with respect to σ^2 and the term $p\left(\mathbf{S}\right)$ does not depend on it, the expression above can be simplified to

$$\mathrm{E}_k = \sum_{s_1} \cdots \sum_{s_\kappa} \ln p\left(\mathbf{y}|\mathbf{S};\sigma^2\right) p\left(\mathbf{S}|\mathbf{y};\widehat{\sigma^2}_k\right) \ . \tag{A.60}$$

We notice that the result above is, in fact, the expected value of $\ln p\left(\mathbf{y}|\mathbf{S};\sigma^2\right)$ with respect to $\mathbf{S}|\mathbf{y}$ evaluated in $\widehat{\sigma^2}_k$. The expression for $\ln p\left(\mathbf{y}|\mathbf{S};\sigma^2\right)$ is given in (2.34) with τ replaced by κ. As a consequence, E_k in (A.60) becomes

$$\begin{aligned}\mathrm{E}_k = &-\frac{\mathbf{y}^H\mathbf{y}}{\sigma^2} + \frac{\mathbf{y}^H \mathrm{E}\left[\mathbf{S}\mathbf{S}^H\right]\mathbf{y}}{\sigma^2\kappa} - \sum_{m=1}^{M} \frac{\mathrm{E}\left[|\mathbf{u}_m|^2\right]}{\kappa\left(\sigma^2+\kappa\lambda_m\right)} \\ &- M\kappa\ln\left(\pi\right) - M\kappa\ln\left(\sigma^2\right) - \sum_{m=1}^{M} \ln\left(1+\frac{\lambda_m\kappa}{\sigma^2}\right) \ ,\end{aligned} \tag{A.61}$$

where the expected value is taken with respect to $\mathbf{S}|\mathbf{y}$ evaluated in $\widehat{\sigma^2}_k$. This proves equation (2.39) in Theorem 2.4.2.

The last step to determine E_k is the evaluation of the expected values in (A.61). For same random variable x, the expected value of x with respect to $\mathbf{S}|\mathbf{y}$ evaluated in $\widehat{\sigma^2}_k$ is given by

$$\mathrm{E}\left[x\right] = \sum_{s_1} \cdots \sum_{s_\kappa} x\, p\left(\mathbf{S}|\mathbf{y}; \widehat{\sigma^2}_k\right) \ . \tag{A.62}$$

Applying the Bayes' theorem, the expression above may be rewritten as

$$\begin{aligned} \mathrm{E}\left[x\right] &= \sum_{s_1} \cdots \sum_{s_\kappa} x\, \frac{p\left(\mathbf{y}|\mathbf{S}; \widehat{\sigma^2}_k\right) p\left(\mathbf{S}\right)}{p\left(\mathbf{y}; \widehat{\sigma^2}_k\right)} \\ &= \frac{1}{N^\kappa p\left(\mathbf{y}; \widehat{\sigma^2}_k\right)} \sum_{s_1} \cdots \sum_{s_\kappa} x\, p\left(\mathbf{y}|\mathbf{S}; \widehat{\sigma^2}_k\right) \ , \end{aligned} \tag{A.63}$$

where $p\left(\mathbf{y}|\mathbf{S}; \widehat{\sigma^2}_k\right)$ is as in (2.32) with τ and σ^2 replaced by κ and $\widehat{\sigma^2}_k$, respectively. Besides, $p\left(\mathbf{y}; \widehat{\sigma^2}_k\right)$ is computed as the marginal distribution of the joint PDF of $\mathbf{y}$ and $\mathbf{S}$:

$$p\left(\mathbf{y}; \widehat{\sigma^2}_k\right) = \sum_{s_1} \cdots \sum_{s_\kappa} p\left(\mathbf{y}, \mathbf{S}; \widehat{\sigma^2}_k\right) = \frac{1}{N^\kappa} \sum_{s_1} \cdots \sum_{s_\kappa} p\left(\mathbf{y}|\mathbf{S}; \widehat{\sigma^2}_k\right) \ . \tag{A.64}$$

Hence, after plugging (A.64) into (A.63), $\mathrm{E}\left[x\right]$ simplifies to (2.40), which proves Theorem 2.4.2. □

A.9 Proof of Theorem 2.4.3:

We proceed as in the model with equal noise variance. The second moment of $\mathbf{y}_m$ is given by

$$\boldsymbol{\mu}_2 = \mathrm{E}\left[\mathbf{y}_m^H \mathbf{y}_m\right] = \mathrm{E}\left[\mathbf{s}^H h_m^* h_m \mathbf{s}\right] + \mathrm{E}\left[\mathbf{s}^H h_m^* \mathbf{n}_m\right] + \mathrm{E}\left[\mathbf{n}_m^H h_m \mathbf{s}\right] + \mathrm{E}\left[\mathbf{n}_m^H \mathbf{n}_m\right] \ , \tag{A.65}$$

where the expectation is taken with respect to the received signal $\mathbf{y}_m$. Considering that h_m and $\mathbf{n}_m$ are independent random variables, the summands in (A.65) are given by

$$\mathrm{E}\left[\mathbf{s}^H h_m^* h_m \mathbf{s}\right] = \mathbf{s}^H \mathrm{E}\left[h_m^* h_m\right] \mathbf{s} = \mathbf{s}^H \mathbf{R}_{mm} \mathbf{s} = \tau \mathbf{R}_{mm} \ , \tag{A.66}$$

$$\mathrm{E}\left[\mathbf{s}^H h_m^* \mathbf{n}_m\right] = \mathbf{s}^H \mathrm{E}\left[h_m^*\right] \mathrm{E}\left[\mathbf{n}_m\right] = 0 \ , \tag{A.67}$$

$$\mathrm{E}\left[\mathbf{n}_m^H h_m \mathbf{s}\right] = \mathrm{E}\left[\mathbf{n}_m^H\right] \mathrm{E}\left[h_m\right] \mathbf{s} = 0 \ , \tag{A.68}$$

$$\mathrm{E}\left[\mathbf{n}_m^H \mathbf{n}_m\right] = \sum_{i=1}^{\tau} \mathrm{E}\left[\mathbf{n}_{m,i}^* \mathbf{n}_{m,i}\right] = \sum_{i=1}^{\tau} \sigma_m^2 = \tau \sigma_m^2 \ . \tag{A.69}$$

By plugging (A.66)-(A.69) into (A.65), the second moment is finally given by

$$\boldsymbol{\mu}_2 = \tau \mathbf{R}_{mm} + \tau \sigma_m^2 \ . \tag{A.70}$$

Hence, we have

$$\sigma_m^2 = \frac{\boldsymbol{\mu}_2}{\tau} - \mathbf{R}_{mm} \ . \tag{A.71}$$

The MM estimator of σ_m^2 is given by (A.71) with the second moment replaced by the given value of $\mathbf{y}_m^H \mathbf{y}_m$. This results in (2.50), which proves Theorem 2.4.3. □

A.10 Proof of Theorem 2.4.4:

Let us define $\mathbf{x}_m$ as the *complete* data set, which includes the *incomplete* data $\mathbf{y}_m$ and the unknown pilot sequence $\mathbf{s}$. As a consequence, E_k is given by

$$\mathrm{E}_k = \mathrm{E}\left[\ln p\left(\mathbf{x}_m; \sigma_m^2\right)\right] \ , \tag{A.72}$$

where the expected value is taken with respect to $\mathbf{x}_m|\mathbf{y}_m$ evaluated in $\widehat{\sigma_{mk}^2}$. Therefore, E_k may be rewritten as

$$\begin{aligned} \mathrm{E}_k &= \int_{\mathbf{x}_m} \ln p\left(\mathbf{x}_m; \sigma_m^2\right) p\left(\mathbf{x}_m|\mathbf{y}_m; \widehat{\sigma_{mk}^2}\right) d\mathbf{x}_m \\ &= \int_{\mathbf{y}_m} \left(\sum_{s_1} \cdots \sum_{s_\kappa} \ln p\left(\mathbf{x}_m; \sigma_m^2\right) p\left(\mathbf{x}_m|\mathbf{y}_m; \widehat{\sigma_{mk}^2}\right) \right) d\mathbf{y}_m \ , \end{aligned} \tag{A.73}$$

where

$$\ln p\left(\mathbf{x}_m; \sigma_m^2\right) = \ln p\left(\mathbf{y}_m, \mathbf{s}; \sigma_m^2\right) = \ln\left(p\left(\mathbf{y}_m|\mathbf{s}; \sigma_m^2\right) p\left(\mathbf{s}\right)\right) \ , \tag{A.74}$$

$$p\left(\mathbf{x}_m|\mathbf{y}_m; \widehat{\sigma_{mk}^2}\right) = p\left(\mathbf{y}_m, \mathbf{s}|\mathbf{y}_m; \widehat{\sigma_{mk}^2}\right) = p\left(\mathbf{s}|\mathbf{y}_m; \widehat{\sigma_{mk}^2}\right) \ . \tag{A.75}$$

Since the received signal $\mathbf{y}_m$ is known, the integral in (A.73) may be removed. By taking this into account, and the expressions in (A.74) an (A.75), E_k becomes

$$\mathrm{E}_k = \sum_{s_1} \cdots \sum_{s_\kappa} \ln\left(p\left(\mathbf{y}_m|\mathbf{s};\sigma_m^2\right)p\left(\mathbf{s}\right)\right) p\left(\mathbf{s}|\mathbf{y}_m;\widehat{\sigma_{mk}^2}\right) \ . \tag{A.76}$$

Furthermore, since $p\left(\mathbf{s}\right)$ does not depend on σ_m^2, E_k simplifies to

$$\mathrm{E}_k = \sum_{s_1} \cdots \sum_{s_\kappa} \ln p\left(\mathbf{y}_m|\mathbf{s};\sigma_m^2\right) p\left(\mathbf{s}|\mathbf{y}_m;\widehat{\sigma_{mk}^2}\right) \ . \tag{A.77}$$

As we observed in the model for equal noise variance, E_k is, in fact, the expected value of $\ln p\left(\mathbf{y}_m|\mathbf{s};\sigma_m^2\right)$ with respect to $\mathbf{s}|\mathbf{y}_m$ evaluated in $\widehat{\sigma_{mk}^2}$. The expression for $\ln p\left(\mathbf{y}_m|\mathbf{s};\sigma_m^2\right)$ is given by (2.47) with τ replaced by κ. Therefore, we have

$$\begin{aligned}\mathrm{E}_k = &-\kappa\ln\left(\pi\right) - \kappa\ln\left(\sigma_m^2\right) - \ln\left(1+\frac{\mathbf{R}_{mm}\kappa}{\sigma_m^2}\right) \\ &- \frac{\mathbf{y}_m^H\mathbf{y}_m}{\sigma_m^2} + \frac{\mathbf{y}_m^H\mathrm{E}\left[\mathbf{s}\mathbf{s}^H\right]\mathbf{y}_m}{\sigma_m^2\kappa} - \frac{\mathbf{y}_m^H\mathrm{E}\left[\mathbf{s}\mathbf{s}^H\right]\mathbf{y}_m}{\kappa\left(\sigma_m^2+\mathbf{R}_{mm}\kappa\right)} \ ,\end{aligned} \tag{A.78}$$

which proves equation (2.51) in Theorem 2.4.4. In (A.78), the expected value of $\mathbf{s}\mathbf{s}^H$, taken with respect to $\mathbf{s}|\mathbf{y}_m$ evaluated in $\widehat{\sigma_{mk}^2}$, is given by

$$\mathrm{E}\left[\mathbf{s}\mathbf{s}^H\right] = \sum_{s_1} \cdots \sum_{s_\kappa} \mathbf{s}\mathbf{s}^H p\left(\mathbf{s}|\mathbf{y}_m;\widehat{\sigma_{mk}^2}\right) \ , \tag{A.79}$$

which, after applying the Bayes' theorem, becomes

$$\mathrm{E}\left[\mathbf{s}\mathbf{s}^H\right] = \frac{1}{N^\kappa p\left(\mathbf{y}_m;\widehat{\sigma_{mk}^2}\right)} \sum_{s_1} \cdots \sum_{s_\kappa} \mathbf{s}\mathbf{s}^H p\left(\mathbf{y}_m|\mathbf{s};\widehat{\sigma_{mk}^2}\right) \ , \tag{A.80}$$

where $p\left(\mathbf{y}_m|\mathbf{s};\widehat{\sigma_{mk}^2}\right)$ is as in (2.46) with τ and σ_m^2 replaced by κ and $\widehat{\sigma_{mk}^2}$, respectively. In (A.80), $p\left(\mathbf{y}_m;\widehat{\sigma_{mk}^2}\right)$ may be rewritten as

$$p\left(\mathbf{y}_m;\widehat{\sigma_{mk}^2}\right) = \sum_{s_1} \cdots \sum_{s_\kappa} p\left(\mathbf{y}_m,\mathbf{s};\widehat{\sigma_{mk}^2}\right) = \frac{1}{N^\kappa}\sum_{s_1} \cdots \sum_{s_\kappa} p\left(\mathbf{y}_m|\mathbf{s};\widehat{\sigma_{mk}^2}\right) \ . \tag{A.81}$$

Hence, after plugging (A.81) into (A.80), $\mathrm{E}\left[\mathbf{s}\mathbf{s}^H\right]$ simplifies to (2.52), which proves Theorem 2.4.4. □