Channel Estimation and Receiver Design in Single- and Multiuser Multiantenna Systems

Gábor Fodor

Dissertation Submitted for the Degree of Doctor of the Hungarian Academy of Sciences

Contents

List of Figures

List of Tables

Summary

As an engineer and researcher, I have been dealing with the research, standardization and industrialization of wireless communication systems since the late 90's. Specifically, I have been witnessing and contributing to the evolution of the 3rd, 4th and currently to the 5th generation of cellular networks. The impact of this evolution on the society, public administration, businesses and individuals has been profound and played a key role in defining the information age and shaping the fully connected societies. Indeed, the technology footprint of cellular networks has lead to unprecedented economies of scale, which, in turn resulted in a rapid growth of technology solutions that enable them to operate with high spectral and energy efficiency in a great number of spectrum bands.

My contributions to the advances of cellular technologies lie in the fields of radio resource management and signal processing for multi-antenna systems, and specifically in the areas of channel estimation and receiver design. In particular, my contributions as a researcher are threefold: (i) conducting research for the purpose of proposing channel estimation and receiver designs that are superior to their state-of-the-art counterparts, (ii) identifying the necessary changes in communication standards that ensure the inter-operability of such novel designs and (iii) developing suitable methodology for the performance analysis of the proposed channel estimation and receiver techniques. The results of these efforts include research papers in internationally recognized journals and book chapters, communication standards specifically developed for the inter-operability of cellular systems and more than 40 internationally granted patents that are deployed cellular systems around the world.

In this thesis, I develop methodology and techniques to develop receiver algorithms that are optimal in terms of minimizing the mean squared error of the received data symbols in the presence of the estimation errors of the prevailing wireless channels through which communication takes place. The proposed methodology and techniques enable me to prove that the state-of-the-art receiver structures are suboptimal in the presence of wireless channel estimation errors, while the proposed receivers are optimal in terms of minimizing the symbol errors at multi-antenna receivers. I also developed methodology that enables the exact analysis of the symbol errors as functions of the resources used for obtaining channel estimates at the wireless receiver and transmitting data through the wireless channel.

These methods have lead to channel estimation techniques and receiver algorithms that significantly improve the spectral and energy efficiency of multi-antenna cellular systems, and simplify the design of receiver algorithms when the number of deployed antennas at cellular infrastructure nodes increases over time.

Összefoglaló

A kilencvenes évek óta a vezetéknélküli távközlő rendszerek kutatásával, szabványosításával és ipari megvalósításával foglalkozom. Munkám során hozzájárultam a harmadik, negyedik és jelenleg az ötödik generációs celluláris rendszerek fejlődéséhez. E fejlődés jelentős hatást gyakorol társadalmunk szinte minden szegmensére, és meghatározó szerepet játszik az információs társadalom formálásában. A celluláris rendszerek széleskörű használata lehetővé tette e rendszerekben alkalmazott csúcstechnológiák gyorsan megtérülő kifejlesztését és gazdaságos bevezetését. Ezeknek a mérnöki megoldásoknak köszönhetően a celluláris hálózatok milliók számára teszik lehetővé az Internetes szolgáltatások elérését.

Kutatási területem a vezetéknélküli távközlő rendszerek rádiós erőforrásainak menedzselését segítő algoritmusok, valamint az ilyen rendszerekben alkalmazható jelfeldolgozó módszerek. Specifikusan, e disszertáció eredményei hozzájárulnak a többantennás vezetéknélküli távközlő rendszerekben alkalmazható csatornabecslő és jelvevő módszerek fejlesztéséhez, ezen módszerek szabványosításához valamint teljesítményelemzéséhez. E területeken elért eredményeimet nemzetközi konferenciákon és tudományos folyóiratokban valamit számos könyvfejezetben tettem közzé. A javasolt rádiós erőforráskezlő és többantennás rendszerekben alkalmazható jelfeldolgozó technikák ipari megvalósitását jelenleg több, mint negyven nemzetközi szabadalom védi.

Jelen disszertáció olyan módszereket és technikákat mutat be, melyek optimális többantennás vevők kifejlesztését teszik lehetővé, abban az értelemben, hogy a vett adatszimbólumok átlagos négyzetes hibája minimális. A disszertációban kidolgozott módszerekről bebizonyítom, hogy a jelenleg alkalmazásban lévő vevőalgoritmusok szuboptimálisak, azaz nem minimalizálják a vett adatszimbólumok becslési hibáját olyan esetekben, melyekben a rádiós csatorna állapota a vevő számára nem pontosan ismert. Az általam javasolt módszer újdonsága, hogy a csatornabecslést és a küldött adatszimbólumok dekódolását együttesen kezeli. Ezáltal lehetővé válik a csatornabecslésre és az adatküldésre szánt rádiós erőforrások együttes kezelése és optimalizálása. Olyan matematikai módszereket dolgoztam ki, melyek ezen együttes csatornebecslő - jelvevő módszerek pontos, zárt képletekkel történő elemzését teszik lehetővé, és ezáltal betekintést nyújtanak az antennák számanak, az alkalmazott adóteljesítménynek és más mérnöki paramétereknek a kommunikáció minőségére és spektrumhatékonyságára gyakorolt hatásának az elemzésére.

A disszertációban javasolt módszerek összességükben nagyban javítják a többantennás celluláris rendszerek spektrális és rádiós erőforrás hatékonyságát és a vezetéknélküli csatornán történő kommunikáció minőségét.

Acknowledgments

First and foremost I would like to thank Miklós Telek for the many years of joint works, support and encouragement. It is a privilage to be one of his disciples in the broad field of mathematical modeling and performance analysis. I am also grateful for the support, encouragement and technical discussions and the many pieces of advice by László Pap and Sándor Imre during the preparation phase of this dissertation. I gratefully acknowledge the numerous pieces of advice and encouragement by Gábor Horváth.

I am grateful to Tamás Henk, Gyula Sallai, József Bíró, Tibor Cinkler, Edit Halász and Do Van Tien for their selfless and wholehearted encouragement and support of all sorts since I graduated at the Department of Mediainformatics in the late 90's. I also thank Norbert Reider and András Rácz for their encouragement, help and the discussions on various technical topics.

Finally, I would like to thank the incredible support, continuous encouragement and the uncountable pieces of advice of all possible kinds by my family, including my mother, my wife Viktória and my children Laura and Sebastian.

Chapter 1
Introduction

1.1 Technological Motivation

In this section I review some of the relevant literature in the areas of information theoretical aspects of wideband communications, Multiple Input Multiple Output (MIMO) transceiver design, pilot based channel estimation techniques and specifically techniques to mitigate the affects of pilot contamination (PC) and Channel State Information (CSI) errors. I also point out my contributions to this line of research.

1.1.1 Information Theoretical Aspects of Wideband Communications and Capacity Analysis

An important insight from the works reported in [1] and [2] is that there is a continuum between the extremes of communicating in non-coherent (without CSI availability) and coherent (perfect CSI) fashions over wireless channels in terms of the achieved spectral efficiency. Specifically, communicating over a completely unknown channel is subject to a penalty of the channel uncertainty, sometimes in the form of training costs. On the one hand, this penalty depends on the knowledge the receiver has of the channel and on the channel's rate of change. On the other hand, reducing this penalty by sending over only a fraction of the available degrees of freedom results in a loss of spectral efficiency.

In practice, the channel coherence time might be long enough to both estimate the fading coefficients and use such estimates to communicate coherently after the estimation period, as well as to achieve performance close to the fully coherent case (as emphasized in [3]).

Reference [2] studies the connection between the channel uncertainty penalty and the coherence length of the channel in MIMO systems. A key observation is that in the low signal-to-noise ratio (SNR) regime, estimating the channel at the receiver may not be possible and hence communication may be desirable without training. More exactly, if the channel coherence length is above a certain antenna- and SNR-dependent threshold, the noncoherent and coherent capacities become the same in the low-SNR regime.

The above results suggest that, depending on the SNR and the number of antennas, there may be a large gap between the coherent and noncoherent extremes in terms of achievable spectral efficiency, and channel learning is key in bridging this gap. Therefore, it is interesting to consider the ultra-wideband (UWB) regime and focus on the case when training signals are used for channel estimation at the receiver. The capacity of this scheme is studied in [4] to investigate the impact of multipath sparsity on achieving coherent capacity. The key results of that paper are a lower bound on the capacity of the training-based communication scheme and the coherence level that can be achieved, and the insights into the impact of channel sparsity on the achievable capacity in the UWB regime.

The work in [5] studies the impact of channel state feedback on the achievable rates in sparse wideband channels. A key insight is that a partial and/or limited feedback scheme, where only one bit per independent degree-of-freedom (DoF) is available at the transmitter, can nearly achieve the performance of a system in which perfect CSI is available at the transmitter. References [6] and [7] focus on acquiring channel state information at the transmitter in multi-user systems where the feedback from each user terminal must be limited. It is shown that the combination of long term channel statistics and instantaneous norm feedback provides sufficient information at the transmitter for efficient scheduling, beamforming and link adaptation in wide-area scenarios. More recently, the work in [8] considers a

case in which a transmitter with two antennas broadcasts to two single-antenna users. It is assumed that the two receiving users have perfect channel information, whereas the transmitter has only statistical information of each user's link (covariance matrix of the vector channel). The paper focuses on the design of beamforming vectors that depend on such statistical information and maximize the ergodic sum-rate delivered to the two users.

1.1.2 Multiple Input Multiple Output Transceiver Design

Reference [9] deals with robust MIMO precoding design with deterministic imperfect channel state information at the transmitter channel state information at the transmitter (CSIT) such that the worst-case received SNR is maximized, or the worst-case error probability is minimized. Reference [10] is concerned with the design of linear MIMO transceivers that are robust to CSI perturbations at both sides of the link that is to errors in CSIT and channel state information at the receiver channel state information at the receiver (CSIR). In that work, the design of the transceiver is based on minimizing the average sum MSE of all data streams and users. That paper assumes a perturbation error (modelled as a Gaussian additive term), but this CSI error is not controlled by pilot power or the training scheme. Therefore, the pilot-data trade-off is not considered in that paper. The model used in [11] builds on the uplink (UL)-downlink (DL) duality in sum MSE under imperfect CSI. In that work, the imperfectness of the channel knowledge is taken into account in the joint minimum mean squared error (MMSE) design. The sum MSE minimization problem for the UL and DL is subject to power constraints. However, the aspect of pilot power is not considered, and the MSE is not derived as a function of the pilot power under a constrained pilot-data budget.

1.1.3 Channel Estimation and the Pilot-to-Data Power Ratio

The seminal work reported in [12] evaluates the difference between the mutual information when the receiver has only an estimate of the channel and when it has perfect knowledge of the channel. Upper and lower bounds are established on this difference and are related to the variance of the channel measurement error. In [13] it is shown how training based channel estimation affects the capacity of the fading channel, recognizing that training imposes a substantial information-theoretic penalty, especially when the coherence interval T is only slightly larger than the number of transmit antennas or when the SNR is low. In these regimes, learning the entire channel is highly suboptimal. Conversely, if the SNR is high, and T is much larger than M, training-based schemes can come very close to achieving capacity. Therefore, the power that should be spent on training and data transmission depends on the relation between T and M. The work in [14] can be seen as a sequel of [13], taking into account intersymbol interference and the receiver technique (equalizer) used on the receiver side. However, none of these works consider the regularized MMSE receiver, and therefore the pilot power setting that minimizes the MSE of a regularized MMSE receiver is not discussed in these papers.

The Multi-user Multiple Input Multiple Output (MU-MIMO) setting is the focus of [15], in which the coherence interval of T symbols is expended for channel training, channel estimation, and precoder computation for DL transmission. Specifically, the optimum number of terminals in terms of the DL spectral efficiency is determined for a given coherence interval, number of base station antennas, and DL/UL signal-to-interference-plus-noise ratio. There is no receiver design involved and the pilot-to-data power trade-off is out of the scope of the optimization process.

Reference [16] investigates the optimization of the pilot overhead for single-user wireless fading channels, and the dependencies of this pilot overhead on various system parameters of interest (e.g. fading rate, SNR) are quantified. By finding an expansion of the spectral efficiency for the overhead

optimization in terms of the fading rate around the perfect-CSI point, the square root dependence of both the overhead and the spectral efficiency penalty is clearly identified.

1.1.4 Contributions of the Dissertation

In this dissertation, I consider the uplink of Single-User Multiple Input Multiple Output (SU-MIMO) and MU-MIMO systems that use pilot based estimation techniques to acquire CSIR and aim at minimizing the MSE of the received data symbols. To this end, in this rather general setting, the dissertation addresses the key design aspects of designing the optimal receiver structure and the finding optimal Pilot-to-Data Power Ratio (PDPR) an MMSE receiver for data reception.

My contributions to the existing literature are as follows:

- I derive the actual MMSE receiver that – in contrast to the classical or naive formula – minimizes the MSE of the estimated uplink data symbols in the presence of PDPR dependent estimation errors.
- I derive a closed form exact expression for the MSE, as a function of not only the PDPR but also the number of antennas. This exact formula allows me to arrive at the key insight that employing the actual MMSE gives large gains as the number of antennas grows large.
- For the case of uncorrelated receive antennas at the base station (BS), I give a closed form expression for the MSE of the estimated data symbols and for the pilot power that minimizes this MSE.
- For the case of correlated receive antennas at the BS, I identify the regularized MMSE receiver structure, and give closed form expressions for the achieved MSE.
- I derive a closed form for MSE of the equalized data symbols for arbitrary correlation structure between the antennas by allowing any covariance matrix of the uplink channel. This form is powerful, because it considers not only the pilot and data transmit power levels and the number of receive antennas at the base station as independent variables, but it also explicitly takes into account antenna spacing and the statistics of the angle-of-arrivals (AoAs), including the angular spread as a parameter. For example, the methodology enables me to study the impact of the PDPR on the UL performance for the popular 3GPP spatial channel model often used to model the wireless channels in cellular systems. The closed form formula takes into account the impact of number of antennas, AoA and angular spread on the MSE and thereby on the PDPR that minimizes the MSE.
- I derive closed forms for both the uplink data MSE and spectral efficiency taking into account the constraints of the comb and block type pilot arrangements. As a major difference with respect to previous works, this closed form result allows me to find the close-to- optimum number of pilot symbols and pilot power for a generic channel estimation method. In particular, I compare Least Square (LS) and MMSE channel estimation in block-type and comb-type pilot arrangement, for a BS that employs a large number of antennas.

These results allow me to study numerically the gains of using the regularized MMSE receiver and optimal pilot power levels over schemes that use the naive receivers and/or suboptimal pilot power levels. A key insight is that the pilot power that minimizes the MSE does not depend on the number of antennas, but heavily depends on the path loss between the BS and the mobile terminal.

1.2 Structure of the Dissertation

Due to the complexity of MU-MIMO systems and the inherent trade-offs associated with CSI acquisition and data reception, different system configurations are appropriate in specific deployment scenarios. Therefore, in this dissertation, I consider a wide range of system configurations, as specified in Table 1.1.

Table 1.1 System Configuration and Main Results of Chapters 3-7

Chapter	Basic System Setup	Receiver Structure	Antenna Correlation	Comment	Main Results
Chapter 3	single user	naive receiver	uncorrelated antennas	LS channel estimation, single pilot symbol	Theorem 3.1
Chapter 4	multiuser	naive and true MMSE	uncorrelated antennas	LS channel estimation, pilot sequence	Theorem 4.3.3, Theorem 4.5.1
Chapter 5	single user	naive receiver	correlated antennas	LS and MMSE channel estimation, pilot sequence	Theorem 5.5.1
Chapter 6	single user	naive receiver	uncorrelated antennas	LS and MMSE channel estimation, comb and block type pilot symbol arrangements	Theorem 6.5.1, Theorem 6.5.2
Chapter 7	multiuser	naive and true MMSE	correlated antennas	LS and MMSE channel estimation	Theorem 7.5.1, Propositions 7.5.2 and 7.5.3

Chapter 2 considers a simplified system as compared with real MU-MIMO systems deployed in practice. Despite this simplification in terms of both channel estimation and data reception, the abstract model of Chapter 2 and the derived results on the MSE of the received data symbols provide valuable insights, and serve as a starting point for the more realistic system models developed in the subsequent chapters. The main result of this chapter is a closed form expression for the MSE in a single user SIMO system in the case of uncorrelated antennas.

Chapter 3 considers a system, in which multiple users share the same time-frequency resource and facilitate CSI acquisition at the BS by transmitting orthogonal pilot sequences in the code domain. The main result of this chapter is the derivation of the receiver that minimizes the MSE of the uplink received data symbols in the presence of MU-MIMO interference and CSI errors at the receiver. As I will point out, the literature uses the terminology of MMSE receiver for receivers that, in fact, do not minimize the MSE unless perfect CSI is available at the receiver. To distinguish the true MMSE receiver from the receiver that requires perfect CSI, in this dissertation I use the terms naive receivers and MMSE receivers, where the latter refers to the receiver that does not require perfect CSI to minimize the MSE, as it is detailed in Chapter 3.

Chapter 4 relaxes the assumption on uncorrelated antennas, and thereby it aims at analyzing a significantly more realistic model than that of the preceding chapters. However, allowing for arbitrary antenna correlation structures by taking into account antenna spacing and AoA conditions of antenna arrays used in practice makes the analysis more involved. The main result of this chapter is the closed form derivation of the MSE taking into account antenna correlation.

Chapter 5 assumes perfectly uncorrelated antennas, but allows for comb and block type pilot arrangements. The results of this chapter include closed form formulas for the MSE of the received data symbols, and approximate closed form formulas for the spectral efficiency of the system. The results of this chapter are important for selecting the best channel estimation scheme and pilot symbol pattern in systems using moderate or large number of antennas.

Finally, Chapter 6 develops a model that can capture the characteristics of MU-MIMO systems in which antenna correlation cannot be ignored and uses the MMSE receiver. This chapter generalizes the results of Chapter 3 by allowing for correlated antennas. It also proves a result that was first conjectured, based on numerical experiments in Chapter 3. This result states that the optimal pilot power that minimizes the MSE is independent of the number of receive antennas when the system uses the true MMSE receiver. This result is an important difference as compared with the naive receiver, whose optimal pilot power is sensitive to the number of antennas, and have to be reconfigured when increasing the number of antennas at deployed BSs.

Chapter 2
Background

2.1 The Evolution of Multi-Antenna Systems: From Single User to Massive Multi-user Multiple Input Multiple Output Systems

Conventional communication systems equipped with a a single transmit antenna and a single receive antenna are called Single Input Single Output (SISO) communication systems (Figure 2.1, upper left). This intuitively clear terminology explicitly refers to a signal model that involves the convolution of the complex impulse response of the wireless channel (typically represented as a random variable h) and the single input x to model the single output y:

$$y = h \star x + n, \tag{2.1}$$

where n is complex baseband additive white Gaussian noise. The above equation is for a single realization of the complex single output y.

The value of multiple antenna systems as a means to improve communications, including improving the overall system capacity and transmission reliability, was recognized in the early ages of wireless communications. Specifically, adaptive transmit or receive beamforming by means of employing multiple antennas either at the transmitter or the receiver roots back to classic papers that appeared in the 1960s and 1970s [17, 18, 19]. In particular, Widrow *et al.* described the Least Mean Square (LMS) adaptive antenna array, which is a technique to adaptively determine the weights that are derived from the received signal to minimize the MSE between the received signal and a reference (pilot) signal [17, 19]. Applebaum proposed a multiple antenna array structure that adaptively suppresses sidelobe energy when the desired signal's AoA is known, such as in a radar system.

Starting from the 1980's, there has been a renewed and increased interest in employing multiple antenna techniques in commercial systems, particularly mobile and cellular systems, where multipath and unintentional interference from simultaneously served users was the main and increasing concern [20]. However, it was not until the cost of digital signal processing was dramatically reduced and commercial wireless systems matured in the late 1990s that adaptive beamforming became commercially feasible, and large scale industrial interest has started to take off.

While traditional SISO systems exploit time- or frequency-domain processing and decoding of the transmitted and received data, the use of additional antenna elements at the cellular BS or user equipment (UE) side opens up the extra spatial dimension to signal precoding and detection. Depending on the availability of multiple antennas at the transmitter and the receiver, such techniques are classified as SIMO, multiple input single output (MISO) or MIMO (Figure 2.1, upper middle and upper right). Specifically, space-time and space-frequency processing methods in SIMO, MISO and MIMO systems make use of the spatial dimension with the aim of improving the link's performance in terms of error rate, data rate or spectral and energy efficiency [19].

In the context of cellular networks, for example, in the scenario of a multi-antenna enabled BS communicating with a single antenna UE, the UL and DL are referred to as SIMO and MISO respectively. When a multi-antenna terminal is involved, a full MIMO link may be obtained, although the term MIMO is sometimes also used in a collective sense including SIMO and MISO as special cases.

A MIMO system, in which the transmitter and receiver are equipped with M and N antennas respectively, is conveniently characterized by the multi-dimensional version of (2.1) as follows:

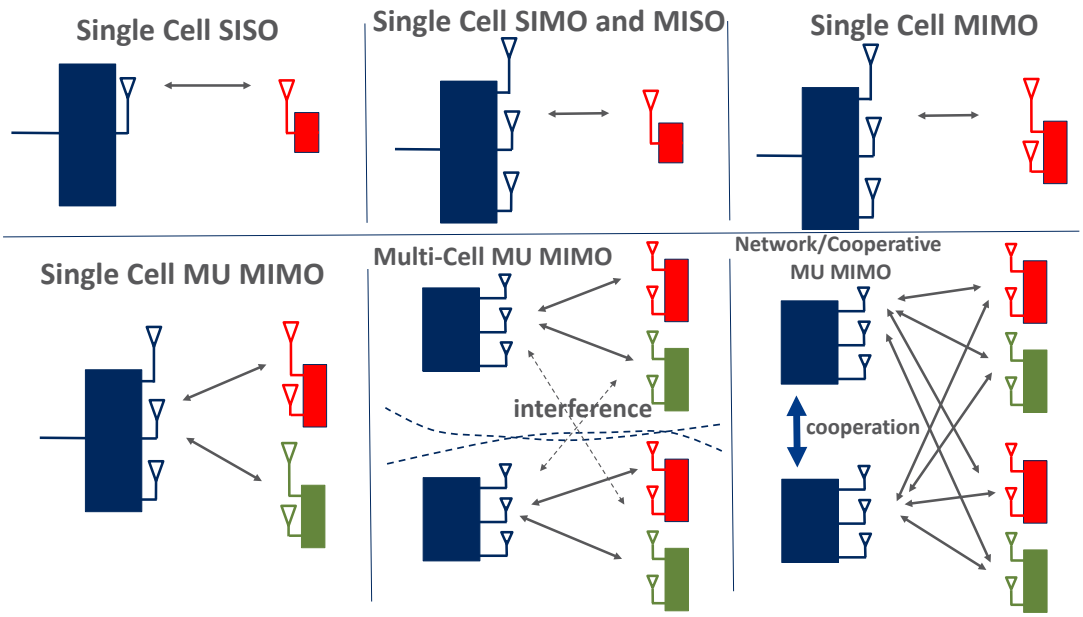

Fig. 2.1 The evolution of multiple antenna systems from single cell single input single output transmissions to cooperative network multiple input multiple output transmissions.

$$\mathbf{y} = \underbrace{\mathbf{H}}_{N\times M}\ \underbrace{\mathbf{x}}_{M\times 1} + \underbrace{\mathbf{n}}_{N\times 1} \quad \in C^{N\times 1}, \tag{2.2}$$

where $\mathbf{x}$ and $\mathbf{y}$ represent the complex M and N dimensional input and output vectors of the MIMO system respectively.

While a point-to-point multiple-antenna link between a BS and a UE is referred to as SU-MIMO, MU-MIMO features several UEs communicating simultaneously using the same frequency- and time-domain resources (Figure 2.1, lower left). By extension, considering a multi-cell system, neighboring BSs sharing their antennas and forming a virtual MIMO system to communicate with the same set of UEs in different cells are called cooperative multi-point (CoMP) or network MIMO transmission/reception (Figure 2.1, lower middle and lower right).

Multiple antenna techniques, as illustrated by Figure 2.1 offer (the combinations of) three advantages over traditional SISO systems:

- Diversity gain: The diversity gain corresponds to the mitigation of the effect of multipath fading, by means of transmitting and/or receiving over multiple wireless channels created by the multiple antennas on the transmit and/or receive sides of the communication link.
- Array gain: The array gain corresponds to a spatial version of the well-known matched-filter gain achieved by time-domain receivers.
- Spatial multiplexing gain: The spatial multiplexing gain refers to the ability to send multiple data streams in parallel and to separate them on the basis of their spatial signature. The spatial multiplexing gain is a particularly attractive gain of MIMO systems over SISO systems, because MIMO data stream multiplexing does not come at the cost of bandwidth expansion and can therefore yield drastic spectral efficiency gains.

As we shall see, the gains associated with multi-antenna systems strongly depend on the availability of CSI – the matrix $\mathbf{H}$ in (2.2) – at the transmitter and the receiver, which motivated the research and standardization communities to develop resource efficient techniques that enable the acquisition of CSIT

and CSIR. Due to their great impact on the achievable gains, these acquisition techniques form an important part of MIMO systems, as discussed in more detail in the next section.

Due to the advances in digital signal processing, antenna theory and the commercial success of MIMO, and in particular, MU-MIMO systems, the research community has been investigating the characteristics of large scale antenna systems, in which the cellular BS is equipped with a great number of antennas. Indeed, evolving wireless standards are expected to support the deployment of several tens or even hundreds of transmit and receive antennas at infrastructure nodes and over ten transmit and receive antennas at commercial UEs. It is worth noting that in the asymptotic regime of such large scale or massive MIMO systems, it turns out that the lack of accurate CSI is the main cause of performance saturation, besides hardware impairments. Therefore, scalable and resource efficient CSI acquisition techniques have been and continues to be in the focus of the MIMO community ever since the large commercial deployments of such systems have started.

2.2 Channel State Information Acquisition and Transceiver Design: Major Challenges in Multiple Input Multiple Output Systems

As noted, the spectral and energy-efficient operation of wireless systems in general, and multiple antenna systems in particular, relies on the acquisition of accurate CSIT and CSIR [21]. The main reasons for this are that (1) transmitters of modern wireless systems adapt the transmitted signal characteristics to the prevailing channel conditions and (2) the effect of the channel on the transmitted signal must be estimated in order to recover the transmitted information. As long as the receiver accurately estimates how the channel modifies the transmitted signal, it can recover the signal from the impacts of the wireless channel. In practice, pilot signal-based data-aided techniques are used not only due to their superior performance in fast fading environments, but also due to their cost efficiency and inter-operability in commercial systems. Consequently, channel estimation methods have been studied extensively and a large number of schemes, including blind, data-aided, and decision-directed non-blind techniques, have been evaluated and proposed in the literature [22, 23, 24].

As the number of antennas at the BS and the simultaneously served users grows large, it is desirable to have pilot based schemes that are scalable in terms of the required pilot symbols and provide high quality CSI for UL data detection and DL precoding. To this end, MU-MIMO systems employing a large number of antennas typically rely on channel reciprocity and employ uplink pilots to acquire CSI at BSs. Although solutions for non-reciprocal systems (such as systems operating in frequency division duplex (FDD) mode) are available [25], it is generally assumed that massive MIMO systems can advantageously operate in time division duplex (TDD) mode exploiting channel reciprocity [26, 27].

Pilot reuse generally causes contamination of the channel estimates, which is known as PC or pilot pollution. As there are a large number of channels to be estimated in MU-MIMO and massive MIMO systems, accurate CSI acquisition scaling with the number of BS antennas becomes a significant challenge due to the potentially limited number of pilots available. Indeed, PC limits the performance gains of non-cooperative MU-MIMO systems [26, 28]. Specifically, PC is known to cause a saturation effect in the signal-to-interference-plus-noise ratio (SINR) as the number of BS antennas increases to a very large value. This is in contrast to the PC exempt scenario where the SINR increases almost linearly with the number of antennas [28]. It is therefore clear that the trade-offs associated with the resources used for pilot signals and those reserved for data transmission is a key design aspect of modern wireless communication systems.

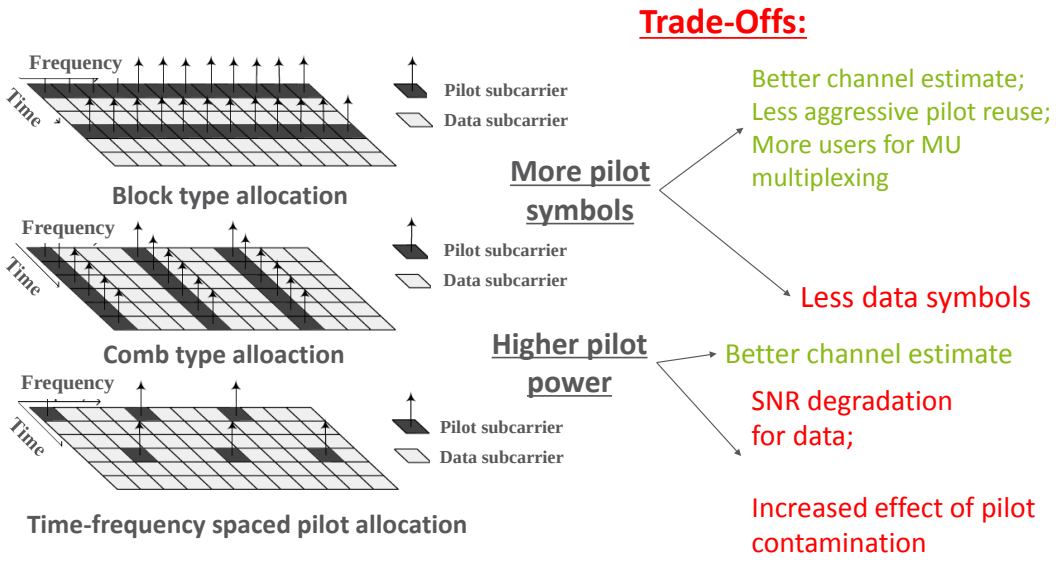

Fig. 2.2 Trade-offs associated with channel estimation, reference (pilot) signal design in MU-MIMO systems

2.3 Fundamental Trade-Offs in the Design of Multi-user Multiple Input Multiple Output Systems

Although pilot-based CSI acquisition is advantageous in fast fading environments, its inherent trade-offs must be taken into account when designing channel estimation techniques for various purposes. These purposes include demodulation, precoding or beamforming, spatial multiplexing and other channel-dependent algorithms such as frequency selective scheduling or adaptive modulation and coding scheme (MCS) selection [29, 30, 15]. The inherent trade-offs between allocating resources to pilot and data symbols include the following, as illustrated in Figure 2.2:

- Increasing the power, time, or frequency resources to pilot signals improves the quality of the channel estimate, but leaves fewer resources for uplink or downlink data transmission [29, 30, 15].
- Constructing long pilot sequences (for example, employing orthogonal symbol sequences such as those based on the well-known Zadoff-Chu sequences in Long Term Evolution (LTE) systems) helps to avoid tight pilot reuse in multi-cell systems), helps to reduce or avoid inter-cell pilot interference. This is because long pilot sequences enable to construct a great number of orthogonal sequences and, consequently, help avoid pilot reuse in neighbor cells, and thereby address the root cause of PC. On the other hand, spending a greater number of symbols on pilots increases the pilot overhead and might violate the coherence bandwidth [15, 16].
- Specifically in MU-MIMO systems, increasing the number of orthogonal pilot sequences may increase the number of spatially multiplexed users at the expense of spending more symbols when creating the orthogonal sequences [29, 30].

In particular, increasing the pilot power increases the SNR of the received pilot signal, and thereby improves the quality of channel estimation in terms of the MSE of the channel estimate [31]. Unfortunately, increasing the pilot power may also lead to the SNR degradation of the data signals, and may exacerbate the PC problem in multi-cell scenarios [32]. In addition to these inherent trade-offs, the arrangement of the pilot symbols in the time, frequency, and spatial domains have been shown to have

a significant impact on the performance of MU-MIMO and massive MIMO systems in practice, see for example [29, 30, 33].

Chapter 3
The Pilot-to-Data Power Ratio in Single User Systems

3.1 Introduction

In this chapter I consider a single input multiple output SIMO system in which the Mobile Station (MS) balances its PDPR, while the BS uses LS channel estimation to initialize a linear MMSE equalizer. The objective of this chapter is to derive a closed form for the MSE of the equalized data symbols. To obtain engineering insight into the inherent pilot-data trade-off as the number of antennas increases at the BS, my objective is to derive an MSE formula that includes not only the pilot and data transmit power levels as independent variables, but also the number of receive antennas at the base station (N_r). This formula allows me to study the impact of N_r on the MSE and thereby on the PDPR that minimizes the MSE. To the best of my knowledge the MSE formula as well as the insights obtained in the numerical section of this chapter are novel.

3.2 System Model

I consider the uplink transmission of a SIMO single cell multi-user wireless system, in which users are scheduled on orthogonal frequency channels. It is assumed that each mobile station (MS) employs an orthogonal pilot sequence, so that no interference between pilots is present in the system. This is a common assumption in massive multi-user MIMO systems in which a single MS may have a single antenna [26]. Since the channel is quasi-static frequency-flat within each transmission block, it is equivalent to model the whole pilot sequence as a single symbol per resource block with power P^p, while each data symbol is transmitted with power P. The BS estimates the channel $\mathbf{h}$ (column vector of dimension N_r, where N_r is the number of receive antennas at the BS) by employing LS channel estimators to initialize linear MMSE equalizers.

3.2.1 Channel Estimation Model

Each MS transmits an orthogonal pilot symbol x_j that is received by the BS. Thus, the column vector of the received pilot signal at the BS from the j^{th} MS is:

$$\mathbf{y}_j^p = \sqrt{P_j^p}\alpha_j \mathbf{h}_j x_j + \mathbf{n}^p, \tag{3.1}$$

where it is assumed that $\mathbf{h}_j$ is a circular symmetric complex normal distributed vector with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{C}_j$ (of size N_r), denoted as $\mathbf{h}_j \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_j)$, α_j accounts for the propagation loss, $\mathbf{n}^p \sim \mathcal{CN}(\mathbf{0}, \sigma^2\,\mathbf{I})$ is the contribution from additive Gaussian noise and the pilot symbol is scaled as $|x_j|^2 = 1, \forall j$. Since I assume orthogonal pilot sequences, the channel estimation process can be assumed independent for each MS, and I can therefore drop the index j. With a LS channel estimator, the BS estimates the channel based on (3.1) assuming

$$\hat{\mathbf{h}} = \frac{\mathbf{y}^p}{\sqrt{P^p}\alpha x},$$

that is:

$$\hat{\mathbf{h}} = \mathbf{h} + \frac{\mathbf{n}^p}{\sqrt{P^p}\alpha x}; \quad |x|^2 = 1. \tag{3.2}$$

It then follows that the estimated channel $\hat{\mathbf{h}}$ is distributed as follows:

$$\hat{\mathbf{h}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}), \tag{3.3}$$

with $\mathbf{R} \triangleq \mathbb{E}\left\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\right\} = \mathbf{C} + \frac{\sigma^2}{P^p\alpha^2}\,\mathbf{I}$.

Further, it follows that the channel estimation error $\mathbf{w} \triangleq \hat{\mathbf{h}} - \mathbf{h}$ is also normally distributed with a covariance inversely proportional to the employed pilot power:

$$\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_\mathbf{w}); \quad \mathbf{C}_\mathbf{w} \triangleq \frac{\sigma^2}{P^p\alpha^2}\,\mathbf{I}_{N_r}.$$

Equations (3.2)-(3.3) imply that $\mathbf{h}$ and $\hat{\mathbf{h}}$ are jointly circular symmetric complex Gaussian (multivariate normal) distributed random variables [34], [35]. Specifically, recall from [34] that the covariance matrix of the joint probability density function (PDF) is composed by autocovariance matrices $\mathbf{C}_{\mathbf{h},\mathbf{h}}$, $\mathbf{C}_{\hat{\mathbf{h}},\hat{\mathbf{h}}}$ and cross covariance matrices $\mathbf{C}_{\mathbf{h},\hat{\mathbf{h}}}$, $\mathbf{C}_{\hat{\mathbf{h}},\mathbf{h}}$ as

$$\begin{bmatrix} \mathbf{C}_{\mathbf{h},\mathbf{h}} & \mathbf{C}_{\mathbf{h},\hat{\mathbf{h}}} \\ \mathbf{C}_{\hat{\mathbf{h}},\mathbf{h}} & \mathbf{C}_{\hat{\mathbf{h}},\hat{\mathbf{h}}} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{C} \\ \mathbf{C} & \mathbf{R} \end{bmatrix},$$

and $\mathbf{R} = \mathbf{C} + \mathbf{C}_\mathbf{w}$.

3.2.2 Determining the Conditional Channel Distribution

From the joint PDF of $\mathbf{h}$ and $\hat{\mathbf{h}}$ I can compute the following conditional distributions.

Result 3.2.1 *Given a random channel realization* $\mathbf{h}$*, the estimated channel* $\hat{\mathbf{h}}$ *conditioned to* $\mathbf{h}$ *can be shown to be distributed as*

$$(\hat{\mathbf{h}} \mid \mathbf{h}) \sim \mathbf{h} + \mathcal{CN}(\mathbf{0}, \mathbf{C}_\mathbf{w}). \tag{3.4}$$

Result 3.2.2 *The distribution of the channel realization* $\mathbf{h}$ *conditioned to the estimate* $\hat{\mathbf{h}}$ *is normally distributed as follows:*

$$(\mathbf{h} \mid \hat{\mathbf{h}}) \sim \mathbf{D}\hat{\mathbf{h}} + \mathcal{CN}\left(\mathbf{0}, \mathbf{Q}\right), \tag{3.5}$$

where $\mathbf{D} = \mathbf{C}\mathbf{R}^{-1}$ *and* $\mathbf{Q} = \mathbf{C} - \mathbf{C}\mathbf{R}^{-1}\mathbf{C}$.

The proofs of these results are provided in the Appendix of this chapter.

To capture the tradeoff between the pilot and data power, I need to calculate the mean square error of the equalized data symbols. To this end, let us consider an equalization model in the next subsection.

3.2.3 Equalizer Model based on the Least Square Channel Estimator

The N_r dimensional data signal received by the BS is

$$\mathbf{y} = \alpha\sqrt{P}\,\mathbf{h}x + \mathbf{n}, \tag{3.6}$$

where $|x|^2 = 1$. I assume that the BS employs a naive MMSE equalizer, where the estimated channel (3.2) is taken as if it was the actual channel:

$$\mathbf{G} = \alpha\sqrt{P}\,\hat{\mathbf{h}}^H(\alpha^2 P\hat{\mathbf{h}}\hat{\mathbf{h}}^H + \sigma^2\mathbf{I}_{N_r})^{-1}. \tag{3.7}$$

Under this assumption, I state the following result as a first step towards determining the MSE.

Result 3.2.3 *Let* $\mathrm{MSE}(\mathbf{h},\hat{\mathbf{h}}) = \mathbb{E}_{x,\mathbf{n}}\left\{|\mathbf{G}\mathbf{y} - x|^2\right\}$ *be the* MSE *for the equalized symbols, given the realizations of* $\mathbf{h}$ *and* $\hat{\mathbf{h}}$*. It is*

$$\mathrm{MSE}(\mathbf{h},\hat{\mathbf{h}}) = \alpha^2 P\mathbf{G}\mathbf{h}\mathbf{h}^H\mathbf{G}^H - 2\alpha\sqrt{P}\mathrm{Re}[\mathbf{G}\mathbf{h}] + \sigma^2\mathbf{G}\mathbf{G}^H + 1. \tag{3.8}$$

The proof is presented in the Appendix. From this, my next result follows directly.

Result 3.2.4 *Let* $\mathrm{MSE}(\hat{\mathbf{h}}) = \mathbb{E}_{\mathbf{h}|\hat{\mathbf{h}}}\left\{\mathrm{MSE}(\mathbf{h},\hat{\mathbf{h}})\right\}$ *be the* MSE *for the equalized symbols, given the estimated channel realization* $\hat{\mathbf{h}}$*. It satisfies*

$$\mathrm{MSE}(\hat{\mathbf{h}}) = \mathbf{G}\left(\alpha^2 P(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q}) + \sigma^2\mathbf{I}_{N_r}\right)\mathbf{G}^H - 2\alpha\sqrt{P}\mathrm{Re}\{\mathbf{G}\mathbf{D}\hat{\mathbf{h}}\} + 1. \tag{3.9}$$

The proof is presented in the Appendix of the chapter.

3.3 Determining the Unconditional Mean Squared Error

Based on the conditional MSE expression of the preceding section, I am now interested in deriving the unconditional expectation of the MSE. To this end, the following two lemmas turn out to be useful.

Lemma 3.1. *Given a channel estimate instance* $\hat{\mathbf{h}}$*, the MMSE weighting matrix* $\mathbf{G}$*, as a function of the number of receive antennas at the base station* (N_r) *can be expressed as follows*

$$\mathbf{G} = \frac{\alpha\sqrt{P}}{\|\hat{\mathbf{h}}\|^2\alpha^2 P + \sigma^2}\,\hat{\mathbf{h}}^H, \tag{3.10}$$

where $\|\hat{\mathbf{h}}\|^2 = \hat{\mathbf{h}}^H\hat{\mathbf{h}} = \sum_{i=1}^{N_r}|\hat{h}_i|^2$.

The proof is presented in the Appendix.

Using this simple expression of $\mathbf{G}$ I can further simplify the conditional expectation of the MSE of the MMSE equalized data symbols.

Lemma 3.2. *When assuming independent channel distributions with identical variances, that is the channel covariance matrix is diagonal in the form of* $\mathbf{C} = \varrho\mathbf{I}$*, where* $\varrho \in \mathbb{R}^+$*, then the covariance matrices* $\mathbf{D}$*,* $\mathbf{Q}$ *are* $\mathbf{D} = d\mathbf{I}$*,* $\mathbf{Q} = q\mathbf{I}$ *with* $d = \varrho(\varrho + \frac{\sigma^2}{P^p\alpha^2})^{-1}$*,* $q = \varrho(1-d)$ *and* (3.9) *simplifies to*

$$\mathrm{MSE}(\hat{\mathbf{h}}) = 1 - \frac{2\|\hat{\mathbf{h}}\|^2 d\alpha^2 P}{\|\hat{\mathbf{h}}\|^2\alpha^2 P + \sigma^2} + \frac{\|\hat{\mathbf{h}}\|^2\alpha^2 P}{\left(\|\hat{\mathbf{h}}\|^2\alpha^2 P + \sigma^2\right)^2}\cdot\left[\|\hat{\mathbf{h}}\|^2 d^2\alpha^2 P + q\alpha^2 P + \sigma^2\right]. \tag{3.11}$$

The proof is presented in the Appendix of this chapter. After these preparations I can state the main theorem about the MSE.

Theorem 3.1. *The expected value of the mean square error of the equalized symbols is*

$$\mathbb{E}\{\text{MSE}\} = d^2 N_r \left(\mathcal{G}(a, 1+N_r) + pr\mathcal{G}(1+N_r, 1+N_r) - 1 \right) +$$
$$+ \frac{b}{pr} \left(\mathcal{G}(a, N_r) + pr\mathcal{G}(N_r, N_r) - 1 \right) - 2d \cdot \left(pr\mathcal{G}(N_r, 1+N_r) \right) + 1;$$

where $p = \alpha^2 P$, $a = \sigma^2$, *and*

$$\mathcal{G}(x, y) \triangleq \frac{1}{pr} e^{\frac{a}{pr}} x E_{in}\left(y, \frac{a}{pr}\right),$$

and $E_{in}(n, z) \triangleq \int_1^\infty e^{-zt}/t^n \, dt$ *is a standard exponential integral function which is commonly available in numerical programming environments (e.g., it is called ExpIntegralE in Mathematica).*

The proof is provided in the Appendix of the chapter.

It is important to note that q and d carry the dependency on the pilot power P^p and p carries the dependency on the data power P in $\mathbb{E}\{\text{MSE}\}$.

3.4 Numerical Results

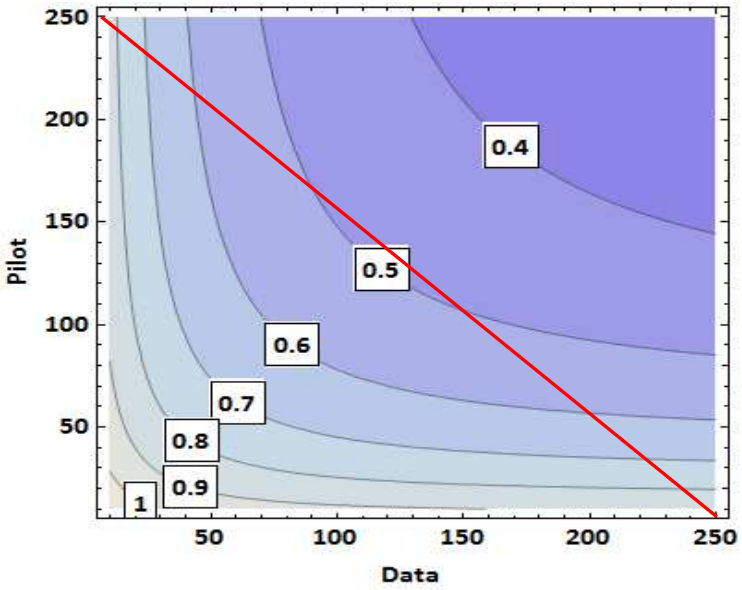

Fig. 3.1 Contour plot of the MSE achieved by specific pilot and data power settings of a SIMO system with $N_r = 2$ receiver antennas. The diagonal line indicates the feasible region of a mobile station of a sum power level of 250 mW.

In this section I consider a single cell SIMO system and concentrate on the performance of a single mobile station (MS) scheduled on a flat fading frequency channel. Unless stated otherwise, I assume that the MS has a power budget of 24 dBm that needs to be shared between the pilot and data symbols, as described in Section 3.2.

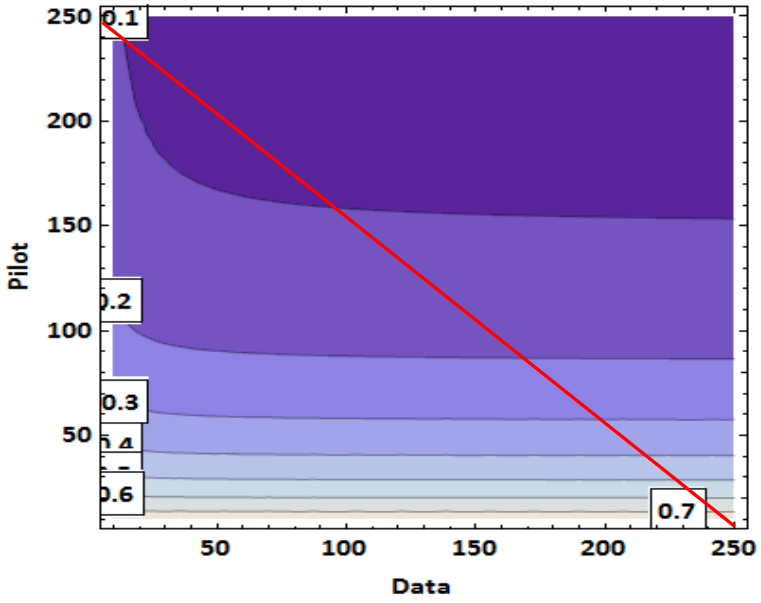

Fig. 3.2 Contour plot of the MSE achieved by specific pilot and data power settings of a SIMO system with 100 receiver antennas. The diagonal line indicates the feasible region of a mobile station of a sum power level of 250 mW.

Figure 3.1-3.2 are contour plots of the MSE of the equalized symbols as a function of the employed pilot and data transmit power levels when the number of receive antennas is $N_r = 2$ and $N_r = 100$ respectively. These figures indicate the pilot-data transmit power level pairs that maintain a given MSE. For example, in Figure 3.1 we can see that the lowest MSE value that is feasible with a 250 mW power budget is 0.5. In contrast, Figure 3.2 shows that when $N_r = 100$, the same power budget can maintain an MSE less than 0.1. From this figure it is also clear that the 'knee' of the MSE curves is shifted toward much lower data power levels, which intuitively suggests a shift in the optimal PDPR. For example, the optimal MSE with 250 mW power budget is attained at around $P^p = 145, P = 105$ on Figure 3.1 and around $P^p = 215, P = 35$ on Figure 3.2.

Figure 3.3 shows the impact of increasing the number of antennas at the base station from 2 to 100 in terms of the MSE performance as the function of the pilot and data transmit power. On the lower plane ($N_r = 100$), the pilot power minimizing the MSE is shifted towards a higher value compared with the $N_r = 2$ case (indicated with a circle) when I assume a power budget of 250 mW.

These results are reinforced by Figure 3.4 that shows the MSE as a function of the allocated pilot power under varying power budget (200mW, 225 mW and 250 mW) and assuming different number of receive antennas ($N_r = 2$, $N_r = 20$ and $N_r = 100$). Here we can clearly see the tendency that as the number of the antennas grows large, the MS needs to allocate a smaller share of the total budget to data transmission and can 'afford' a larger share of the budget for pilot transmission. This basic insight is in line with the classical observation by Marzetta predicting a diminishing data transmit power required for maintaining an SNR target [26].

Figure 3.5 shows the MSE as a function of the data power and the path loss for two antenna configurations ($N_r = 2$ and $N_r = 100$). We can observe that the data power level that minimizes the MSE is not only dependent on N_r, but also on the path loss. Specifically, for larger path loss (cell edge) users, more data power (i.e. less pilot power) minimizes the MSE than for cell center users. However, this effect becomes less pronounced as the number of antennas increases.

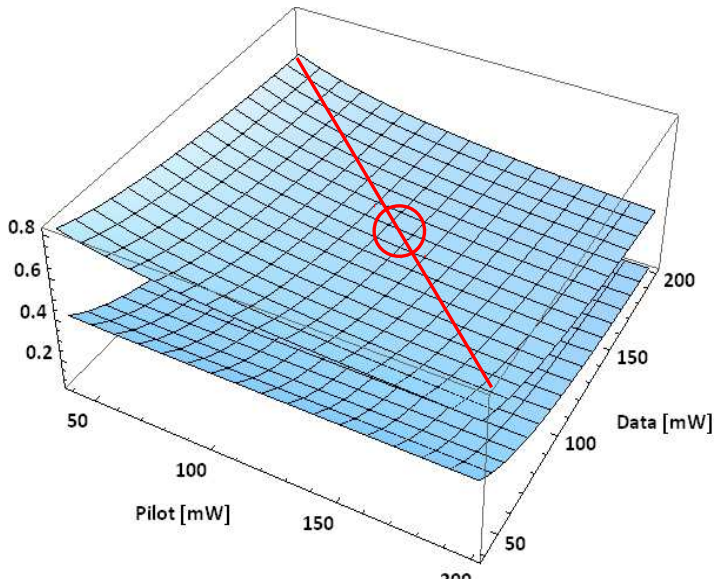

Fig. 3.3 The MSE of a SIMO system of 2 and 100 antennas. The circle indicates the optimal pilot and data power setting for the 2 antenna system with a sum power constraint of 250 mW.

3.5 Concluding Remarks

The main contribution of this chapter is the derivation of the MSE as the function of the employed pilot and data power levels as well as the number of receive antennas in SIMO systems. The numerical results provide two key insights. First, as the number of antennas at the base station increases, the MSE is minimized when a larger portion of the total transmit power budget is allocated for pilot transmission. This result is in line with the results from massive MIMO systems that suggest that the required transmit energy per bit vanishes as the number of antennas grows large. Secondly, as the path loss between the MS transmitting the pilot and base station increases, a smaller portion of the power budget needs to be spent on the pilot power. This second effect becomes less pronounced as the number of antennas at the base station increases. My summary is therefore that the PDPR that minimizes the MSE of the equalized symbols heavily depends on both the number of antennas and the MS position within the cell. An important future work is to investigate multicell systems, in which greater pilot power does not only imply lower available power for data transmission, but also a higher level of pilot contamination [26]. Therefore, these conclusions from the single cell analysis need to be reexamined in multicell systems.

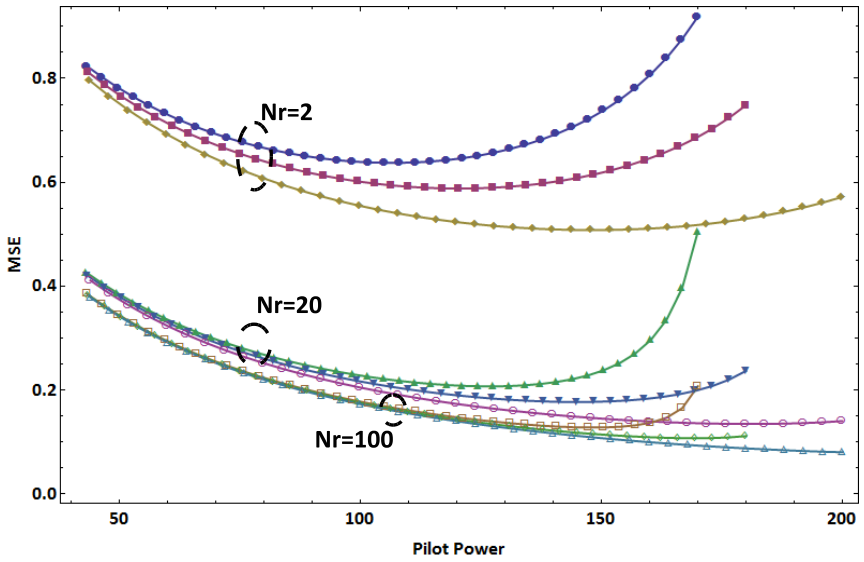

Fig. 3.4 The MSE as a function of the pilot power of a SIMO system with $N_r = 2, 20, 100$ antennas respectively, for 3 different sum power constraints (200 mW, 225 mW and 250 mW). As the number of antennas increases, the optimal pilot power increases.

Appendix of Chapter 3

Proof of Result 3.2.1

Proof. To prove the result I apply (10.24)-(10.28) of [34], but in contrast with [34], (3.4) and (3.5), in this proof I explicitly distinguish between the condition, $\mathbf{h_0}$, and the unconditional random vector, $\mathbf{h}$. According to [34] the conditional distribution of $\hat{\mathbf{h}}|\mathbf{h_0}$ is complex normal with the following properties:

$$E(\hat{\mathbf{h}}|\mathbf{h_0}) = \underbrace{E(\hat{\mathbf{h}})}_{\mathbf{0}} + \underbrace{\mathbf{C}_{\hat{\mathbf{h}},\mathbf{h}}}_{\mathbf{C}} \underbrace{\mathbf{C}_{\mathbf{h},\mathbf{h}}^{-1}}_{\mathbf{C}^{-1}} (\mathbf{h_0} - \underbrace{E(\mathbf{h})}_{\mathbf{0}}) = \mathbf{h_0}.$$

$$\mathbf{C}_{\hat{\mathbf{h}}|\mathbf{h_0}} = \mathbf{C}_{\hat{\mathbf{h}},\hat{\mathbf{h}}} - \mathbf{C}_{\hat{\mathbf{h}},\mathbf{h}}\mathbf{C}_{\mathbf{h},\mathbf{h}}^{-1}\mathbf{C}_{\mathbf{h},\hat{\mathbf{h}}} = \mathbf{R} - \mathbf{C} = \mathbf{C}_{\mathbf{w}};$$

□

Proof of Result 3.2.2

Proof. Similarly to the proof of Result 3.2.1, $\mathbf{h}|\hat{\mathbf{h}}_\mathbf{0}$ is complex normal distributed with the following mean and covariance [34]

$$E(\mathbf{h}|\hat{\mathbf{h}}_\mathbf{0}) = E(\mathbf{h}) + \mathbf{C}_{\mathbf{h},\hat{\mathbf{h}}}\mathbf{C}_{\hat{\mathbf{h}},\hat{\mathbf{h}}}^{-1} (\hat{\mathbf{h}}_\mathbf{0} - E(\hat{\mathbf{h}})) = \mathbf{C}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathbf{0};$$

$$\mathbf{C}_{\mathbf{h}|\hat{\mathbf{h}}_\mathbf{0}} = \mathbf{C}_{\mathbf{h},\mathbf{h}} - \mathbf{C}_{\mathbf{h},\hat{\mathbf{h}}}\mathbf{C}_{\hat{\mathbf{h}},\hat{\mathbf{h}}}^{-1}\mathbf{C}_{\hat{\mathbf{h}},\mathbf{h}} = \mathbf{C} - \mathbf{C}\mathbf{R}^{-1}\mathbf{C}.$$

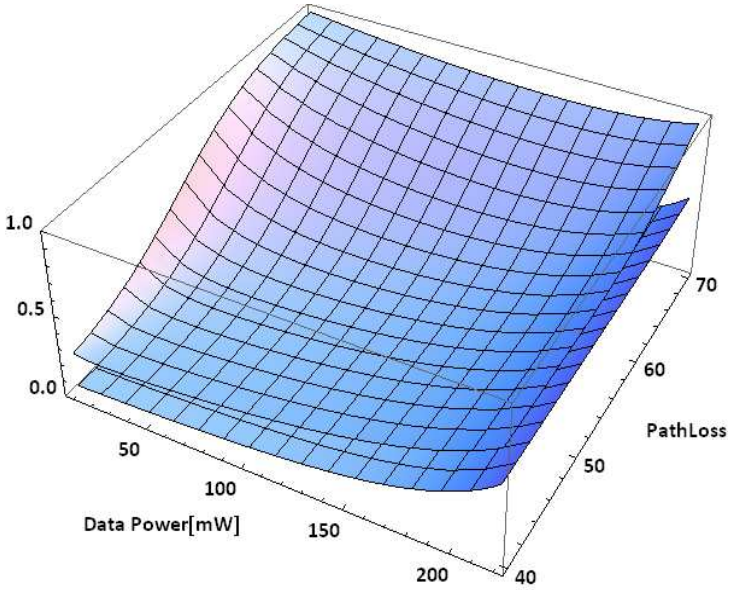

Fig. 3.5 The MSE as a function of the data power and the distance dependent path loss of a sum power constrained (250 mW) SIMO system with $N_r = 2$ and $N_r = 100$ antennas. For $N_r = 2$, as the path loss increases, the data power level that minimizes the MSE increases. However, this effect is not visible for $N_r = 100$.

□

Proof of Result 3.2.3

Proof. Having $\mathbf{y} = \alpha\sqrt{P}\mathbf{h}x + \mathbf{n}$ the mean square error of the equalized symbols, given a specific set of realizations $\mathbf{h}$ and $\hat{\mathbf{h}}$ can be calculated as:

$$\mathrm{MSE}\left(\mathbf{h}, \hat{\mathbf{h}}\right) = \mathbb{E}_{x,\mathbf{n}}\left\{|\mathbf{G}\mathbf{y} - x|^2\right\} = \mathbb{E}_{x,\mathbf{n}}\left\{|\underbrace{(\mathbf{G}\alpha\sqrt{P}\mathbf{h} - 1)x}_{a} + \underbrace{\mathbf{G}\mathbf{n}}_{b}|^2\right\}. \tag{3.12}$$

Using $|a+b|^2 = (a+b)(a^H + b^H) = aa^H + ab^H + a^H b + bb^H$ I further have

$$\begin{aligned}
\mathrm{MSE}\left(\mathbf{h},\hat{\mathbf{h}}\right) &= \mathbb{E}_{x,\mathbf{n}}\left\{((\mathbf{G}\alpha\sqrt{P}\mathbf{h}-1)x)((\mathbf{G}\alpha\sqrt{P}\mathbf{h}-1)x)^H\right\}+\\
&+\mathbb{E}_{x,\mathbf{n}}\left\{((\mathbf{G}\alpha\sqrt{P}\mathbf{h}-1)x)(\mathbf{G}\mathbf{n})^H\right\}+\mathbb{E}_{x,\mathbf{n}}\left\{\mathbf{G}\mathbf{n}((\mathbf{G}\alpha\sqrt{P}\mathbf{h}-1)x)^H\right\}+\\
&+\mathbb{E}_{x,\mathbf{n}}\left\{\mathbf{G}\mathbf{n}(\mathbf{G}\mathbf{n})^H\right\}=\\
&=(\mathbf{G}\alpha\sqrt{P}\mathbf{h}-1)\underbrace{\mathbb{E}_{x,\mathbf{n}}\{xx^H\}}_{1}(\mathbf{G}\alpha\sqrt{P}\mathbf{h}-1)^H+\\
&+(\mathbf{G}\alpha\sqrt{P}\mathbf{h}-1)\underbrace{\mathbb{E}_{x,\mathbf{n}}\{x\}}_{0}\mathbb{E}_{x,\mathbf{n}}\{\mathbf{n}^H\}\mathbf{G}^H+\\
&+\mathbf{G}\mathbb{E}_{x,\mathbf{n}}\{\mathbf{n}\}\underbrace{\mathbb{E}_{x,\mathbf{n}}\{x\}}_{0}(\mathbf{G}\alpha\sqrt{P}\mathbf{h}-1)^H+\\
&+\mathbf{G}\underbrace{\mathbb{E}_{x,\mathbf{n}}\{\mathbf{n}\mathbf{n}^H\}}_{\sigma^2\mathbf{I}}\mathbf{G}^H==\left|\mathbf{G}\mathbf{h}\alpha\sqrt{P}-1\right|^2+\sigma^2\mathbf{G}\mathbf{G}^H.
\end{aligned}$$

Finally

$$\begin{aligned}
\mathrm{MSE}\left(\mathbf{h},\hat{\mathbf{h}}\right) &= \left|\mathbf{G}\mathbf{h}\alpha\sqrt{P}-1\right|^2+\sigma^2\mathbf{G}\mathbf{G}^H=\\
&=(\mathbf{G}\mathbf{h}\alpha\sqrt{P}-1)\cdot(\mathbf{G}\mathbf{h}\alpha\sqrt{P}-1)^H+\sigma^2\mathbf{G}\mathbf{G}^H=\\
&=(\mathbf{G}\mathbf{h}\alpha\sqrt{P}-1)\cdot(\mathbf{h}^H\mathbf{G}^H\alpha\sqrt{P}-1)+\sigma^2\mathbf{G}\mathbf{G}^H=\\
&=\alpha^2P\mathbf{G}\mathbf{h}\mathbf{h}^H\mathbf{G}^H-\mathbf{G}\mathbf{h}\alpha\sqrt{P}-\mathbf{h}^H\mathbf{G}^H\alpha\sqrt{P}+1+\sigma^2\mathbf{G}\mathbf{G}^H=\\
&=\alpha^2P\mathbf{G}\mathbf{h}\mathbf{h}^H\mathbf{G}^H-\alpha\sqrt{P}\left(\mathbf{G}\mathbf{h}+\mathbf{h}^H\mathbf{G}^H\right)+1+\sigma^2\mathbf{G}\mathbf{G}^H=\\
&=\alpha^2P\mathbf{G}\mathbf{h}\mathbf{h}^H\mathbf{G}^H-\alpha\sqrt{P}\,2\mathrm{Re}\left[\mathbf{G}\mathbf{h}\right]+1+\sigma^2\mathbf{G}\mathbf{G}^H.
\end{aligned}$$

□

Proof of Result 3.2.4

Proof. I can compute $\mathbb{E}_{\mathbf{h}|\hat{\mathbf{h}}}\left\{\mathrm{MSE}\left(\mathbf{h},\hat{\mathbf{h}}\right)\right\}$ based on the conditional distribution given in (3.5):

$$(\mathbf{h}|\hat{\mathbf{h}})\sim\mathbf{D}\hat{\mathbf{h}}+\mathcal{CN}(\mathbf{0},\mathbf{Q}).$$

Recall that for a complex random column vector $\mathbf{X}$:

$$\mathbb{E}(\mathbf{X}\mathbf{X}^H)=\mathbb{E}(\mathbf{X})\;\mathbb{E}(\mathbf{X})^H+Cov(\mathbf{X}).$$

Consequently,

$$\mathbb{E}\{\mathbf{h}|\hat{\mathbf{h}}\}=\mathbf{D}\hat{\mathbf{h}}\;\text{ and }\;\mathbb{E}\{\mathbf{h}\mathbf{h}^H|\hat{\mathbf{h}}\}=\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H+\mathbf{Q}\,.$$

The expectation reads as:

$$\begin{aligned}
\mathrm{MSE}\left(\hat{\mathbf{h}}\right) &= \mathbb{E}_{\mathbf{h}|\hat{\mathbf{h}}}\left\{\mathrm{MSE}\left(\mathbf{h},\hat{\mathbf{h}}\right)\right\} = \\
&= \mathbb{E}_{\mathbf{h}|\hat{\mathbf{h}}}\left\{\mathbf{G}\mathbf{h}\mathbf{h}^H\mathbf{G}^H\alpha^2 P\right\} - \mathbb{E}_{\mathbf{h}|\hat{\mathbf{h}}}\left\{2\alpha\sqrt{P}Re[\mathbf{G}\mathbf{h}]\right\} + \\
&+ \sigma^2\mathbf{G}\mathbf{G}^H + 1 = \\
&= \mathbf{G}\mathbb{E}_{\mathbf{h}|\hat{\mathbf{h}}}\left\{\mathbf{h}\mathbf{h}^H\right\}\mathbf{G}^H\alpha^2 P - 2\alpha\sqrt{P}Re[\mathbf{G}\mathbb{E}_{\mathbf{h}|\hat{\mathbf{h}}}\{\mathbf{h}\}] + \\
&+ \sigma^2\mathbf{G}\mathbf{G}^H + 1 = \\
&= \mathbf{G}(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q})\mathbf{G}^H\alpha^2 P - 2\alpha\sqrt{P}Re[\mathbf{G}\mathbf{D}\hat{\mathbf{h}}] + \\
&+ \sigma^2\mathbf{G}\mathbf{G}^H + 1 = \\
&= \mathbf{G}\left(\alpha^2 P(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q}) + \sigma^2\mathbf{I}\right)\mathbf{G}^H - \\
&- 2\alpha\sqrt{P}Re[\mathbf{G}\mathbf{D}\hat{\mathbf{h}}] + 1.
\end{aligned} \tag{3.13}$$

I now focus on the first term of the above and make use of the following. From (3.5) I know that:

$$(\mathbf{h}|\hat{\mathbf{h}}) \sim \mathcal{CN}(\mathbf{D}\hat{\mathbf{h}}, \mathbf{Q}).$$

Then:

$$(\mathbf{h}\mathbf{h}^H|\hat{\mathbf{h}}) \sim \mathcal{CN}(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q}, \mathbf{C}_{\mathbf{h}\mathbf{h}^H|\hat{\mathbf{h}}}).$$

With this, the first term of (3.13) becomes:

$$\mathbf{G}\left(\alpha^2 P(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q}) + \sigma^2\mathbf{I}\right)\mathbf{G}^H = \mathbf{G}\Big(\alpha^2 P(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q})\Big)\mathbf{G}^H.$$

Now focusing on the second term of (3.13):

$$\begin{aligned}
-2\alpha\sqrt{P}\mathrm{Re}[\mathbf{G}\mathbf{D}\hat{\mathbf{h}}] &= -2\alpha\sqrt{P}\mathbb{E}_{x,\mathbf{n}}\left\{\mathrm{Re}\{\mathbf{G}\mathbf{h}\}\Big|\{\hat{\mathbf{h}}\}\right\} \\
&= -2\alpha\sqrt{P}\mathrm{Re}\left\{\mathbf{G}\mathbb{E}(\mathbf{h}|\hat{\mathbf{h}})\right\} = -2\alpha\sqrt{P}\mathrm{Re}\left\{\mathbf{G}\mathbf{D}\hat{\mathbf{h}}\right\}.
\end{aligned} \tag{3.14}$$

Putting the first and second term together, the right hand side of (3.13) finally takes the following form:

$$\mathrm{MSE}\left(\hat{\mathbf{h}}\right) = \mathbf{G}\Big(\alpha^2 P\left(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q}\right)\Big)\mathbf{G}^H - 2\alpha\sqrt{P}\mathrm{Re}\left[\mathbf{G}\mathbf{D}\hat{\mathbf{h}}\right] + \sigma^2\mathbf{G}\mathbf{G}^H + 1. \tag{3.15}$$

□

Proof of Lemma 3.1

Proof. From (3.7) I have

$$\mathbf{G}\Big(\alpha^2 P\hat{\mathbf{h}}\hat{\mathbf{h}}^H + \sigma^2\mathbf{I}\Big) = \alpha\sqrt{P}\hat{\mathbf{h}}^H, \tag{3.16}$$

whose solution for $\mathbf{G}$ is (3.10). To show this, I substitute (3.10) into the left hand side of (3.16) and obtain

$$\frac{\alpha\sqrt{P}}{\|\hat{\mathbf{h}}\|^2\alpha^2P+\sigma^2}\,\hat{\mathbf{h}}^H\left(\alpha^2P\hat{\mathbf{h}}\hat{\mathbf{h}}^H+\sigma^2\mathbf{I}\right)=$$
$$\frac{\alpha\sqrt{P}}{\|\hat{\mathbf{h}}\|^2\alpha^2P+\sigma^2}\,\alpha^2P\underbrace{\hat{\mathbf{h}}^H\hat{\mathbf{h}}}_{\|\hat{\mathbf{h}}\|^2}\hat{\mathbf{h}}^H+\frac{\alpha\sqrt{P}}{\|\hat{\mathbf{h}}\|^2\alpha^2P+\sigma^2}\sigma^2\hat{\mathbf{h}}^H\mathbf{I}=$$
$$\alpha\sqrt{P}\left(\frac{\alpha^2P\|\hat{\mathbf{h}}\|^2}{\|\hat{\mathbf{h}}\|^2\alpha^2P+\sigma^2}\,\hat{\mathbf{h}}^H+\frac{\sigma^2}{\|\hat{\mathbf{h}}\|^2\alpha^2P+\sigma^2}\hat{\mathbf{h}}^H\right)=$$
$$\alpha\sqrt{P}\hat{\mathbf{h}}^H\,,$$

which is indeed the right hand side of (3.16). According to the matrix inversion lemma for matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ of size $n\times n$, $n\times m$, $m\times m$, $m\times n$, respectively, I have

$$(\mathbf{A}+\mathbf{BCD})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{DA}^{-1}\mathbf{B}+\mathbf{C}^{-1}\right)^{-1}\mathbf{DA}^{-1}\,.$$

Substituting $\mathbf{A}=\sigma^2\mathbf{I}$, $\mathbf{B}=\alpha\sqrt{P}\hat{\mathbf{h}}$, $\mathbf{C}=1$, $\mathbf{D}=\alpha\sqrt{P}\hat{\mathbf{h}}^H$, I have

$$(\sigma^2\mathbf{I}+\alpha^2P\hat{\mathbf{h}}\hat{\mathbf{h}}^H)^{-1}=\frac{1}{\sigma^2}\mathbf{I}-\frac{1}{\sigma^2}\mathbf{I}\alpha\sqrt{P}\hat{\mathbf{h}}\cdot\quad\cdot\left(\alpha\sqrt{P}\hat{\mathbf{h}}^H\frac{1}{\sigma^2}\mathbf{I}\alpha\sqrt{P}\hat{\mathbf{h}}+1\right)^{-1}\alpha\sqrt{P}\hat{\mathbf{h}}^H\frac{1}{\sigma^2}\mathbf{I}=$$
$$=\frac{1}{\sigma^2}\mathbf{I}-\frac{\frac{\alpha^2P}{\sigma^4}\hat{\mathbf{h}}\hat{\mathbf{h}}^H}{\frac{\alpha^2P}{\sigma^2}\hat{\mathbf{h}}^H\hat{\mathbf{h}}+1},$$

where $\hat{\mathbf{h}}^H\hat{\mathbf{h}}=\|\hat{\mathbf{h}}\|^2$. Finally,

$$\mathbf{G}=\alpha\sqrt{P}\hat{\mathbf{h}}^H(\alpha^2P\hat{\mathbf{h}}\hat{\mathbf{h}}^H+\sigma^2\mathbf{I})^{-1}=\frac{\alpha\sqrt{P}}{\sigma^2}\hat{\mathbf{h}}^H-\frac{\alpha\sqrt{P}\frac{\alpha^2P}{\sigma^4}\hat{\mathbf{h}}^H\hat{\mathbf{h}}}{\frac{\alpha^2P}{\sigma^2}\|\hat{\mathbf{h}}\|^2+1}\hat{\mathbf{h}}^H=$$
$$=\frac{\alpha\sqrt{P}}{\sigma^2}\hat{\mathbf{h}}^H\left(1-\frac{\alpha^2P\|\hat{\mathbf{h}}\|^2}{\alpha^2P\|\hat{\mathbf{h}}\|^2+\sigma^2}\right)=\frac{\alpha\sqrt{P}}{\sigma^2}\hat{\mathbf{h}}^H\left(\frac{\sigma^2}{\alpha^2P\|\hat{\mathbf{h}}\|^2+\sigma^2}\right)=$$
$$=\frac{\alpha\sqrt{P}}{\alpha^2P\|\hat{\mathbf{h}}\|^2+\sigma^2}\hat{\mathbf{h}}^H.$$

□

Proof of Lemma 3.2

Proof. To simplify the notation I introduce z such that

$$\mathbf{G}=\frac{\alpha\sqrt{P}}{\|\hat{\mathbf{h}}\|^2\alpha^2P+\sigma^2}\,\hat{\mathbf{h}}^H=z\,\hat{\mathbf{h}}^H$$

Substituting this into (3.9) and using $\mathbf{D}=d\mathbf{I}$, $\mathbf{Q}=q\mathbf{I}$ yields (since $d\in\mathbb{R}^+$):

$$\begin{aligned}
\mathrm{MSE}\left(\hat{\mathbf{h}}\right) &= z\hat{\mathbf{h}}^H\left[\alpha^2 P\left(d\mathbf{I}\hat{\mathbf{h}}\hat{\mathbf{h}}^H d\mathbf{I}+q\mathbf{I}\right)+\sigma^2\mathbf{I}\right]z\hat{\mathbf{h}}-2\alpha\sqrt{P}\mathrm{Re}\left\{z\hat{\mathbf{h}}^H d\mathbf{I}\hat{\mathbf{h}}\right\}+1 = \\
&= z\hat{\mathbf{h}}^H\left[\alpha^2 P\left(d^2\hat{\mathbf{h}}\hat{\mathbf{h}}^H+q\mathbf{I}\right)+\sigma^2\mathbf{I}\right]z\hat{\mathbf{h}}-2\alpha\sqrt{P}\mathrm{Re}\left\{zd\hat{\mathbf{h}}^H\hat{\mathbf{h}}\right\}+1 = \\
&= z^2\alpha^2 Pd^2\hat{\mathbf{h}}^H\hat{\mathbf{h}}\hat{\mathbf{h}}^H\hat{\mathbf{h}}+z^2\alpha^2 Pq\hat{\mathbf{h}}^H\hat{\mathbf{h}}+z^2\sigma^2\hat{\mathbf{h}}^H\hat{\mathbf{h}}-2\alpha\sqrt{P}\mathrm{Re}\left\{zd\|\hat{\mathbf{h}}\|^2\right\}+1 = \\
&= z^2\alpha^2 Pd^2\|\hat{\mathbf{h}}\|^2\|\hat{\mathbf{h}}\|^2+z^2\alpha^2 Pq\|\hat{\mathbf{h}}\|^2+z^2\sigma^2\|\hat{\mathbf{h}}\|^2-2\alpha\sqrt{P}zd\|\hat{\mathbf{h}}\|^2+1 = \\
&= z^2\|\hat{\mathbf{h}}\|^2\left(\alpha^2 Pd^2\|\hat{\mathbf{h}}\|^2+\alpha^2 Pq+\sigma^2\right)-2\alpha\sqrt{P}zd\|\hat{\mathbf{h}}\|^2+1\,,
\end{aligned}$$

which is equivalent with (3.11). □

Proof of Theorem 3.1

Proof. To shorten the notation I introduce $Y=\|\hat{\mathbf{h}}\|^2$, $p=\alpha^2 P$ and $b=qp+\sigma^2$ and rewrite (3.11) with these notations.

$$\mathrm{MSE}(\hat{\mathbf{h}}) = d^2 p\frac{pY^2}{(\sigma^2+pY)^2}+b\frac{pY}{(\sigma^2+pY)^2}-2d\frac{pY}{\sigma^2+pY}+1. \tag{3.17}$$

Next, I observe that $\mathrm{MSE}(\hat{\mathbf{h}})$ depends on $\hat{\mathbf{h}}$ only through $\|\hat{\mathbf{h}}\|^2$ and $\|\hat{\mathbf{h}}\|^4$. Since $\hat{\mathbf{h}}$ is a complex normal random vector (3.3) with covariance matrix $\mathbf{R}=r\mathbf{I}$ and $r=\varrho+\frac{\sigma^2}{Pp\alpha^2}$, it follows that $Y=\|\hat{\mathbf{h}}\|^2$ follows the Gamma$(N_r,1/r)$ distribution (the sum of N_r independent r.v. which are exponentially distributed with parameter $1/r$) with probability density function:

$$f_Y(x)=\frac{r^{-N_r}x^{N_r-1}e^{-x/r}}{(N_r-1)!}\quad x>0. \tag{3.18}$$

I will make use of the following integrals:

$$\int_{x=0}^{\infty}\frac{r^{-N_r}x^{N_r-1}e^{-x/r}}{(N_r-1)!}\cdot\frac{px^2}{(a+px)^2}dx=\frac{N_r\left(-pr+e^{\frac{a}{pr}}\left(a+(1+N_r)pr\right)E_{in}\left(1+N_r,\frac{a}{pr}\right)\right)}{p^2r}; \tag{3.19}$$

$$\int_{x=0}^{\infty}\frac{r^{-N_r}x^{N_r-1}e^{-x/r}}{(N_r-1)!}\cdot\frac{px}{(a+px)^2}dx=\frac{-pr+e^{\frac{a}{pr}}\left(a+N_r pr\right)E_{in}\left(N_r,\frac{a}{pr}\right)}{p^2r^2}; \tag{3.20}$$

and

$$\int_{x=0}^{\infty}\frac{r^{-N_r}x^{N_r-1}e^{-x/r}}{(N_r-1)!}\cdot\frac{px}{(a+px)}dx=e^{\frac{a}{pr}}N_r E_{in}\left(1+N_r,\frac{a}{pr}\right), \tag{3.21}$$

where $E_{in}(n,z)\triangleq\int_1^{\infty}e^{-zt}/t^n\,dt$ is a standard exponential integral function.

Theorem 3.1 follows from averaging (3.17) according to the density function (3.18) and using (3.19)-(3.21):

$$
\begin{aligned}
\mathbb{E}\{\mathrm{MSE}\} &= d^2 p \int_{x=0}^{\infty} \frac{px^2}{(a+px)^2} f_Y(x)dx + b\int_{x=0}^{\infty} \frac{px}{(a+px)^2} f_Y(x)dx - \\
&- 2d\int_{x=0}^{\infty} \frac{px}{a+px} f_Y(x)dx + 1 = \\
&= d^2 p \cdot N_r \left(e^{\frac{a}{pr}} \left(a + (1+N_r)pr \right) \cdot E_{in}\left(1+N_r, \frac{a}{pr}\right) - pr \right) \frac{1}{p^2 r} + \\
&+ b \cdot \left(e^{\frac{a}{pr}} \left(a + N_r pr \right) \cdot E_{in}\left(N_r, \frac{a}{pr}\right) - pr \right) \frac{1}{p^2 r^2} - \\
&- 2d \cdot \left(e^{\frac{a}{pr}} N_r E_{in}\left(1+N_r, \frac{a}{pr}\right) \right) + 1 = \\
&= d^2 N_r \frac{1}{pr} \left(e^{\frac{a}{pr}} a E_{in}\left(1+N_r, \frac{a}{pr}\right) + e^{\frac{a}{pr}} (1+N_r) pr E_{in}\left(1+N_r, \frac{a}{pr}\right) - pr \right) + \\
&+ b \frac{1}{p^2 r^2} \left(e^{\frac{a}{pr}} a E_{in}\left(N_r, \frac{a}{pr}\right) + e^{\frac{a}{pr}} N_r pr E_{in}\left(N_r, \frac{a}{pr}\right) - pr \right) - \\
&- 2d \cdot \left(e^{\frac{a}{pr}} N_r E_{in}\left(1+N_r, \frac{a}{pr}\right) \right) + 1 = \\
&= d^2 N_r \left(\mathcal{G}(a, 1+N_r) + pr\mathcal{G}(1+N_r, 1+N_r) - 1 \right) + \\
&+ \frac{b}{pr} \left(\mathcal{G}(a, N_r) + pr\mathcal{G}(N_r, N_r) - 1 \right) - 2d \cdot \left(pr\mathcal{G}(N_r, 1+N_r) \right) + 1;
\end{aligned}
$$

where

$$
\mathcal{G}(x,y) \triangleq \frac{1}{pr} e^{\frac{a}{pr}} x E_{in}\left(y, \frac{a}{pr}\right).
$$

□

Chapter 4
The Minimum Mean Squared Error Receiver in the Presence of Channel Estimation Errors

4.1 Introduction

As discussed, in MU-MIMO systems, the fundamental trade-off between spending resources CSI acquisition and data transmission is known to affect the performance in terms of spectral and energy efficiency [16], [36]. Therefore, balancing the pilot-to-data power ratio (PDPR) [29] and determining the number of pilot and data symbols are important aspects of designing MU-MIMO systems [15], [37], [38]. From a different perspective, a related work combined a transmitter employing a linear dispersion code (LDC) and a linear MMSE detector at the receiver [39]. It has been found that optimizing the average normalized MSE is relevant for detectors employing a linear front end and helps designing optimal transmit strategies. In this chapter I build on the results on SU-MIMO systems in Chapter 2 and consider the uplink of a MU-MIMO system employing an MMSE receiver for data reception [40]. Similarly to the receiver studied in the previous chapter, the MU-MIMO MMSE receiver is initialized by the estimates of the CSI rather than assuming the availability of perfect CSI. Thus, the contribution of this chapter to the existing literature is two-fold:

1. I derive the actual MMSE receiver that, – in contrast to the classical or *naïve* formula [41] – minimizes the MSE of the estimated uplink data symbols in the presence of PDPR dependent estimation errors.
2. Secondly, I derive a closed form exact expression for the MSE, as a function of not only the PDPR but also the number of antennas. This exact formula allows me to arrive at the key insight that employing the actual MMSE gives large gains as the number of antennas grows large.

4.2 System and Channel Estimation Model

In this chapter I consider the uplink of a MU-MIMO system, in which the MS transmit orthogonal pilot sequences $\mathbf{s} = \left[s_1, ..., s_{\tau_p}\right]^T \in \mathbb{C}^{\tau_p \times 1}$, in which each pilot symbol is scaled as $|s_i|^2 = 1$, for $i = 1, .., \tau_p$. The pilot sequences are constructed such that they remain orthogonal as long as the number of spatially multiplexed users is maximum τ_p. Specifically, without loss of generality, it is assumed that the number MU-MIMO users is $K \leq \tau_p$. In practice, $K \ll N_r$, where N_r is the number of antennas at the BS.

In this chapter I assume a comb type arrangement of the pilot symbols [42]. Given F subcarriers in the coherence bandwidth, a fraction of τ_p subcarriers are allocated to the pilot and $F_d = F - \tau_p$ subcarriers are allocated to the data symbols. Each MS transmits at a constant power P_{tot}, however, the transmission power can be distributed unequally in each subcarrier. In particular, considering a transmitted power P_p for each pilot symbol and P for each data symbol transmission, the sum constraint of $\tau_p P_p + (F - \tau_p)P = P_{tot}$ is enforced. Thus, the $N_r \times \tau_p$ matrix of the received pilot signal from a specific MS at the BS can be conveniently written as:

$$\mathbf{Y}^p = \alpha\sqrt{P_p}\mathbf{h}\mathbf{s}^T + \mathbf{N}, \tag{4.1}$$

where it is assumed that $\mathbf{h} \in \mathbb{C}^{N_r \times 1}$ is a circular symmetric complex normal distributed column vector with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{C}$ (of size N_r), denoted as $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, α accounts for the

propagation loss, $\mathbf{N} \in \mathbb{C}^{N_r \times \tau_p}$ is the spatially and temporally additive white Gaussian noise (AWGN) with element-wise variance σ_p^2, where the index p refers to the noise power on the received *pilot* signal.

In this chapter I assume that the BS uses the popular least square (LS) estimator that relies on correlating the received signal with the known pilot sequence. Note that my methodology to determine the MSE of the received data is not confined to the LS estimator, but is directly applicable to an MMSE or other channel estimation techniques as well. For each MS, the BS utilizes pilot sequence orthogonality and estimates the channel based on (4.1) assuming:

$$\hat{\mathbf{h}} = \mathbf{h} + \mathbf{w} = \frac{1}{\alpha\sqrt{P_p}}\mathbf{Y}^p\mathbf{s}^*(\mathbf{s}^T\mathbf{s}^*)^{-1} = \mathbf{h} + \frac{1}{\alpha\sqrt{P_p}\tau_p}\mathbf{N}\mathbf{s}^*, \tag{4.2}$$

where $\mathbf{s}^* = \left[s_1^*, ..., s_{\tau_p}^*\right]^T \in \mathbb{C}^{\tau_p \times 1}$ denotes the vector of pilot symbols and $(\mathbf{s}^T\mathbf{s}^*) = \tau_p$. By considering $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, it follows that the estimated channel $\hat{\mathbf{h}}$ is a circular symmetric complex normal distributed vector $\hat{\mathbf{h}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$, with

$$\mathbf{R} \triangleq \mathbb{E}\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\} = \mathbf{C} + \frac{\sigma_p^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r}. \tag{4.3}$$

As it was shown in [42], the distribution of the channel realization $\mathbf{h}$ conditioned on the estimate $\hat{\mathbf{h}}$ is normally distributed as follows:

$$(\mathbf{h} \mid \hat{\mathbf{h}}) \sim \mathbf{D}\hat{\mathbf{h}} + \mathcal{CN}\left(\mathbf{0}, \mathbf{Q}\right), \tag{4.4}$$

where $\mathbf{D} \triangleq \mathbf{C}\mathbf{R}^{-1}$ and $\mathbf{Q} \triangleq \mathbf{C} - \mathbf{C}\mathbf{R}^{-1}\mathbf{C}$.

4.2.1 Perfect Channel Estimation

With perfect channel estimation:

$$(\mathbf{h} \mid \hat{\mathbf{h}}) \sim \mathbf{D}\hat{\mathbf{h}} + \mathcal{CN}\left(\mathbf{0}, \mathbf{Q}\right), \tag{4.5}$$

where $\mathbf{D} \triangleq \mathbf{C}\mathbf{R}^{-1} = \mathbf{I}_{N_r}$ and $\mathbf{Q} \triangleq \mathbf{C} - \mathbf{C}\mathbf{R}^{-1}\mathbf{C} = \mathbf{0}_{N_r}$.

That is:

$$\mathbf{G}^\star = \mathbf{G}^{\text{naive}}, \quad \mathbf{G}_\kappa^\star = \mathbf{G}_\kappa^{\text{naive}}\mathbf{G}_\kappa$$

where

$$\mathbf{G}^{\text{naïve}} = \mathbf{G}^{\text{naïve}}(\hat{\mathbf{h}}) = \frac{\alpha\sqrt{P}\hat{\mathbf{h}}^H}{\alpha^2 P||\hat{\mathbf{h}}||^2 + \sigma^2}$$

Note that the squared error is:

$$|\mathbf{G}\mathbf{y} - x|^2$$

4.3 The Linear Minimum Mean Squared Error Receiver

4.3.1 Received Data Signal Model

The MU-MIMO received data signal at the BS can be written as:

$$\mathbf{y} = \underbrace{\alpha_\kappa \mathbf{h}_\kappa \sqrt{P_\kappa} x_\kappa}_{\text{User-}\kappa} + \underbrace{\sum_{k\neq\kappa}^{K} \alpha_k \mathbf{h}_k \sqrt{P_k} x_k}_{\text{Other users}} + \mathbf{n}_d, \tag{4.6}$$

where $\alpha_k \cdot \mathbf{h}_k$ is the $M \times 1$ vector channel including large and small scale fading between User-k and the BS (α_k and $\mathbf{h}_k$ respectively), x_k is the transmitted data symbol by User-k and $\mathbf{n}_d$ emphasizes the noise on the received *data* signal.

4.3.2 Employing a Minimum Mean Squared Error Receiver at the Base Station

In this chapter the BS employs an MMSE receiver $\mathbf{G}_\kappa \in \mathbb{C}^{1\times N_r}$ to estimate the data symbol transmitted by User-κ. Recall that the MMSE receiver aims at minimizing the mean-square error between the estimate $\mathbf{G}_\kappa \mathbf{y}$ and the transmitted symbol x_κ:

$$\mathbf{G}_\kappa \triangleq \arg\min_{\mathbf{G}} \mathbb{E}\{\text{MSE}\} = \arg\min_{\mathbf{G}} \mathbb{E}\{|\mathbf{G}\mathbf{y} - x_\kappa|^2\}. \tag{4.7}$$

When the BS employs a naïve receiver, the estimated channel is taken as if it was the actual channel:

$$\mathbf{G}_\kappa^{\text{naive}} = \alpha_\kappa \sqrt{P_\kappa} \hat{\mathbf{h}}_\kappa^H (\alpha_\kappa^2 P_\kappa \hat{\mathbf{h}}_\kappa \hat{\mathbf{h}}_\kappa^H + \sigma_d^2 \mathbf{I})^{-1}. \tag{4.8}$$

As we shall see, this receiver does not minimize the MSE.

4.3.3 Determining the Actual Minimum Mean Squared Error Receiver Matrix

This section is concerned with determining the MMSE receiver matrix $\mathbf{G}_k$ that the BS should use to demodulate the received data signal such that the data estimation error for User-κ is minimized taking explicitly account that the BS has access only to the estimated channels $\hat{\mathbf{h}}_\kappa$, as opposed to the naïve receiver that minimizes the MSE only when perfect channel estimation is assumed. To this end, I consider the MSE of the estimated data symbols of the tagged User-κ: obtained from the signal model of (4.6) using a receiver vector $\mathbf{G}_\kappa$:

$$\begin{aligned}
&\text{MSE}\,(\mathbf{G}_\kappa, \mathbf{h}_1, \ldots, \mathbf{h}_K) = \mathbb{E}_{x,\mathbf{n}_d}\{|\mathbf{G}_\kappa \mathbf{y} - x_\kappa|^2\} = \\
&= \mathbb{E}_{x,\mathbf{n}_d} \left| (\mathbf{G}_\kappa \alpha_\kappa \mathbf{h}_\kappa \sqrt{P_\kappa} - 1) x_\kappa + \sum_{k\neq\kappa}^{K} \mathbf{G}_\kappa \alpha_k \mathbf{h}_k \sqrt{P_k} x_k + \mathbf{G}_\kappa \mathbf{n}_d \right|^2 = \\
&= \mathbb{E}_{x,\mathbf{n}_d} \left| (\mathbf{G}_\kappa \alpha_\kappa \mathbf{h}_\kappa \sqrt{P_\kappa} - 1) x_\kappa \right|^2 + \sum_{k\neq\kappa}^{K} P_k \mathbb{E}_{x,\mathbf{n}_d} |\mathbf{G}_\kappa \alpha_k \mathbf{h}_k x_k|^2 + \\
&\quad + \mathbb{E}_{x,\mathbf{n}_d} |\mathbf{G}_\kappa \mathbf{n}_d|^2,
\end{aligned} \tag{4.9}$$

where I utilized that $\mathbb{E}\{x_k\} = 0$ and $\mathbb{E}\{\mathbf{n}_d\} = \mathbf{0}$.

Additionally, utilizing $\mathbb{E}\{x_k x_k^*\} = 1$ and $\mathbb{E}\{\mathbf{n}_d \mathbf{n}_d^H\} = \sigma_d^2 \mathbf{I}_{N_r}$, I have:

$$\mathrm{MSE}\left(\mathbf{G}_\kappa, \mathbf{h}_1, \ldots, \mathbf{h}_K\right) = \left|\mathbf{G}_\kappa \alpha_\kappa \mathbf{h}_\kappa \sqrt{P_\kappa} - 1\right|^2 + \sum_{k \neq \kappa}^{K} P_k |\mathbf{G}_\kappa \alpha_k \mathbf{h}_k|^2 + \sigma_d^2 \mathbf{G}_\kappa \mathbf{G}_\kappa^H, \tag{4.10}$$

from which my first result follows.

Result 4.3.1 *When the BS uses the receiver vector* $\mathbf{G}_\kappa$*, the* MSE *of the received data symbols of the tagged user* κ *assuming perfect channel state information at the base station is:*

$$\begin{aligned}
&\mathrm{MSE}\left(\mathbf{G}_\kappa, \mathbf{h}_\kappa\right) = \\
&\mathbb{E}_{\mathbf{h}_1, \ldots, \mathbf{h}_{\kappa-1}, \mathbf{h}_{\kappa+1}, \ldots, \mathbf{h}_K} \{\mathrm{MSE}\left(\mathbf{G}_\kappa, \mathbf{h}_1, \ldots, \mathbf{h}_K\right)\} = \\
&\alpha_\kappa^2 P_\kappa \mathbf{G}_\kappa \mathbf{h}_\kappa \mathbf{h}_\kappa^H \mathbf{G}_\kappa^H - \alpha_\kappa \sqrt{P_\kappa} (\mathbf{G}_\kappa \mathbf{h}_\kappa + \mathbf{h}_\kappa^H \mathbf{G}_\kappa^H) + 1 + \\
&+ \underbrace{\sum_{k \neq \kappa}^{K} \alpha_k^2 P_k \mathbf{G}_\kappa \mathbf{C}_k \mathbf{G}_\kappa^H}_{\text{Multi-User Interference}} + \sigma_d^2 \mathbf{G}_\kappa \mathbf{G}_\kappa^H .
\end{aligned} \tag{4.11}$$

Although this result is useful, I need an expression for the MSE as a function of $\hat{\mathbf{h}}$, rather than $\mathbf{h}$.

Result 4.3.2 *The* MSE *of the received data symbols of the tagged user* κ *as a function of the estimated channel at the BS is:*

$$\begin{aligned}
&\mathrm{MSE}\left(\mathbf{G}_\kappa, \hat{\mathbf{h}}_\kappa\right) = \mathbb{E}_{\mathbf{h}_\kappa | \hat{\mathbf{h}}_\kappa} \mathrm{MSE}\left(\mathbf{G}_\kappa, \mathbf{h}_\kappa\right) = \\
&\alpha_\kappa^2 P_\kappa \mathbf{G}_\kappa (\mathbf{D}_\kappa \hat{\mathbf{h}}_\kappa \hat{\mathbf{h}}_\kappa^H {\mathbf{D}_\kappa}^H + \mathbf{Q}_\kappa) \mathbf{G}_\kappa^H \\
&- \alpha_\kappa \sqrt{P_\kappa} (\mathbf{G}_\kappa \mathbf{D}_\kappa \hat{\mathbf{h}}_\kappa + \hat{\mathbf{h}}_\kappa^H {\mathbf{D}_\kappa}^H \mathbf{G}_\kappa^H) + 1 + \\
&\sum_{k \neq \kappa}^{K} \alpha_k^2 P_k \mathbf{G}_\kappa \mathbf{C}_k \mathbf{G}_\kappa^H + \sigma_d^2 \mathbf{G}_\kappa \mathbf{G}_\kappa^H .
\end{aligned} \tag{4.12}$$

Using these results, I am in the position of deriving the optimal MU-MIMO receiver vector for User-κ:

Theorem 4.3.3 *The optimal* $\mathbf{G}_\kappa^\star$ *can be derived as:*

$$\begin{aligned}
\mathbf{G}_\kappa^\star &= \alpha_\kappa \sqrt{P_\kappa} \hat{\mathbf{h}}_\kappa^H {\mathbf{D}_\kappa}^H \cdot \\
&\cdot \left(\alpha_\kappa^2 P_\kappa \left(\mathbf{D}_\kappa \hat{\mathbf{h}}_\kappa \hat{\mathbf{h}}_\kappa^H {\mathbf{D}_\kappa}^H + \mathbf{Q}_\kappa \right) + \sum_{k \neq \kappa}^{K} \alpha_k^2 P_k \mathbf{C}_k + \sigma_d^2 \mathbf{I} \right)^{-1} .
\end{aligned} \tag{4.13}$$

The proof is in the Appendix of the present chapter.

4.4 Determining the Mean Squared Error of the Received Data Symbols with Optimal $\mathbf{G}^\star$

In the case of proper antenna spacing, the channel covariance matrices can be modeled as $\mathbf{C}_\kappa = c_\kappa \mathbf{I}$, which for $k = \kappa$ implies $\mathbf{D}_\kappa = d_\kappa \mathbf{I}$, $\mathbf{Q}_\kappa = q_\kappa \mathbf{I}$. In this case, the MSE as a function of the estimated channel can be obtained as follows.

Lemma 4.4.1 *In the case of uncorrelated antennas at the* BS*, when the BS employs the optimal receiver* $\mathbf{G}_\kappa^\star$*, the mean square error of the received data symbols can be expressed as:*

$$\text{MSE}\left(\hat{\mathbf{h}}_\kappa\right) = -2\alpha_\kappa\sqrt{P_\kappa}g_\kappa d_\kappa||\hat{\mathbf{h}}_\kappa||^2 + 1 + \\ g_\kappa^2\cdot\left(\alpha_\kappa^2 P_\kappa d_\kappa^2||\hat{\mathbf{h}}_\kappa||^4 + \left(\alpha_\kappa^2 P_\kappa q_\kappa + \sum_{k\neq\kappa}^{K}\alpha_k^2 P_k c_k + \sigma_d^2\right)||\hat{\mathbf{h}}_\kappa||^2\right)$$

where

$$g_\kappa \triangleq \frac{\alpha_\kappa\sqrt{P_\kappa}d_\kappa}{\alpha_\kappa^2 P_\kappa\left(d_\kappa^2||\hat{\mathbf{h}}_\kappa||^2 + q_\kappa\right) + \sum_{k\neq\kappa}^{K}\alpha_k^2 P_k c_k + \sigma_d^2}. \tag{4.14}$$

I can now derive the unconditional MSE from $\text{MSE} = \mathbb{E}_{\hat{\mathbf{h}}_\kappa}\text{MSE}\left(\hat{\mathbf{h}}_\kappa\right)$ based on the distribution of $\hat{\mathbf{h}}_\kappa$ which I recall from (4.3) as $\hat{\mathbf{h}}_\kappa \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}_\kappa\right)$.

4.5 Calculating the Unconditional Mean Squared Error

To calculate the unconditional MSE, notice that the $\text{MSE}(\hat{\mathbf{h}}_\kappa)$ depends on $\hat{\mathbf{h}}_\kappa$ only through $||\hat{\mathbf{h}}_\kappa||^2$. Thus, I can conveniently introduce $Y_\kappa \triangleq ||\hat{\mathbf{h}}_\kappa||^2$, substitute g_κ into (4.14) and, by inspecting (4.14), introduce the following notations:

$$T_1 \triangleq g_\kappa^2\left(\alpha_\kappa^2 P_\kappa d_\kappa^2||\hat{\mathbf{h}}_\kappa||^4\right) = s_\kappa\frac{s_\kappa Y_\kappa^2}{(b_\kappa + s_\kappa Y_\kappa)^2}, \tag{4.15}$$

where I introduced the notation $s_\kappa \triangleq d_\kappa^2 p_\kappa$, $p_\kappa \triangleq \alpha_\kappa^2 P_\kappa$, $\sigma_\kappa^2 \triangleq \sum_{k\neq\kappa}^{K}\alpha_\kappa^2 P_\kappa c_\kappa + \sigma_d^2$ and $Y_\kappa \triangleq ||\hat{\mathbf{h}}_\kappa||^2$ and $b_\kappa \triangleq q_\kappa p_\kappa + \sigma_\kappa^2$. Similarly:

$$T_2 \triangleq g_\kappa^2\left(\alpha_\kappa^2 P_\kappa q_\kappa + \sum_{k\neq\kappa}^{K}\alpha_k^2 P_k c_k + \sigma_d^2\right)||\hat{\mathbf{h}}_\kappa||^2 = \\ = b_\kappa\frac{s_\kappa Y_\kappa}{(b_\kappa + s_\kappa Y_\kappa)^2}, \tag{4.16}$$

$$T_3 \triangleq 2d_\kappa\alpha_\kappa\sqrt{P_\kappa}||\hat{\mathbf{h}}_\kappa||^2\cdot g_\kappa = \\ = 2\frac{d_\kappa^2 p_\kappa Y_\kappa}{\sigma_\kappa^2 + p_\kappa\left(d_\kappa^2 Y_\kappa + q_\kappa\right)} = 2\frac{s_\kappa Y_\kappa}{b_\kappa + s_\kappa Y_\kappa}. \tag{4.17}$$

I can now prove the following proposition, which will serve as the basis for numerical evaluations.

Theorem 4.5.1 *The unconditional* MSE *of the received data symbols of User-κ when the BS uses the optimal* $\mathbf{G}_\kappa^\star$ *receiver is as follows.*

$$\begin{aligned}\text{MSE} = & \\ s_\kappa \cdot \frac{N_r\left(-s_\kappa r + e^{\frac{b_\kappa}{s_\kappa r}}\left(b_\kappa + (1+N_r)s_\kappa r\right)E_{in}\left(1+N_r, \frac{b_\kappa}{s_\kappa r}\right)\right)}{s_\kappa^2 r} + & \\ + b_\kappa \cdot \frac{-s_\kappa r + e^{\frac{b_\kappa}{s_\kappa r}}\left(b_\kappa + N_r s_\kappa r\right)E_{in}\left(N_r, \frac{b_\kappa}{s_\kappa r}\right)}{s_\kappa^2 r^2} - & \\ -2 \cdot e^{\frac{b_\kappa}{s_\kappa r}} N_r E_{in}\left(1+N_r, \frac{b_\kappa}{s_\kappa r}\right) + 1, & \end{aligned} \tag{4.18}$$

where $E_{in}(n,z) \triangleq \int_1^\infty e^{-zt}/t^n \, dt$ *is a standard exponential integral function.*

The proof is in the Appendix of this chapter.

4.6 Numerical Results and Concluding Remarks

Table 4.1 System Parameters

Parameter	Value
Number of antennas	$N_r = 2, 4, 8, 10, 20, 50, 100, 500$
Path Loss of tagged User-κ	$\alpha = 40, 45, 50$ dB
Number of pilot and data symbols	$\tau_p = 1;\ \tau_d = 11$
Power budget	$\tau_p P_p + \tau_d P = P_{tot}$ =250 mW.

In this section I consider a single cell single user MIMO system, in which the mobile terminal is equipped with a single transmit antenna, whereas the BS employs N_r receive antennas. Note that the performance characteristics of the proposed MMSE receiver as compared with the naïve receiver are similar in the multi-user MIMO case from the perspective of the tagged user, since the proposed receiver treats the multi-user interference as noise according to (4.13). The key input parameters to this system that are necessary to obtain numerical results using the MSE derivation in this chapter (ultimately relying on Theorem 4.5.1) are listed in Table 4.1.

Figure 4.1 compares the performance of the system in which the number of antennas at the BS grows large ($N_r = 500$). As expected, given a fix sum power budget of $\tau_p P_p + \tau_d P = P_{tot}$ =250 mW, the optimal pilot-data power allocation becomes non trivial as it depends on the number of antennas, path loss and the employed receiver structure. The minimum value of the MSE in all cases are marked with a dot, which clearly indicate that the achievable minimum MSE with this power budget is significantly lower when employing the MMSE receiver.

Figure 4.2 shows the achievable minimum MSE value and the optimal pilot power setting as the function of the number of antennas at the base station. First, notice that the gain in terms of achievable minimum MSE increases as the number of antennas increases.

For example, at $N_r = 500$ the gain is around 6 dB. Interestingly, the pilot power setting that minimizes the MSE does not depend on the number of antennas when using the MMSE receiver, whereas it increases with the number of antennas in the case of the naïve receiver. The intuitive explanation for this is that in the case of uncorrelated antennas, according to equation (4.3), the diagonal elements of the covariance of the CSI error does not depend on the number of antennas, although the size of the matrix does. Thus, the pilot-data ration when using the MMSE receiver does not depend on the number of antennas, as opposed to the naïve receiver case, which does not minimize the MSE. The formal proof of this phenomenon will be presented in Chapter 7.

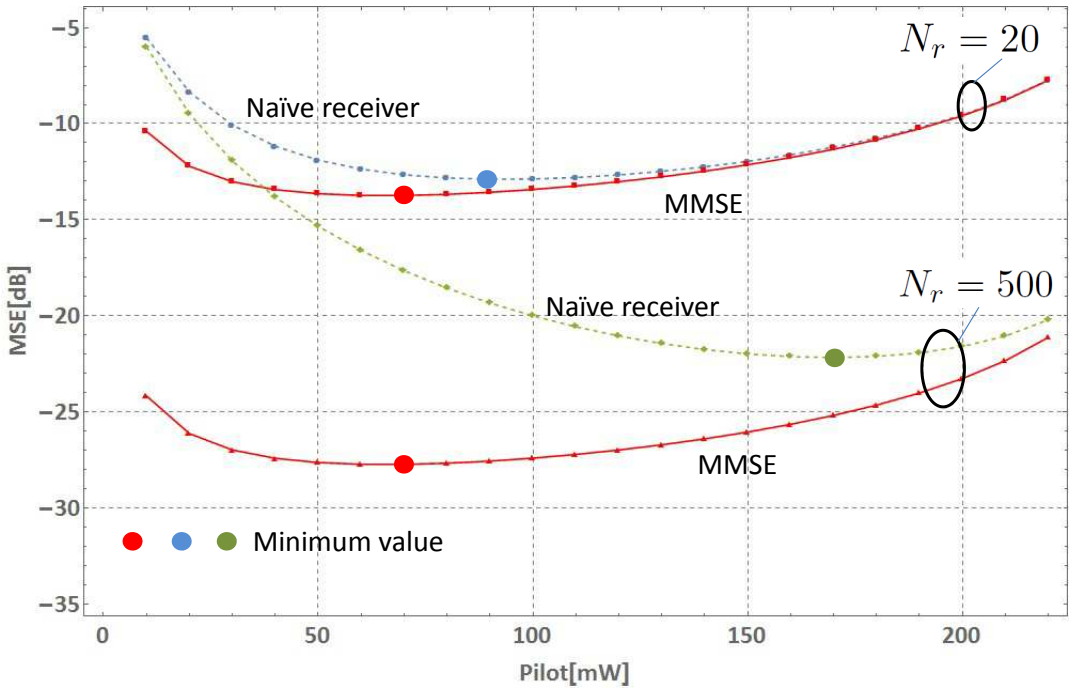

Fig. 4.1 MSE as the function of the pilot power P_p assuming a fixed pilot+data power budget with $N_r = 20$ and $N_r = 500$ number of antennas when using the naïve receiver and the MMSE receiver.

Appendix of Chapter 4

Appendix I: Proof of Result 4.3.1

From (4.10) it follows, that focusing on the tagged User-κ:

$$\begin{aligned}&\mathrm{MSE}\left(\mathbf{G}_\kappa, \mathbf{h}_\kappa\right) = \mathbb{E}_{\mathbf{h}_1,\ldots,\mathbf{h}_{\kappa-1},\mathbf{h}_{\kappa+1},\ldots,\mathbf{h}_K} \mathrm{MSE}\left(\mathbf{G}_\kappa, \mathbf{h}_1, \ldots, \mathbf{h}_K\right) = \\ &\left|\mathbf{G}_\kappa \alpha_\kappa \mathbf{h}_\kappa \sqrt{P_\kappa} - 1\right|^2 + \sum_{k,k\neq\kappa} \alpha_k^2 P_k \mathbb{E}_{\mathbf{h}_k} |\mathbf{G}_\kappa \mathbf{h}_k|^2 + \sigma_d^2 \mathbf{G}_\kappa \mathbf{G}_\kappa^H. \end{aligned} \tag{4.19}$$

Recognizing that [42]:

$$\left|\mathbf{G}_\kappa \mathbf{h}_\kappa \alpha \sqrt{P} - 1\right|^2 = \alpha^2 P \mathbf{G}_\kappa \mathbf{h}_\kappa \mathbf{h}_\kappa^H \mathbf{G}_\kappa^H - \alpha \sqrt{P} (\mathbf{G}_\kappa \mathbf{h}_\kappa + \mathbf{h}_\kappa^H \mathbf{G}_\kappa^H) + 1,$$

and $\mathbb{E}_{\mathbf{h}_k} |\mathbf{G}_\kappa \mathbf{h}_k|^2 = \mathbf{G}_\kappa \mathbb{E}_{\mathbf{h}_k} |\mathbf{h}_k|^2 \mathbf{G}_\kappa^H = \mathbf{G}_\kappa \mathbf{C}_k \mathbf{G}_\kappa^H$, the result follows.

Appendix II: Proof of Result 4.3.2

Utilizing $(\mathbf{h}_\kappa \mid \hat{\mathbf{h}}_\kappa) \sim \mathbf{D}_\kappa \hat{\mathbf{h}}_\kappa + \mathcal{CN}\left(\mathbf{0}, \mathbf{Q}_\kappa\right)$, where $\mathbf{D}_\kappa = \mathbf{C}_\kappa \mathbf{R}_\kappa^{-1}$, $\mathbf{R}_\kappa = \mathbf{C}_\kappa + \mathbf{C}_\kappa^{\mathbf{w}}$ and $\mathbf{Q}_\kappa = \mathbf{C}_\kappa - \mathbf{C}_\kappa \mathbf{R}_\kappa^{-1} \mathbf{C}_\kappa$, and, by averaging over $\mathbf{h}_\kappa | \hat{\mathbf{h}}_\kappa$, and following the technique proposed in [42], the result follows.

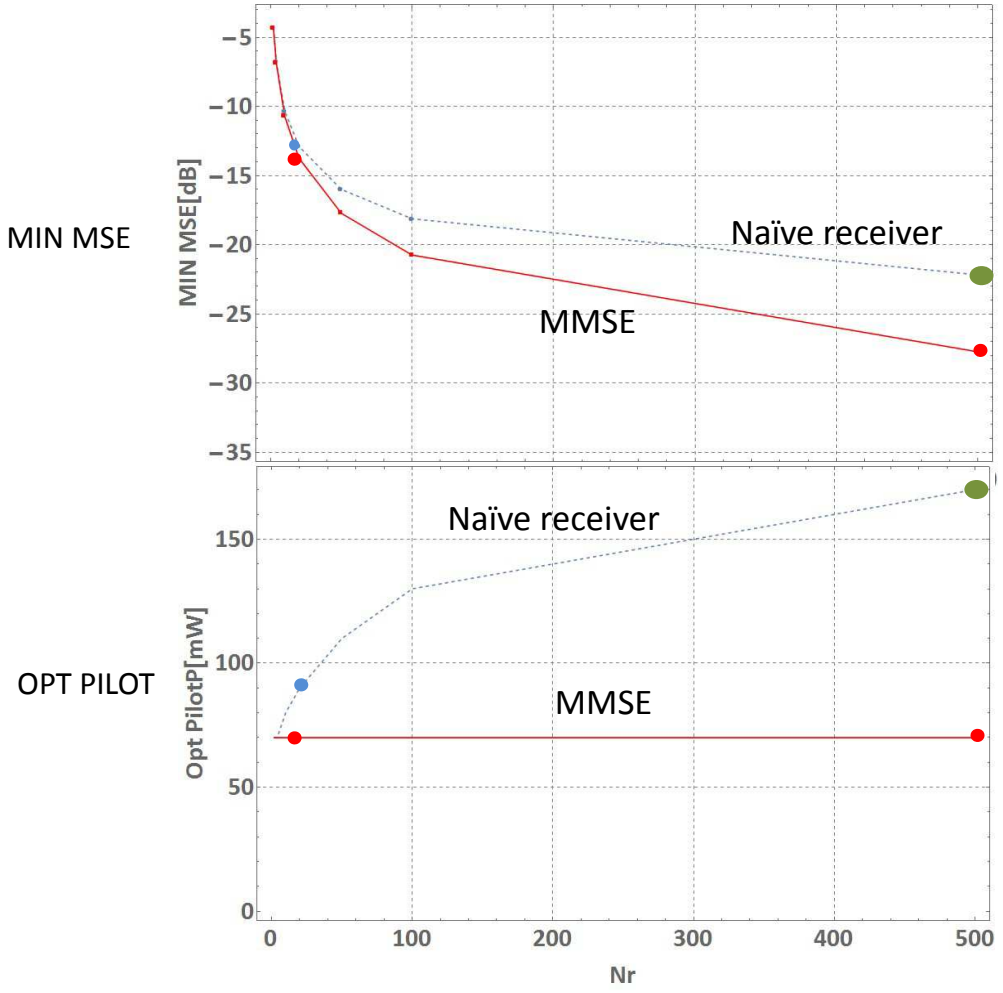

Fig. 4.2 The achievable minimum MSE and the optimum pilot power as the function of the number of the base station antennas when employing the naïve receiver and the MMSE receiver. The dots in the figure correspond to the case of $N_r = 20$ and $N_r = 500$ antennas.

Appendix III: Proof of Theorem 4.3.3

To derive the optimal $\mathbf{G}_\kappa$, I rewrite $\mathrm{MSE}\left(\mathbf{G}_\kappa, \hat{\mathbf{h}}_\kappa\right)$ in quadratic form of $(\mathbf{x}\mathbf{A}\mathbf{x}^H - \mathbf{x}\mathbf{B} - \mathbf{B}^H\mathbf{x}^H + 1)$:

$$\mathrm{MSE}\left(\mathbf{G}_\kappa, \hat{\mathbf{h}}_\kappa\right) = -\underbrace{\mathbf{G}_\kappa}_{\mathbf{x}}\underbrace{\alpha_\kappa\sqrt{P_\kappa}\mathbf{D}_\kappa\hat{\mathbf{h}}_\kappa}_{\mathbf{B}} - \alpha_\kappa\sqrt{P_\kappa}\hat{\mathbf{h}}_\kappa^H\mathbf{D}_\kappa^H\mathbf{G}_\kappa^H + 1 + \\ + \underbrace{\mathbf{G}_\kappa\left(\alpha_\kappa^2 P_\kappa\left(\mathbf{D}_\kappa\hat{\mathbf{h}}_\kappa\hat{\mathbf{h}}_\kappa^H\mathbf{D}_\kappa{}^H + \mathbf{Q}_\kappa\right) + \sum_{k\neq\kappa}^{K}\alpha_k^2 P_k\mathbf{C}_k + \sigma_d^2\mathbf{I}\right)}_{\mathbf{A}}\mathbf{G}_\kappa^H \tag{4.20}$$

Based on this quadratic form, the optimal receiver ($\mathbf{x}^\star = \mathbf{B}^H\mathbf{A}^{-1}$) is as in (4.13).

Appendix IV: Proof of Lemma 4.4.1

If $\mathbf{C}_\kappa = c_\kappa \mathbf{I}$, implying $\mathbf{D}_\kappa = d_\kappa \mathbf{I}$, $\mathbf{Q}_\kappa = q_\kappa \mathbf{I}$ and the optimal $\mathbf{G}_\kappa^\star$ can be written as:

$$\begin{aligned}\mathbf{G}_\kappa^\star &= \frac{\alpha_\kappa \sqrt{P_\kappa} d_\kappa}{\alpha_\kappa^2 P_\kappa \left(d_\kappa^2 ||\hat{\mathbf{h}}_\kappa||^2 + q_\kappa\right) + \sum_{k\neq\kappa}^K \alpha_k^2 P_k c_k + \sigma_d^2} \ \hat{\mathbf{h}}_\kappa^H \\ &\triangleq g_\kappa \cdot \hat{\mathbf{h}}_\kappa^H. \end{aligned} \tag{4.21}$$

Substituting $\mathbf{G}_\kappa^\star$ into the MSE of Result 4.3.2 gives the lemma.

Appendix V: Proof of Theorem 4.5.1

Recognizing that Y_κ is Gamma distributed, the density function of $Y_\kappa \forall \kappa$ is given by (dropping the index κ for convenience):

$$f_Y(x) = \frac{r^{-N_r} x^{N_r - 1} e^{-x/r}}{(N_r - 1)!} \quad x > 0. \tag{4.22}$$

Theorem (4.5.1) follows from Lemma (4.4.1) taking the average of MSE $\left(\hat{\mathbf{h}}_\kappa\right)$ using the the following integrals:

$$\begin{aligned}&\int_{x=0}^{\infty} T_1 f_{Y_\kappa}(x) dx = \\ &s_\kappa \cdot \frac{N_r \left(- s_\kappa r + e^{\frac{b_\kappa}{s_\kappa r}} \left(b_\kappa + (1+N_r) s_\kappa r\right) E_{in}\left(1 + N_r, \frac{b_\kappa}{s_\kappa r}\right)\right)}{s_\kappa^2 r};\end{aligned} \tag{4.23}$$

$$\int_{x=0}^{\infty} T_2 f_{Y_\kappa}(x) dx = b_\kappa \cdot \frac{-s_\kappa r + e^{\frac{b_\kappa}{s_\kappa r}} (b_\kappa + N_r s_\kappa r) E_{in}\left(N_r, \frac{b_\kappa}{s_\kappa r}\right)}{s_\kappa^2 r^2}; \tag{4.24}$$

$$\int_{x=0}^{\infty} T_3 f_{Y_\kappa}(x) dx = 2 \cdot e^{\frac{b_\kappa}{s_\kappa r}} N_r E_{in}\left(1 + N_r, \frac{b_\kappa}{s_\kappa r}\right). \tag{4.25}$$

where $E_{in}(n, z) \triangleq \int_1^\infty e^{-zt}/t^n \, dt$ is a standard exponential integral function.

Chapter 5
The Impact of Antenna Correlation on the Pilot-to-Data Power Ratio

5.1 Introduction

The preceding chapters investigated the optimal PDPR in the case of uncorrelated antennas giving rise to a diagonal covariance matrix. An isolated cell without modeling antenna correlation was also considered in [31, 42], while the multi-cell case was studied in [43], where it was found that the so called distributed iterative channel inversion (DICI) algorithm originally proposed by [44] can be advantageously extended taking into account the pilot-data power trade off. However, none of the aforementioned works captures the impact of antenna correlation on the performance of SIMO systems.

In this chapter, I turn my attention to a SIMO system in which the MS balances its PDPR, while the base station uses LS or MMSE channel estimation to initialize a linear MMSE equalizer. The specific contributions of this chapter to the line of related works are the derivations of a closed form for the MSE of the equalized data symbols for arbitrary correlation structure between the antennas by allowing any covariance matrix of the uplink channel. Similarly to the preceding chapter, this more general form is powerful, because it considers not only the pilot and data transmit power levels and the number of receive antennas at the base station (N_r) as independent variables, but it also explicitly takes into account antenna spacing and the statistics of the AoAs, including the angular spread as a parameter. For example, this methodology enables me to study the impact of the PDPR on the UL performance for the popular 3GPP spatial channel model (SCM) often used to model the wireless channels in cellular systems. The closed form formula takes into account the impact of N_r, AoA and angular spread on the MSE and thereby on the PDPR that minimizes the MSE. To the best of my knowledge the analytical result as well as the insights obtained in the numerical section of this chapter are novel.

The system model is defined in Section 5.2. In this section, for the sake of completeness and readability, I restate and reuse some results of [42]. Next, Section 5.3 describes the channel estimation models for least square and minimum mean square error channel estimators. Section 5.4 is concerned with deriving the conditional mean square error of the uplink equalized data symbols using either of the channel estimation techniques and assuming MMSE equalization. Based on the results of this section, the unconditional MSE with arbitrary channel covariance matrix is determined in Section 5.5. Numerical results are studied in Section 5.6. Section 5.7 concludes this chapter.

5.2 System Model

I consider the uplink transmission of a multi-antenna single cell wireless system, in which users are scheduled on orthogonal frequency channels. It is assumed that each mobile station (MS) employs an orthogonal pilot sequence, so that no interference between pilots is present in the system. This is a common assumption in massive MU-MIMO systems in which a single MS may have a single antenna. The BS estimates the channel $\mathbf{h}$ (column vector of dimension N_r, where N_r is the number of receive antennas at the BS) by either LS or MMSE channel estimators to initialize an MMSE equalizer for uplink data reception. Since I assume orthogonal pilot sequences, the channel estimation process can be assumed independent for each MS. I consider a time-frequency resource of T time slots in the channel coherence time, and F subcarriers in the coherence bandwidth, with a total number of symbols

$\tau_p + \tau_d = F \cdot T$, where I denote by τ_p the number of symbols allocated to pilot, and by τ_d the number of data symbols allocated to data ($\tau_p + \tau_d = \tau$). Moreover, I consider a transmission power level P_p and P for each pilot and data symbol, respectively. With this setup, I consider two pilot symbol allocation methods, namely block type and comb type, which will be discussed in the following subsections.

5.2.1 Block Type Pilot Allocation

The block type pilot arrangement consists of allocating one or more time slots for pilot transmission, by using all subcarriers in those time slots. This approach is a suitable strategy for slow time-varying channels. Given T slots, a fraction of T_p *slots* are allocated to the pilot and $T_d = T - T_p$ slots are allocated to the data symbols. Note that a maximum transmission power P_{tot} is allowed in each time slot, among all F subcarriers. This power constraint is then identical for both the pilot (P_p) and data power (P), i.e.,

$$FP_p \leq P_{tot} \quad FP \leq P_{tot}. \tag{5.1}$$

The power cannot be traded between pilot and data, but the *energy* budget can be distributed by tuning the number of time slots T_p and T_d, i.e., $\tau_p = FT_p$ and $\tau_d = FT_d$.

5.2.2 Comb Type Pilot Allocation

In the comb type pilot arrangement a certain number of subcarriers are allocated to pilot symbols, continuously in time. This approach is a suitable strategy for non-frequency selective channels. Given F subcarriers in the coherence bandwidth, a fraction of F_p subcarriers are allocated to the pilot and $F_d = F - F_p$ subcarriers are allocated to the data symbols.

Each MS transmits at a constant power P_{tot}, however, the transmission power can be distributed unequally in each subcarrier. In particular, considering a transmitted power P_p for each pilot symbol and P for each data symbol transmission, the following constraint is enforced:

$$\frac{\tau_p}{T} P_p + \frac{\tau_d}{T} P = P_{tot} \tag{5.2}$$

The total number of symbols for pilot is $\tau_p = TF_p$ and for data is $\tau_d = TF_d$. However, with comb type pilot arrangement, the trade off between pilot and data signals includes the trade-offs between the number of frequency channels *and* between the transmit power levels, which is an additional degree of freedom compared with the block type arrangement. With fixed (given or standardized) τ_p and τ_d the engineering freedom includes the tuning of the P_p and P power levels, which is the topic of the present chapter.

5.3 Channel Estimation

Let us consider a MS that transmits an orthogonal pilot sequence $\mathbf{s} = [s_1, ..., s_{\tau_p}]^T$, where each symbol is scaled as $|s_i|^2 = 1$, for $i = 1, .., \tau_p$. Thus, the $N_r \times \tau_p$ matrix of the received pilot signal at the BS from the MS is

$$\mathbf{Y}^p = \alpha \sqrt{P_p} \mathbf{h} \mathbf{s}^T + \mathbf{N}, \tag{5.3}$$

where I assume that $\mathbf{h}$ is a circular symmetric complex normal distributed vector with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{C}$ (of size N_r), denoted as $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, α accounts for the propagation loss, $\mathbf{N} \in \mathbb{C}^{N_r \times \tau_p}$ is the spatially and temporally additive white Gaussian noise (AWGN) with element-wise variance σ^2.

In this chapter, I consider two techniques, i.e., the LS and the MMSE channel estimation that are detailed in the following subsections.

5.3.1 Least Square Estimation

Conventional LS estimation relies on correlating the received signal with the known pilot sequence. The BS estimates the channel based on (5.3) assuming

$$\hat{\mathbf{h}}_{LS} = \mathbf{h} + \tilde{\mathbf{h}}_{LS} = \frac{1}{\alpha\sqrt{P_p}}\mathbf{Y}^p\mathbf{s}^*(\mathbf{s}^T\mathbf{s}^*)^{-1} = \mathbf{h} + \frac{1}{\alpha\sqrt{P_p}\tau_p}\mathbf{N}\mathbf{s}^*. \tag{5.4}$$

Note that $\mathbf{N}\mathbf{s}^* = \left[\sum_{i=1}^{\tau_p} s_i^* n_{i,1}, ..., \sum_{i=1}^{\tau_p} s_i^* n_{i,N_r}\right]^T$, then $\mathbf{N}\mathbf{s}^* \sim \mathcal{CN}(\mathbf{0}, \tau_p\sigma^2\mathbf{I}_{N_r})$.

By considering $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, it follows that the estimated channel $\hat{\mathbf{h}}_{LS}$ is a circular symmetric complex normal distributed vector $\hat{\mathbf{h}}_{LS} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{LS})$, with

$$\mathbf{R}_{LS} = \mathbb{E}\{\hat{\mathbf{h}}_{LS}\hat{\mathbf{h}}_{LS}^H\} = \mathbf{C} + \frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r}. \tag{5.5}$$

The channel estimation error is defined as $\tilde{\mathbf{h}}_{LS} = \mathbf{h} - \hat{\mathbf{h}}_{LS}$, so that $\tilde{\mathbf{h}}_{LS} \sim \mathcal{CN}(\mathbf{0}, \mathbf{W}_{LS})$ with

$$\mathbf{W}_{LS} = \frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r}$$

and the estimation mean square error (MSE) is derived as

$$\varepsilon_{LS} = \mathbb{E}\{||\tilde{\mathbf{h}}_{LS}||_F^2\} = \mathrm{tr}\{\mathbf{W}_{LS}\} = \frac{N_r\sigma^2}{\alpha^2 P_p \tau_p}, \tag{5.6}$$

where $||\cdot||_F^2$ is the Frobenius norm.

5.3.2 Minimum Mean Squared Error Estimation

In this case I find it convenient to define a training matrix $\mathbf{S} = \mathbf{s} \otimes \mathbf{I}_{N_r}$ (of size $\tau_p N_r \times N_r$), so that $\mathbf{S}^H\mathbf{S} = \tau_p\mathbf{I}_{N_r}$. The $\tau_p N_r \times 1$ vector of received signal (5.3) can be conveniently rewritten as

$$\tilde{\mathbf{Y}}^p = \alpha\sqrt{P_p}\mathbf{S}\mathbf{h} + \tilde{\mathbf{N}}. \tag{5.7}$$

where $\tilde{\mathbf{Y}}^p, \tilde{\mathbf{N}} \in \mathbb{C}^{\tau_p N_r \times 1}$.
The MMSE equalizer aims at minimizing the MSE between the estimate $\hat{\mathbf{h}}_{MMSE} = \mathbf{H}\tilde{\mathbf{Y}}^p$ and the channel $\mathbf{h}$. More precisely,

$$\begin{aligned}\mathbf{H} =& \arg\min_{\mathbf{H}} \mathbb{E}\{||\mathbf{H}\tilde{\mathbf{Y}}^p - \mathbf{h}||_F^2\} \\ =& \alpha\sqrt{P_p}(\sigma^2\mathbf{I}_{N_r} + \alpha^2 P_p \mathbf{C}\mathbf{S}^H\mathbf{S})^{-1}\mathbf{C}\mathbf{S}^H; \quad \mathbf{H} \in \mathbb{C}^{N_r \times \tau_p N_r}.\end{aligned}$$

The MMSE estimate is then expressed as

$$\begin{aligned}\hat{\mathbf{h}}_{\text{MMSE}} &= \alpha\sqrt{P_p}(\sigma^2\mathbf{I}_{N_r} + \alpha^2 P_p \tau_p \mathbf{C})^{-1}\mathbf{C}\mathbf{S}^H(\alpha\sqrt{P_p}\mathbf{S}\mathbf{h} + \tilde{\mathbf{N}}) \\ &= \left(\frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r} + \mathbf{C}\right)^{-1}\mathbf{C}\left(\mathbf{h} + \frac{1}{\alpha\sqrt{P_p}\tau_p}\mathbf{S}^H\tilde{\mathbf{N}}\right). \end{aligned} \tag{5.8}$$

Notice that $\mathbf{S}^H\tilde{\mathbf{N}} \sim \mathcal{CN}(\mathbf{0}, \tau_p\sigma^2\mathbf{I}_{N_r})$, therefore the estimated channel $\hat{\mathbf{h}}_{\text{MMSE}}$ is also a circular symmetric complex normal distributed vector $\hat{\mathbf{h}}_{\text{MMSE}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{\text{MMSE}})$, that is

$$\hat{\mathbf{h}}_{\text{MMSE}} = \mathbf{h} + \tilde{\mathbf{h}}_{\text{MMSE}}, \tag{5.9}$$

and

$$\mathbf{R}_{\text{MMSE}} = \mathbf{C}^2\left(\frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r} + \mathbf{C}\right)^{-1}, \tag{5.10}$$

where I considered $\mathbf{C} = \mathbf{C}^H$ and applied the commutativity of $\mathbf{C}$ and $\mathbf{I}_{N_r}$ to substitute

$$\left(\frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r} + \mathbf{C}\right)^{-1}\mathbf{C} = \mathbf{C}\left(\frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r} + \mathbf{C}\right)^{-1}.$$

The channel estimation error is $\tilde{\mathbf{h}}_{\text{MMSE}} = \mathbf{h} - \hat{\mathbf{h}}_{\text{MMSE}}$ so that $\tilde{\mathbf{h}}_{\text{MMSE}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{W}_{\text{MMSE}})$ with

$$\mathbf{W}_{\text{MMSE}} = \mathbf{C}\left(\mathbf{I}_{N_r} + \frac{\alpha^2 P_p \tau_p}{\sigma^2}\mathbf{C}\right)^{-1}$$

and the estimation MSE simply follows as

$$\varepsilon_{\text{MMSE}} = \text{tr}\left\{\mathbf{C}\left(\mathbf{I}_{N_r} + \frac{\alpha^2 P_p \tau_p}{\sigma^2}\mathbf{C}\right)^{-1}\right\}. \tag{5.11}$$

Notice that for both the LS and MMSE channel estimations, the estimation MSE is a monotonically decreasing function of the pilot energy per antenna $P_p\tau_p$. The quantity $P_p\tau_p$ can be regarded as the total pilot power (energy) budget, that - assuming a fixed τ_p - can be tuned by tuning P_p.

5.4 Determining the Conditional Mean Square Error

5.4.1 A Key Observation

Equations (5.4) and (5.9) imply that $\mathbf{h}$ and $\hat{\mathbf{h}}$ are jointly circular symmetric complex Gaussian (multivariate normal) distributed random variables [34], [35] . Specifically, recall from [34] that the covariance matrix of the joint PDF is composed by autocovariance matrices $\mathbf{C}_{\mathbf{h},\mathbf{h}}$, $\mathbf{C}_{\hat{\mathbf{h}},\hat{\mathbf{h}}}$ and cross covariance matrices $\mathbf{C}_{\mathbf{h},\hat{\mathbf{h}}}$, $\mathbf{C}_{\hat{\mathbf{h}},\mathbf{h}}$ as

$$\begin{bmatrix} \mathbf{C}_{\mathbf{h},\mathbf{h}} & \mathbf{C}_{\mathbf{h},\hat{\mathbf{h}}} \\ \mathbf{C}_{\hat{\mathbf{h}},\mathbf{h}} & \mathbf{C}_{\hat{\mathbf{h}},\hat{\mathbf{h}}} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{C} \\ \mathbf{C} & \mathbf{R} \end{bmatrix},$$

and

$$\mathbf{R} = \begin{cases} \mathbf{R}_{LS} & \text{for LS estimation,} \\ \mathbf{R}_{\text{MMSE}} & \text{for MMSE estimation.} \end{cases}$$

5.4.2 Determining the Conditional Channel Distribution

From the joint PDF of $\mathbf{h}$ and $\hat{\mathbf{h}}$ I can compute the following conditional distributions.

Result 5.4.1 *Given a random channel realization* $\mathbf{h}$*, the estimated channel* $\hat{\mathbf{h}}$ *conditioned to* $\mathbf{h}$ *can be shown to be distributed as*

$$(\hat{\mathbf{h}} \mid \mathbf{h}) \sim \mathbf{h} + C\mathcal{N}(\mathbf{0}, \mathbf{R} - \mathbf{C}). \tag{5.12}$$

Result 5.4.2 *The distribution of the channel realization* $\mathbf{h}$ *conditioned to the estimate* $\hat{\mathbf{h}}$ *is normally distributed as follows:*

$$(\mathbf{h} \mid \hat{\mathbf{h}}) \sim \mathbf{D}\hat{\mathbf{h}} + C\mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right), \tag{5.13}$$

where $\mathbf{D} = \mathbf{C}\mathbf{R}^{-1}$ *and* $\mathbf{Q} = \mathbf{C} - \mathbf{C}\mathbf{R}^{-1}\mathbf{C}$.

Both of these results can be easily verified by exploiting the basic characteristics of multivariate Gaussian random variables [34]. To capture the tradeoff between the pilot and data power, I need to calculate the mean square error of the equalized data symbols. To this end, I consider an equalization model in the next subsection.

5.4.3 Equalizer Model Based on Least Square or Minimum Mean Squared Error Channel Estimation

The data signal received by the BS is

$$\mathbf{y} = \alpha\sqrt{P}\mathbf{h}x + \mathbf{n}, \tag{5.14}$$

where $|x|^2 = 1$. I assume that the BS employs a naive MMSE equalizer, where the estimated channel (either $\hat{\mathbf{h}}_{LS}$ or $\hat{\mathbf{h}}_{\text{MMSE}}$) is taken as if it was the actual channel:

$$\mathbf{G} = \alpha\sqrt{P}\hat{\mathbf{h}}^H(\alpha^2 P\hat{\mathbf{h}}\hat{\mathbf{h}}^H + \sigma^2\mathbf{I})^{-1}. \tag{5.15}$$

Under this assumption, recall the following result from [31] as a first step towards determining the unconditional MSE.

Result 5.4.3 *Let* $MSE(\mathbf{h}, \hat{\mathbf{h}}) = \mathbb{E}_{x,\mathbf{n}}\left\{|\mathbf{G}\mathbf{y} - x|^2\right\}$ *be the MSE for the equalized symbols, given the realizations of* $\mathbf{h}$ *and* $\hat{\mathbf{h}}$*. It is*

$$\begin{aligned} MSE(\mathbf{h}, \hat{\mathbf{h}}) = {} & \alpha^2 P\mathbf{G}\mathbf{h}\mathbf{h}^H\mathbf{G}^H - \\ & - 2\alpha\sqrt{P}Re[\mathbf{G}\mathbf{h}] + \sigma^2\mathbf{G}\mathbf{G}^H + 1. \end{aligned} \tag{5.16}$$

From this, my next result follows directly.

Result 5.4.4 *Let* $MSE(\hat{\mathbf{h}}) = \mathbb{E}_{\mathbf{h}|\hat{\mathbf{h}}}\left\{MSE(\mathbf{h},\hat{\mathbf{h}})\right\}$ *be the MSE for each equalized data symbol, given the estimated channel realization* $\hat{\mathbf{h}}$*. It satisfies (see also [45]):*

$$MSE(\hat{\mathbf{h}}) = \mathbf{G}\Big(\alpha^2 P(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q}) + \sigma^2\mathbf{I}\Big)\mathbf{G}^H - \\ - 2\alpha\sqrt{P}Re\{\mathbf{G}\mathbf{D}\hat{\mathbf{h}}\} + 1. \tag{5.17}$$

5.5 Derivation of the Unconditional Mean Squared Error

Based on the conditional MSE expression of the preceding section, I am now interested in deriving the unconditional expectation of the MSE. To this end, the following two lemmas turn out to be useful.

Lemma 5.1. *Given a channel estimate instance* $\hat{\mathbf{h}}$*, the MMSE weighting matrix* $\mathbf{G}$*, as a function of the number of receive antennas at the base station (*N_r*) can be expressed as follows*

$$\mathbf{G} = \frac{\alpha\sqrt{P}}{||\hat{\mathbf{h}}||^2\alpha^2 P + \sigma^2}\,\hat{\mathbf{h}}^H, \tag{5.18}$$

where $||\hat{\mathbf{h}}||^2 = \hat{\mathbf{h}}^H\hat{\mathbf{h}} = \sum_{i=1}^{N_r}|\hat{h}_i|^2$.

The proof of the lemma is straightforward based on Chapter 3 of this dissertation [31]. Using $z = \frac{\alpha\sqrt{P}}{||\hat{\mathbf{h}}||^2\alpha^2 P+\sigma^2}$ and this simple expression of $\mathbf{G}$ I can express the conditional expectation of the MSE of the MMSE equalized data symbols as a function of the channel covariance, $\mathbf{C}$.

Lemma 5.2.

$$MSE(\hat{\mathbf{h}}) = z^2\alpha^2 P\,\hat{\mathbf{h}}^H\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H\hat{\mathbf{h}} + z^2\alpha^2 P\,\hat{\mathbf{h}}^H\mathbf{Q}\hat{\mathbf{h}} + z^2\sigma^2\hat{\mathbf{h}}^H\hat{\mathbf{h}} - 2z\alpha\sqrt{P}Re\{\hat{\mathbf{h}}^H\mathbf{D}\hat{\mathbf{h}}\} + 1.$$

where $z = \frac{\alpha\sqrt{P}}{||\hat{\mathbf{h}}||^2\alpha^2 P+\sigma^2}$ *is a function of* $||\hat{\mathbf{h}}||^2$.

The proof is in the Appendix.

5.5.1 Computing z

Computing z in Lemma 5.2 is essential in calculating the unconditional MSE. This can be done using the following lemma.

Lemma 5.3.

$$z = \frac{1}{\alpha\sqrt{P}}\cdot\int_{x=0}^{\infty} e^{-x(\hat{\mathbf{h}}^H\hat{\mathbf{h}}+\sigma^2/(\alpha^2 P))}dx$$

and

$$z^2 = \frac{1}{\alpha^2 P}\cdot\int_{x=0}^{\infty} xe^{-x(\hat{\mathbf{h}}^H\hat{\mathbf{h}}+\sigma^2/(\alpha^2 P))}dx.$$

The proof is in the appendix of this chapter.

5.5.2 Singular Value Decomposition

To take the expectation with respect to $\hat{\mathbf{h}}$ of the terms of MSE($\hat{\mathbf{h}}$) as in Lemma 5.2, I introduce the following decomposition. $\hat{\mathbf{h}}$ is $\mathcal{CN}(\mathbf{0},\mathbf{R})$ distributed with $\mathbf{R} = \mathbf{C} + \frac{\sigma^2}{P_p\alpha^2}\ \mathbf{I}$ and recall from Result 5.4.2 that $\mathbf{D} = \mathbf{C}\mathbf{R}^{-1}$. Let $\mathbf{C} = \mathbf{\Theta}^H\mathbf{S_C}\mathbf{\Theta}$ be the singular value decomposition of $\mathbf{C}$. Then $\mathbf{R} = \mathbf{\Theta}^H\mathbf{S_R}\mathbf{\Theta}$, $\mathbf{D} = \mathbf{\Theta}^H\mathbf{S_D}\mathbf{\Theta}$ and $\mathbf{Q} = \mathbf{\Theta}^H\mathbf{S_Q}\mathbf{\Theta}$ with $\mathbf{S_R} = \mathbf{S_C} + \frac{\sigma^2}{P_p\alpha^2}\ \mathbf{I}$, $\mathbf{S_D} = \mathbf{S_C}\mathbf{S_R}^{-1}$, and $\mathbf{S_Q} = \mathbf{S_C} - \mathbf{S_C}\mathbf{S_R}^{-1}\mathbf{S_C}$ where matrices $\mathbf{S}_\bullet$ are real non-negative diagonal matrices. Specifically, I will refer to the diagonal elements of $\mathbf{S_D}$ and $\mathbf{S_R}$ using the notations $d_k = \mathbf{S_D}_{kk}$ and $r_k = \mathbf{S_R}_{kk}$, respectively. Let $\mathbf{v} = \mathbf{\Theta}\hat{\mathbf{h}}$, then $\mathbf{v}$ is a random vector with distribution $\mathcal{CN}(\mathbf{0},\mathbf{S_R})$, since

$$\begin{aligned}\mathbb{E}(\mathbf{v}\mathbf{v}^H) = \mathbb{E}(\mathbf{\Theta}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{\Theta}^H) = \mathbf{\Theta}\mathbb{E}(\hat{\mathbf{h}}\hat{\mathbf{h}}^H)\mathbf{\Theta}^H = \\ = \mathbf{\Theta}\mathbf{R}\mathbf{\Theta}^H = \mathbf{\Theta}\mathbf{\Theta}^H\mathbf{S_R}\mathbf{\Theta}\mathbf{\Theta}^H = \mathbf{S_R}\ .\end{aligned}$$

That is, the elements of $\mathbf{v}$ are independent, but they have different variances.

5.5.3 Terms of the MSE = $\mathbb{E}_{\hat{\mathbf{h}}}\left\{MSE(\hat{\mathbf{h}})\right\}$

To compute $\mathbb{E}_{\hat{\mathbf{h}}}\left\{\text{MSE}(\hat{\mathbf{h}})\right\}$ based on Lemma 5.2, I need to calculate the following terms (T_1, T_2 and T_3):

$$\begin{aligned}\mathbb{E}_{\hat{\mathbf{h}}}\left\{\text{MSE}(\hat{\mathbf{h}})\right\} = &\underbrace{\alpha^2 P\ \mathbb{E}_{\hat{\mathbf{h}}}\left\{z^2\ \hat{\mathbf{h}}^H\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H\hat{\mathbf{h}}\right\}}_{\triangleq T_1} + \\ &+\underbrace{\mathbb{E}_{\hat{\mathbf{h}}}\left\{z^2\ \hat{\mathbf{h}}^H\left(\alpha^2 P\mathbf{Q}+\sigma^2\mathbf{I}\right)\hat{\mathbf{h}}\right\}}_{\triangleq T_2} + \\ &-\underbrace{2\alpha\sqrt{P}\ \mathbb{E}_{\hat{\mathbf{h}}}\left\{z\ Re\{\hat{\mathbf{h}}^H\mathbf{D}\hat{\mathbf{h}}\}\right\}}_{\triangleq T_3} + 1.\end{aligned}$$

With Lemmas 5.1, 5.2 and 5.3 in my hands, I can state the main result on calculating the terms T_1, T_2 and T_3.

Theorem 5.5.1 *The mean square error (MSE) of the uplink received data with arbitrary correlation matrix* $\mathbf{C}$ *of the uplink channel can be calculated as in the sum of the terms* T_1, T_2 *and* T_3 *plus 1, where*

$$T_1 = \sum_k \sum_{\ell,\ell\neq k} d_k d_\ell \cdot$$

$$\cdot \int_{x=0}^{\infty} x e^{-x\sigma^2/(\alpha^2 P)} \frac{1}{x+r_k}\ \frac{1}{x+r_\ell}\ \prod_i \frac{r_i}{x+r_i}\ dx +$$

$$+ \sum_k d_k^2 \int_{x=0}^{\infty} x e^{-x\sigma^2/(\alpha^2 P)}\ \frac{2}{(x+r_k)^2}\ \prod_i \frac{r_i}{x+r_i}\ dx;$$

$$T_2 = \frac{1}{\alpha^2 P}\sum_k m_k \int_{x=0}^{\infty} x e^{-x\frac{\sigma^2}{\alpha^2 P}} \frac{1}{x+r_k} \prod_i \frac{r_i}{x+r_i} dx;$$

and

$$T_3 = 2 \sum_k d_k \int_{x=0}^{\infty} e^{-x \frac{\sigma^2}{\alpha^2 P}} \frac{1}{x + r_k} \prod_i \frac{r_i}{x + r_i} dx,$$

where $\mathbf{S_M} = \alpha^2 P \mathbf{S_Q} + \sigma^2 \mathbf{I}$ *is a diagonal matrix with diagonal elements* $m_k = \mathbf{S_M}_{kk} = \alpha^2 P q_k + \sigma^2$*, where* $q_k = \mathbf{S_Q}_{kk}$.

The proof of Theorem 5.5.1 is in the Appendix of this chapter.

5.6 Numerical Results

5.6.1 Channel Model and Covariance Matrix

Table 5.1 System Parameters

Parameter	**Value**
Number of antennas	$N_r = 2, 4, 8, 10, 20$
Path Loss	$\alpha = 50, 55, 60$ dB
Power budget	$\tau_p P_p + \tau_d P = P_{tot}$ mW, as in Eq. (5.2).
Antenna spacing	$D/\lambda = 0.15, \ldots, 1.5$
Mean Angle of Arrival (AoA)	$\bar{\theta} = 70°$
Angular spread	$2 \cdot \theta_\Delta = 5, \ldots, 45°$

In this section I consider a single cell system, in which MSs use orthogonal pilots to facilitate the estimation of the uplink channel by the BS. Recall from Section 5.3 that the channel estimation process is independent for each MS and I can therefore focus on a single user. The covariance matrix $\mathbf{C}$ of the channel $\mathbf{h}$ as the function of the antenna spacing, mean angle of arrival and angular spread is modeled as by the well known spatial channel model, which is known to be accurate in non-line-of-sight environment with rich scattering and all antenna elements identically polarized, see [46]. For uniformly distributed angle of arrivals, the (m,n) element of the covariance matrix $\mathbf{C}$ are given by

$$\mathbf{C}_{m,n} = \frac{1}{2\theta_\Delta} \int_{-\theta_\Delta}^{\theta_\Delta} e^{j \cdot 2\pi \cdot \frac{D}{\lambda}(n-m)\cos(\bar{\theta}+x)} dx,$$

where the system parameters are given in Table 5.1. The covariance matrix $\mathbf{C}$ becomes practically diagonal as the antenna spacing and the angular spread grows beyond $D\lambda > 1$ and $\theta_\Delta > 30°$. In contrast, with critically spaced antennas $D\lambda = 0.5$ and $\theta_\Delta < 10°$, the antenna correlation in terms of the off-diagonal elements of $\mathbf{C}$ can be considered strong. A validation of the proposed channel model with data from a realistic channel simulator was included in [45].

5.6.2 Numerical Results

Recall that the MSE of the received data symbols according to Theorem 5.5.1 depends on the pilot through the product $t_p P_p$ (which I call the pilot budget) and the data power P for each transmitted data symbol. Figures 5.1 and 5.2 compare the MSE that can be achieved by a particular setting of the pilot and data power levels in the case of uncorrelated and highly correlated antennas (for the case with $N_r = 20$ receive antennas). The impact of high antenna correlation is that in order to reach the same MSE, a higher power level for both the pilot and data transmission must be used.

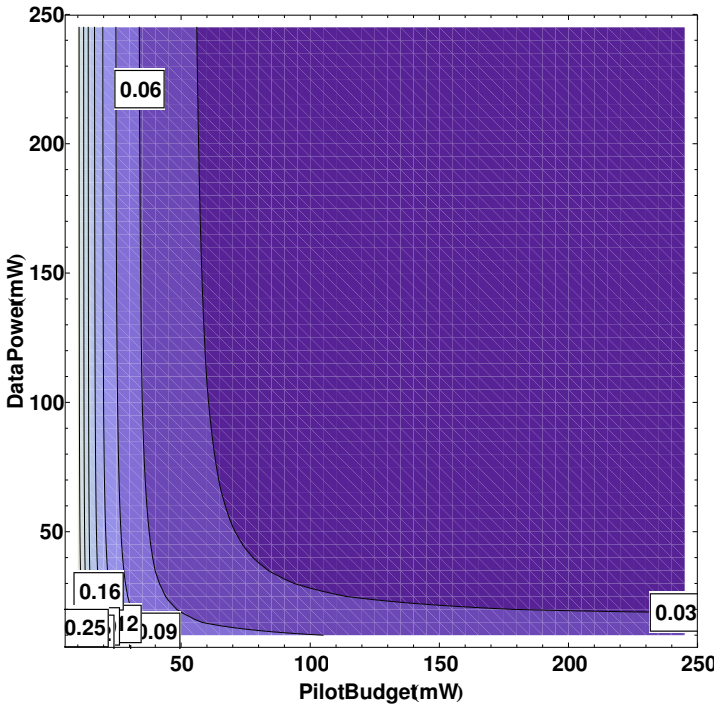

Fig. 5.1 Contour plot of the MSE achieved by specific pilot and data power settings of a SIMO system with $N_r = 20$ *uncorrelated* receiver antennas at a fix path loss position of 50 dB. For example an MSE value less than 0.03 can be reached by setting the pilot power budget to $\tau_p P_p \geq 70$ mW and the data power $P \geq 60$ mW, or by $P_p \geq 200$ mW and $P \geq 20$ mW. We can see that with a total power budget of 250 mW, and with proper pilot-data balancing, a minimum MSE that is clearly less than 0.03 can be reached.

Figures 5.3 and 5.4 compare the MSE that can be achieved by a particular setting of the data power levels in the case of uncorrelated and highly correlated antennas (for the case with $N_r = 20$ receive antennas) at different path loss positions. Although the impact of high antenna correlation is clearly high, the proper setting of the pilot-to-data-power ratio has a more pronounced effect.

5.7 Conclusions

This chapter developed a model of a single cell system, in which MSs use orthogonal pilots to facilitate uplink channel estimation by the BS. I developed a methodology to calculate the MSE of the uplink equalized data symbols and derived a closed form for the MSE as a function of not only the number of antennas, the pilot power and the data transmit power, but also the path loss and other parameters that determine the covariance matrix of the fast fading channel between the MS and the BS. With this methodology, through numerical examples, I found that although the impact of antenna correlation on the MSE performance can be significant, this impact can be compensated by setting the correct pilot-to-data power ratio (PDPR) in case of a total power budget. Furthermore, I found that as the number of antennas grows, a higher ratio of the power budget should be spent on pilots, virtually independently of the antenna correlation. This can be seen in line with the findings of massive MIMO systems that suggest that the data transmit power at the MS can be significantly lower as the number of antennas at the BS grows large.

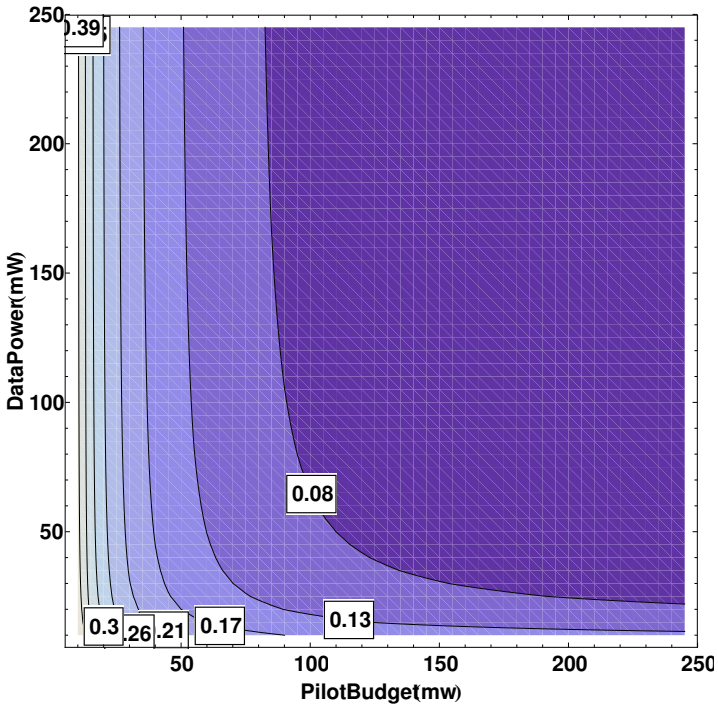

Fig. 5.2 Contour plot of the MSE achieved by specific pilot and data power settings of a SIMO system with $N_r = 20$ *correlated* receiver antennas at a fix path loss position of 50 dB. Compared with Figure 5.1, we can see that with similar sum power budget, the MSE value that can be reached is somewhat higher. For example, with a power budget of 250 mW, an MSE value that is less than 0.08 can be realized ($\tau_p P_p = 150$ mW and $P = 100$ mW).

Appendix of Chapter 5

Proof of Lemma 5.2

$$\begin{aligned}\mathrm{MSE}(\hat{\mathbf{h}}) = z\,\hat{\mathbf{h}}^H\left(\alpha^2 P(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q}) + \sigma^2\mathbf{I}\right)z\,\hat{\mathbf{h}} - \\ - 2\alpha\sqrt{P}Re\{z\,\hat{\mathbf{h}}^H\mathbf{D}\hat{\mathbf{h}}\} + 1 = \\ = z^2\alpha^2 P\,\hat{\mathbf{h}}^H\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H\hat{\mathbf{h}} + z^2\alpha^2 P\,\hat{\mathbf{h}}^H\mathbf{Q}\hat{\mathbf{h}} + \\ + z^2\sigma^2\hat{\mathbf{h}}^H\hat{\mathbf{h}} - 2z\alpha\sqrt{P}Re\{\hat{\mathbf{h}}^H\mathbf{D}\hat{\mathbf{h}}\} + 1.\end{aligned}$$

Proof or Lemma 1

According to the matrix inversion lemma for matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ of size $n\times n$, $n\times m$, $m\times m$, $m\times n$, respectively, I have

$$\begin{aligned}&(\mathbf{A}+\mathbf{BCD})^{-1} = \\ &\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1}\right)^{-1}\mathbf{DA}^{-1}\ .\end{aligned}$$

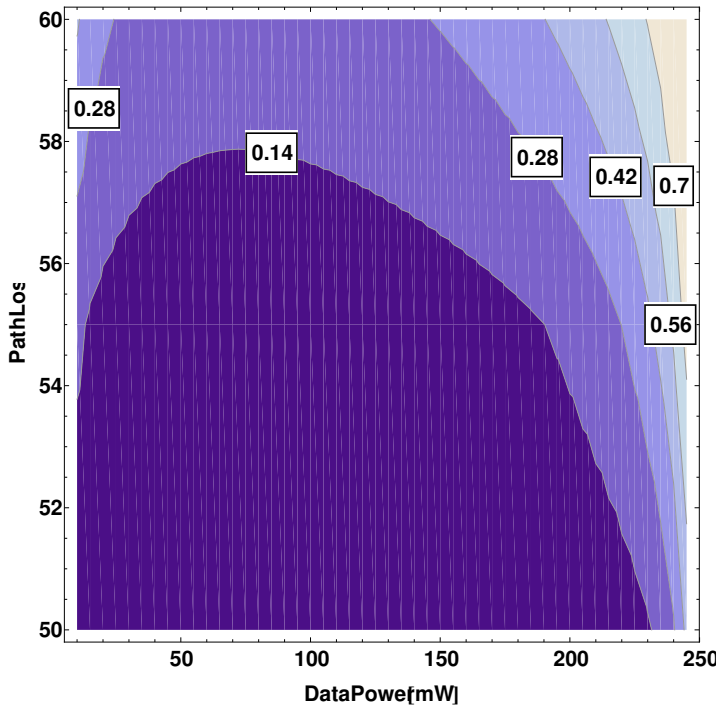

Fig. 5.3 Contour plot of the MSE as the function of the data power and the path loss under a total power (250 mW) constraint with $N_r = 20$ uncorrelated antennas. For example, with the near optimal data power setting and MSE value of 0.14 can be reached at 58 dB path loss.

Substituting $\mathbf{A} = \sigma^2\mathbf{I}$, $\mathbf{B} = \alpha\sqrt{P}\hat{\mathbf{h}}$, $\mathbf{C} = 1$, $\mathbf{D} = \alpha\sqrt{P}\hat{\mathbf{h}}^H$ I have

$$
\begin{aligned}
(\sigma^2\mathbf{I} + \alpha^2 P\hat{\mathbf{h}}\hat{\mathbf{h}}^H)^{-1} &= \frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^2}\mathbf{I}\alpha\sqrt{P}\hat{\mathbf{h}}\cdot \\
&\cdot\left(\alpha\sqrt{P}\hat{\mathbf{h}}^H\frac{1}{\sigma^2}\mathbf{I}\alpha\sqrt{P}\hat{\mathbf{h}} + 1\right)^{-1}\alpha\sqrt{P}\hat{\mathbf{h}}^H\frac{1}{\sigma^2}\mathbf{I} = \\
&= \frac{1}{\sigma^2}\mathbf{I} - \frac{\frac{\alpha^2 P}{\sigma^4}\hat{\mathbf{h}}\hat{\mathbf{h}}^H}{\frac{\alpha^2 P}{\sigma^2}\hat{\mathbf{h}}^H\hat{\mathbf{h}} + 1},
\end{aligned}
$$

where $\hat{\mathbf{h}}^H\hat{\mathbf{h}} = \|\hat{\mathbf{h}}\|^2$. Finally,

$$
\begin{aligned}
\mathbf{G} &= \alpha\sqrt{P}\hat{\mathbf{h}}^H(\alpha^2 P\hat{\mathbf{h}}\hat{\mathbf{h}}^H + \sigma^2\mathbf{I})^{-1} \\
&= \frac{\alpha\sqrt{P}}{\sigma^2}\hat{\mathbf{h}}^H - \frac{\alpha\sqrt{P}\frac{\alpha^2 P}{\sigma^4}\hat{\mathbf{h}}^H\hat{\mathbf{h}}}{\frac{\alpha^2 P}{\sigma^2}\|\hat{\mathbf{h}}\|^2 + 1}\hat{\mathbf{h}}^H \\
&= \frac{\alpha\sqrt{P}}{\sigma^2}\hat{\mathbf{h}}^H\left(1 - \frac{\alpha^2 P\|\hat{\mathbf{h}}\|^2}{\alpha^2 P\|\hat{\mathbf{h}}\|^2 + \sigma^2}\right) \\
&= \frac{\alpha\sqrt{P}}{\sigma^2}\hat{\mathbf{h}}^H\left(\frac{\sigma^2}{\alpha^2 P\|\hat{\mathbf{h}}\|^2 + \sigma^2}\right) \\
&= \frac{\alpha\sqrt{P}}{\alpha^2 P\|\hat{\mathbf{h}}\|^2 + \sigma^2}\hat{\mathbf{h}}^H.
\end{aligned}
$$

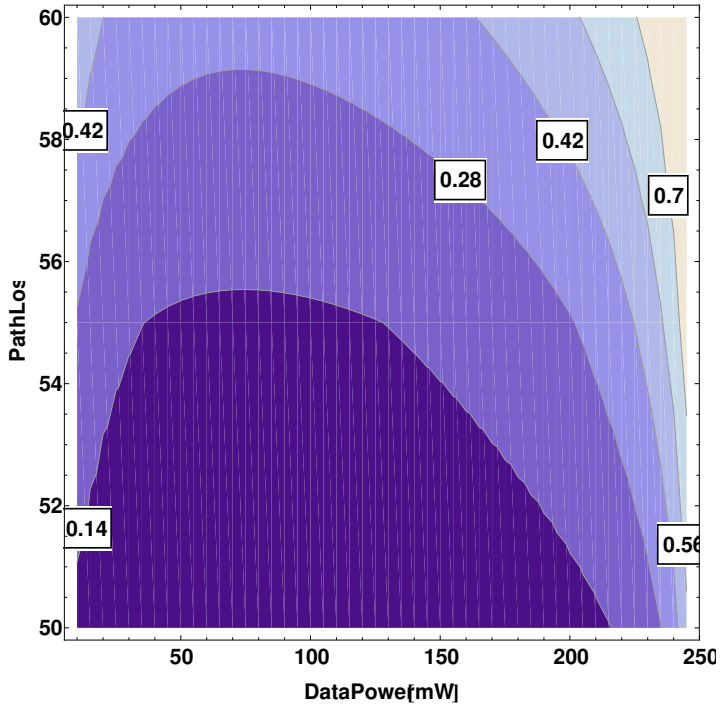

Fig. 5.4 Contour plot of the MSE as the function of the data power and the path loss under a total power (250 mW) constraint with $N_r = 20$ highly correlated antennas. For example, with the near optimal data power setting and MSE value of 0.14 can be reached at 55 dB path loss.

Proof of Lemma 5.3

First, notice that according to equation 3.326 (2) of [47]:

$$\int_{x=0}^{\infty} x^m e^{-yx^n}\,dx = \frac{\Gamma(\gamma)}{ny^{\gamma}}; \quad \gamma = \frac{m+1}{n},$$

which specifically for $n = 1$ means:

$$\frac{1}{y^{\gamma}} = \int_{x=0}^{\infty} \frac{x^{\gamma-1}}{\Gamma(\gamma)} e^{-xy}\,dx.$$

that is, for $\gamma = 1$ and $y = \|\hat{\mathbf{h}}\|^2 + \sigma^2/(\alpha^2 P)$:

$$z = \frac{\alpha\sqrt{P}}{\|\hat{\mathbf{h}}\|^2\alpha^2 P + \sigma^2} = \frac{1}{\alpha\sqrt{P}} \frac{1}{\|\hat{\mathbf{h}}\|^2 + \sigma^2/(\alpha^2 P)} =$$
$$= \frac{1}{\alpha\sqrt{P}\underbrace{\Gamma(1)}_{1}} \cdot \int_{x=0}^{\infty} e^{-x\left(\hat{\mathbf{h}}^H\hat{\mathbf{h}} + \sigma^2/(\alpha^2 P)\right)}\,dx$$

and, for $\gamma = 2$:

$$z^2 = \frac{\alpha^2 P}{(\|\hat{\mathbf{h}}\|^2 \alpha^2 P + \sigma^2)^2} = \\ = \frac{1}{\alpha^2 P \underbrace{\Gamma(2)}_{1}} \cdot \int_{x=0}^{\infty} x e^{-x\left(\hat{\mathbf{h}}^H \hat{\mathbf{h}} + \sigma^2/(\alpha^2 P)\right)} dx.$$

Term T_1

Using the singular value decomposition based transformation as in Section 5.5.2 $\mathbf{v} = \boldsymbol{\Theta}\hat{\mathbf{h}}$ and Lemma 5.3, I can write:

$$T_1 = \alpha^2 P \mathbb{E}_{\hat{\mathbf{h}}} \left\{ z^2 \, \hat{\mathbf{h}}^H \mathbf{D} \hat{\mathbf{h}} \hat{\mathbf{h}}^H \mathbf{D}^H \hat{\mathbf{h}} \right\} =$$

$$= \mathbb{E}_{\hat{\mathbf{h}}} \left\{ \int_{x=0}^{\infty} x e^{-x(\hat{\mathbf{h}}^H \hat{\mathbf{h}} + \sigma^2/(\alpha^2 P))} dx \; \hat{\mathbf{h}}^H \mathbf{D} \hat{\mathbf{h}} \hat{\mathbf{h}}^H \mathbf{D}^H \hat{\mathbf{h}} \right\} =$$

$$= \mathbb{E}_{\mathbf{v}} \left\{ \int_{x=0}^{\infty} x e^{-x(\mathbf{v}^H \mathbf{v} + \sigma^2/(\alpha^2 P))} dx \; \mathbf{v}^H \mathbf{S_D} \mathbf{v} \mathbf{v}^H \mathbf{S_D}^H \mathbf{v} \right\} =$$
$$= \mathbb{E}_{\mathbf{v}} \left\{ \int_{x=0}^{\infty} x e^{-x\left(\sum_i v_i^H v_i + \sigma^2/(\alpha^2 P)\right)} \, dx \cdot \right.$$
$$\left. \cdot \left(\sum_k v_k^H \mathbf{S}_{\mathbf{D}kk} v_k \right) \left(\sum_\ell v_\ell^H \mathbf{S}_{\mathbf{D}\ell\ell} v_\ell \right) \right\} =$$
$$= \int_{x=0}^{\infty} x e^{-x\sigma^2/(\alpha^2 P)} \cdot \mathbb{E}_{\mathbf{v}} \left\{ \prod_i e^{-x(v_i^H v_i)} \; \cdot \right.$$
$$\left. \cdot \left(\sum_k v_k^H \mathbf{S}_{\mathbf{D}kk} v_k \right) \left(\sum_\ell v_\ell^H \mathbf{S}_{\mathbf{D}\ell\ell} v_\ell \right) \right\} \, dx. \tag{5.19}$$

The expectation in (5.19) can be computed as follows:

$$\mathbb{E}_{\mathbf{v}} \left\{ \prod_i e^{-x|v_i|^2} dx \left(\sum_k \mathbf{S}_{\mathbf{D}kk} |v_k|^2 \right) \left(\sum_\ell \mathbf{S}_{\mathbf{D}\ell\ell} |v_\ell|^2 \right) \right\} =$$
$$= \sum_k \sum_{\ell, \ell \neq k} \mathbf{S}_{\mathbf{D}kk} \mathbf{S}_{\mathbf{D}\ell\ell} \cdot$$

$$\cdot \mathbb{E}_{\mathbf{v}} \left\{ e^{-x|v_k|^2} |v_k|^2 e^{-x|v_\ell|^2} |v_\ell|^2 \underbrace{\prod_{i, i \neq k, i \neq \ell} e^{-x|v_i|^2}}_{n-2 \text{ terms}} \right\} +$$

$$+\sum_k \mathbf{S}_{\mathbf{D}kk}^2 \;\mathbb{E}_{\mathbf{v}}\left\{e^{-x|v_k|^2}|v_k|^4 \underbrace{\prod_{i,i\neq k} e^{-x|v_i|^2}}_{n-1 \text{ terms}}\right\} =$$

$$= \sum_k \sum_{\ell,\ell\neq k} \mathbf{S}_{\mathbf{D}kk}\mathbf{S}_{\mathbf{D}\ell\ell}\; \mathbb{E}_{v_k}\left\{|v_k|^2 e^{-x|v_k|^2}\right\} \cdot$$
$$\cdot\, \mathbb{E}_{v_\ell}\left\{|v_\ell|^2 e^{-x|v_\ell|^2}\right\} \prod_{i,i\neq k,i\neq \ell} \mathbb{E}_{v_i}\left\{e^{-x|v_i|^2}\right\} +$$
$$+\sum_k \mathbf{S}_{\mathbf{D}kk}^2 \;\mathbb{E}_{v_k}\left\{|v_k|^4 e^{-x|v_k|^2}\right\} \prod_{i,i\neq k} \mathbb{E}_{v_i}\left\{e^{-x|v_i|^2}\right\},$$

where $y = |v_i|^2$ is exponentially distributed with parameter $r_i = \mathbf{S}_{\mathbf{R}ii}$. Consequently:

$$\mathbb{E}_{v_i}\left\{e^{-x|v_i|^2}\right\} = \int_{y=0}^{\infty} r_i e^{-r_i y} e^{-xy}\, dy = \frac{r_i}{x+r_i};$$

$$\mathbb{E}_{v_i}\left\{|v_i|^2 e^{-x|v_i|^2}\right\} = \int_{y=0}^{\infty} r_i e^{-r_i y} y e^{-xy}\, dy = \frac{r_i}{(x+r_i)^2};$$

$$\mathbb{E}_{v_i}\left\{|v_i|^4 e^{-x|v_i|^2}\right\} = \int_{y=0}^{\infty} r_i e^{-r_i y} y^2 e^{-xy}\, dy = \frac{2r_i}{(x+r_i)^3}.$$

Substituting all of these into the expectation and using $d_i = \mathbf{S}_{\mathbf{D}ii}$, I have:

$$\sum_k \sum_{\ell,\ell\neq k} d_k d_\ell\; \mathbb{E}_{v_k}\left\{|v_k|^2 e^{-x|v_k|^2}\right\} \mathbb{E}_{v_\ell}\left\{|v_\ell|^2 e^{-x|v_\ell|^2}\right\} \cdot$$
$$\cdot \prod_{i,i\neq k,i\neq \ell} \mathbb{E}_{v_i}\left\{e^{-x|v_i|^2}\right\} +$$
$$+\sum_k d_k^2 \;\mathbb{E}_{v_k}\left\{|v_k|^4 e^{-x|v_k|^2}\right\} \prod_{i,i\neq k} \mathbb{E}_{v_i}\left\{e^{-x|v_i|^2}\right\} =$$
$$= \sum_k \sum_{\ell,\ell\neq k} d_k d_\ell\; \frac{r_k}{(x+r_k)^2}\; \frac{r_\ell}{(x+r_\ell)^2} \prod_{i,i\neq k,i\neq \ell} \frac{r_i}{x+r_i} +$$
$$+\sum_k d_k^2 \;\frac{2r_k}{(x+r_k)^3} \prod_{i,i\neq k} \frac{r_i}{x+r_i}.$$

Multiplying and dividing with the missing terms of the products, the expectation in (5.19) can be written as:

$$\sum_k \sum_{\ell,\ell\neq k} d_k d_\ell\; \frac{1}{x+r_k}\; \frac{1}{x+r_\ell} \prod_i \frac{r_i}{x+r_i}$$
$$+\sum_k d_k^2 \;\frac{2}{(x+r_k)^2} \prod_i \frac{r_i}{x+r_i}.$$

Substituting this expression for the expectation, I finally get for $T1$:

$$T1 = \int_{x=0}^{\infty} x e^{-x\sigma^2/(\alpha^2 P)} \cdot \mathbb{E}_{\mathbf{v}} \Bigg\{ \prod_i e^{-x(v_i^H v_i)} \cdot \Bigg(\sum_k v_k^H \mathbf{S}_{\mathbf{D}kk} v_k \Bigg) \Bigg(\sum_\ell v_\ell^H \mathbf{S}_{\mathbf{D}\ell\ell} v_\ell \Bigg) \Bigg\} \, dx =$$

$$= \sum_k \sum_{\ell,\ell\neq k} d_k d_\ell \int_{x=0}^{\infty} x e^{-x\sigma^2/(\alpha^2 P)} \frac{r_k}{(x+r_k)^2} \frac{r_\ell}{(x+r_\ell)^2} \cdot \prod_{i,i\neq k,i\neq \ell} \frac{r_i}{x+r_i} \, dx +$$
$$+ \sum_k d_k^2 \int_{x=0}^{\infty} x e^{-x\sigma^2/(\alpha^2 P)} \frac{2r_k}{(x+r_k)^3} \prod_{i,i\neq k} \frac{r_i}{x+r_i} \, dx =$$

$$= \sum_k \sum_{\ell,\ell\neq k} d_k d_\ell \cdot \int_{x=0}^{\infty} x e^{-x\sigma^2/(\alpha^2 P)} \frac{1}{x+r_k} \frac{1}{x+r_\ell} \prod_i \frac{r_i}{x+r_i} \, dx +$$
$$+ \sum_k d_k^2 \int_{x=0}^{\infty} x e^{-x\sigma^2/(\alpha^2 P)} \frac{2}{(x+r_k)^2} \prod_i \frac{r_i}{x+r_i} \, dx.$$

All of these integrals can be computed efficiently, but they do not have nice closed forms.

Term T_2

$$\begin{aligned} T_2 &= \mathbb{E}_{\hat{\mathbf{h}}} \left\{ z^2 \, \hat{\mathbf{h}}^H \left(\alpha^2 P \mathbf{Q} + \sigma^2 \mathbf{I} \right) \hat{\mathbf{h}} \right\} = \\ &= \mathbb{E}_{\hat{\mathbf{h}}} \left\{ z^2 \, \hat{\mathbf{h}}^H \left(\alpha^2 P \mathbf{\Theta}^H \mathbf{S}_{\mathbf{Q}} \mathbf{\Theta} + \sigma^2 \mathbf{\Theta}^H \mathbf{\Theta} \right) \hat{\mathbf{h}} \right\} = \\ &= \mathbb{E}_{\hat{\mathbf{h}}} \left\{ z^2 \, \hat{\mathbf{h}}^H \mathbf{\Theta}^H \left(\alpha^2 P \mathbf{S}_{\mathbf{Q}} + \sigma^2 \mathbf{I} \right) \mathbf{\Theta} \hat{\mathbf{h}} \right\} = \\ &= \mathbb{E}_{\mathbf{v}} \left\{ z^2 \, \mathbf{v}^H \left(\alpha^2 P \mathbf{S}_{\mathbf{Q}} + \sigma^2 \mathbf{I} \right) \mathbf{v} \right\} = \\ &= \mathbb{E}_{\mathbf{v}} \left\{ z^2 \, \mathbf{v}^H \mathbf{S}_{\mathbf{M}} \mathbf{v} \right\}, \end{aligned}$$

where $\mathbf{S}_{\mathbf{M}} = \alpha^2 P \mathbf{S}_{\mathbf{Q}} + \sigma^2 \mathbf{I}$ is a diagonal matrix with diagonal elements $\mathbf{S}_{\mathbf{M}ii} = \alpha^2 P \mathbf{S}_{\mathbf{Q}ii} + \sigma^2$. Substituting z:

$$\begin{aligned} T_2 &= \frac{1}{\alpha^2 P} \mathbb{E}_{\mathbf{v}} \left\{ \int_{x=0}^{\infty} x e^{-x\left(\mathbf{v}^H \mathbf{v} + \frac{\sigma^2}{\alpha^2 P} \right)} \, dx \mathbf{v}^H \mathbf{S}_{\mathbf{M}} \mathbf{v} \right\} = \\ &= \frac{1}{\alpha^2 P} \mathbb{E}_{\mathbf{v}} \left\{ \int_{x=0}^{\infty} x e^{-x(\mathbf{v}^H \mathbf{v})} e^{-x\frac{\sigma^2}{\alpha^2 P}} \, dx \mathbf{v}^H \mathbf{S}_{\mathbf{M}} \mathbf{v} \right\} = \\ &= \frac{1}{\alpha^2 P} \int_{x=0}^{\infty} x e^{-x\frac{\sigma^2}{\alpha^2 P}} \mathbb{E}_{\mathbf{v}} \left\{ \prod_i e^{-x v_i^H v_i} \sum_k v_k^H \mathbf{S}_{\mathbf{M}kk} v_k \right\} dx. \end{aligned} \tag{5.20}$$

Similarly to the derivation of the expectation expression in T_1, the expectation in (5.20) can be written as:

$$\mathbb{E}_{\mathbf{v}}\left\{\prod_i e^{-x(v_i^H v_i)} \sum_k v_k^H \mathbf{S}_{\mathbf{M}kk} v_k\right\} =$$
$$= \sum_k \mathbf{S}_{\mathbf{M}kk} \mathbb{E}_{v_k}\left\{|v_k|^2 e^{-x|v_k|^2}\right\} \prod_{i\neq k} \mathbb{E}_{v_i}\left\{e^{-x|v_i|^2}\right\} =$$

$$= \sum_k \mathbf{S}_{\mathbf{M}kk} \frac{r_k}{(x+r_k)^2} \prod_{i\neq k} \frac{r_i}{x+r_i} =$$

$$= \sum_k \mathbf{S}_{\mathbf{M}kk} \frac{1}{x+r_k} \prod_i \frac{r_i}{x+r_i}.$$

Finally T_2 is

$$T_2 = \frac{1}{\alpha^2 P} \sum_k \mathbf{S}_{\mathbf{M}kk} \int_{x=0}^{\infty} x e^{-x\frac{\sigma^2}{\alpha^2 P}} \frac{1}{x+r_k} \prod_i \frac{r_i}{x+r_i} dx.$$

Term T_3

$$T_3 = 2\alpha\sqrt{P}\, \mathbb{E}_{\hat{\mathbf{h}}}\left\{z\ Re\{\hat{\mathbf{h}}^H \mathbf{D} \hat{\mathbf{h}}\}\right\} =$$
$$= 2\alpha\sqrt{P}\, \mathbb{E}_{\hat{\mathbf{h}}}\left\{z\ Re\{\hat{\mathbf{h}}^H \mathbf{\Theta}^H \mathbf{S}_{\mathbf{D}} \mathbf{\Theta} \hat{\mathbf{h}}\}\right\} =$$
$$= 2\alpha\sqrt{P}\, \mathbb{E}_{\mathbf{v}}\left\{z\ Re\{\underbrace{\mathbf{v}^H \mathbf{S}_{\mathbf{D}} \mathbf{v}}_{\text{real}}\}\right\} = 2\alpha\sqrt{P}\, \mathbb{E}_{\mathbf{v}}\left\{z\ \mathbf{v}^H \mathbf{S}_{\mathbf{D}} \mathbf{v}\}\right\}.$$

and, for the expectation, the same approach is applicable as above. That is:

$$T_3 = 2\alpha\sqrt{P} \frac{1}{\alpha\sqrt{P}}\, \mathbb{E}_{\mathbf{v}}\left\{\int_{x=0}^{\infty} e^{-x\left(\mathbf{v}^H \mathbf{v} + \frac{\sigma^2}{\alpha^2 P}\right)}\, dx\ \mathbf{v}^H \mathbf{S}_{\mathbf{D}} \mathbf{v}\right\} =$$

$$= 2\ \mathbb{E}_{\mathbf{v}}\left\{\int_{x=0}^{\infty} e^{-x(\mathbf{v}^H \mathbf{v})} e^{-x\frac{\sigma^2}{\alpha^2 P}}\, dx\ \mathbf{v}^H \mathbf{S}_{\mathbf{D}} \mathbf{v}\right\} =$$

$$= 2 \int_{x=0}^{\infty} e^{-x\frac{\sigma^2}{\alpha^2 P}}\, \mathbb{E}_{\mathbf{v}}\left\{\prod_i e^{-x v_i^H v_i} \sum_k v_k^H \mathbf{S}_{\mathbf{D}kk} v_k\right\} dx =$$

$$= 2 \int_{x=0}^{\infty} e^{-x\frac{\sigma^2}{\alpha^2 P}} \sum_k \mathbf{S}_{\mathbf{D}kk} \frac{1}{x+r_k} \prod_i \frac{r_i}{x+r_i} dx =$$
$$= 2 \sum_k \mathbf{S}_{\mathbf{D}kk} \int_{x=0}^{\infty} e^{-x\frac{\sigma^2}{\alpha^2 P}} \frac{1}{x+r_k} \prod_i \frac{r_i}{x+r_i} dx.$$

Chapter 6
Block and Comb Type Channel Estimation

6.1 Introduction

Since the seminal work by Hassibi and Hochwald [13], a number of papers investigated the trade-off between the resources used for CSI acquisition and data transmission. For example, assuming a block fading reciprocal channel, a finite number of symbols in the time and frequency domains are available for CSI acquisition, and uplink as well as DL precoding and data transmission [15], [48]. Also, under a fixed power budget, pilot symbols reduce the transmitted energy for data symbols, as it has been pointed out in [29] and [30] where the optimal PDPR is investigated for various pilot patterns and receiver structures. The results of [30], for example, indicate that the optimal PDPR provides about 2-3 dB gain compared with equal power for pilot and data symbols. Subsequently, [49] derived a closed form of the optimal PDPR for MMSE channel estimation and showed that a tight bound lying in the quasi-optimal region provides a good approximation for the optimal PDPR. More recently, [50] derived a closed form PDPR that maximizes the capacity bound of MIMO orthogonal frequency division modulation (MIMO-OFDM) systems and studied the impact of carrier frequency offset (CFO) on maximizing power allocation.

In [16] and [51], equivalence conditions for the achievable spectral efficiency between block-fading channels and continuous-fading channels are discussed. An approximate closed-form analytical expression of the spectral efficiency is derived in the hypothesis of MMSE estimation.

In a previous work [31], I investigated the effects of the PDPR on the MSE, assuming a single pilot and a single data symbol under a fixed sum power budget with least square (LS) channel estimation at the base station (BS). While [31] provides insight into the trade-off related to PDPR, it does not consider the trade-off related to the number of symbols used for CSI acquisition and data transmission, which is critical for the spectral efficiency. Therefore, the purpose of the present chapter is to devise a methodology to find the optimum number of pilot and data symbols *and* the optimum PDPR. It turns out that the constraints for these trade-offs depend on the pilot pattern that is used in the time and frequency domains. Specifically, the so called *block* type arrangement dedicates all frequency channels within a given time slot to either channel estimation or data transmission whereas the *comb* pilot pattern employs pilot and data symbols mixed in the frequency domain within a single time slot.

The design of the uplink demodulation reference signals (DMRS) specifically in 3GPP Long Term Evolution Advanced (LTE-A) systems is described in [52]. In the LTE uplink, DMRS:s are used to facilitate channel estimation for the coherent demodulation of the physical uplink shared and control channels. The LTE DMRS:s occupy specific Orthogonal Frequency Division Multiplexing (OFDM) symbols within the uplink subframe according to the block type arrangement and support a large number of user equipment utilizing cyclic extensions of the well known Zadoff-Chu sequences [53].

My key contribution in the present chapter is the derivation of a closed form solution for both the uplink data MSE and spectral efficiency specifically taking into account the constraints of the comb and block type pilot arrangements. As a major difference with respect to previous works (i.e., [13]–[51], this closed form result allows me to find the close-to-optimum number of pilot symbols and pilot power for a generic channel estimation method. In particular, I compare LS and MMSE channel estimation in block-type and comb-type pilot arrangement, for a BS employing a large number of antennas. This approach enables me to arrive at some insights that are novel in the massive MIMO literature.

The system model, including the description of the block and comb type pilot patterns, is defined in Sect. 6.2. Next, Section III, discusses the LS and MMSE channel estimation algorithms specifically in block or comb systems and, in Section IV, I introduce the the MSE and spectral efficiency for the

equalized uplink data symbols assuming linear MMSE reception. My main analytical results are derived in Section V. Section VI presents numerical results and Section VII concludes the chapter.

6.2 System Model

I consider the uplink transmission of a multi-antenna single cell wireless system, in which users are scheduled on orthogonal frequency channels. It is assumed that each mobile station (MS) employs an orthogonal pilot sequence, so that no interference between pilots is present in the system. This is a common assumption in massive multi-user MIMO systems in which a single MS may have a single antenna. The base station (BS) estimates the channel $\mathbf{h}$ (column vector of dimension N_r, where N_r is the number of receive antennas at the BS) by either LS or MMSE channel estimation to initialize an MMSE equalizer for uplink data reception. Since I assume orthogonal pilot sequences, the channel estimation process can be assumed independent for each MS. I consider a time-frequency resource of T time slots in the channel coherence time, and F subcarriers in the coherence bandwidth, with a total number of symbols $\tau = F \cdot T$. I denote by τ_p the number of symbols allocated to pilots, and by τ_d the number of symbols allocated to data ($\tau_p + \tau_d = \tau$). Moreover, I consider a transmission power level P_p and P for each pilot and data symbol, respectively. With this setup, I consider two pilot symbol allocation methods, namely block type and comb type, which I discuss in the following subsections.

6.2.1 Block Type Pilot Allocation

The block type pilot arrangement consists of allocating one or more time slots for pilot transmission, by using all subcarriers in those time slots. This approach is a suitable strategy for slow time-varying channels. Given T slots, a fraction of T_p *slots* are allocated to the pilot and $T_d = T - T_p$ slots are allocated to the data symbols. Note that a maximum transmission power P_{tot} is allowed in each time slot, among all F subcarriers. This power constraint is then identical for both the pilot (P_p) and data power (P), i.e.,

$$FP_p \leq P_{tot} \quad FP \leq P_{tot}. \tag{6.1}$$

The power cannot be traded between pilot and data, but the *energy* budget can be distributed by tuning the number of time slots T_p and T_d, i.e., $\tau_p = FT_p$ and $\tau_d = FT_d$.

6.2.2 Comb Type Pilot Allocation

In the comb type pilot arrangement a certain number of subcarriers are allocated to pilot symbols, continuously in time. This approach is a suitable strategy for non-frequency selective channels. Given F subcarriers in the coherence bandwidth, a fraction of F_p subcarriers are allocated to the pilot and $F_d = F - F_p$ subcarriers are allocated to the data symbols.

Each MS transmits at a constant power P_{tot}, however, the transmission power can be distributed unequally in each subcarrier. In particular, let us consider a transmitted power P_p for each pilot symbol and P for each data symbol transmission; then the following constraint is enforced:

$$F_p P_p + (F - F_p)P = P_{tot}. \tag{6.2}$$

The total number of symbols for pilots is $\tau_p = TF_p$ and for data is $\tau_d = TF_d$. However, with comb type pilot arrangement, the trade-off between pilot and data signals includes the trade-offs between the

number of frequency channels *and* between the transmit power levels, which is an additional degree of freedom compared with the block type arrangement.

6.3 Channel Estimation

Let us consider a MS that transmits an orthogonal pilot sequence $\mathbf{s} = [s_1, ..., s_{\tau_p}]^T$, where each symbol is scaled as $|s_i|^2 = 1$, for $i = 1, .., \tau_p$. Thus, the $N_r \times \tau_p$ matrix of the received pilot signal at the BS from the MS is

$$\mathbf{Y}^p = \alpha\sqrt{P_p}\mathbf{h}\mathbf{s}^T + \mathbf{N}, \tag{6.3}$$

where I assume that $\mathbf{h}$ is a circular symmetric complex normal distributed vector of r.v. with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{C}$ (of size N_r), denoted as $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, α accounts for the propagation loss, $\mathbf{N} \in \mathbb{C}^{N_r \times \tau_p}$ is the spatially and temporally additive white Gaussian noise (AWGN) with element-wise variance σ^2.

In this chapter, I consider two techniques, i.e., the least square (LS) and the minimum mean-square error (MMSE) channel estimation that are detailed in the following subsections.

6.3.1 Least Square Estimation

Conventional LS estimation relies on correlating the received signal with the known pilot sequence. The BS estimates the channel based on (6.3) assuming

$$\hat{\mathbf{h}}_{LS} = \mathbf{h} + \tilde{\mathbf{h}}_{\mathbf{LS}} = \frac{1}{\alpha\sqrt{P_p}}\mathbf{Y}^p\mathbf{s}^*(\mathbf{s}^T\mathbf{s}^*)^{-1} = \mathbf{h} + \frac{1}{\alpha\sqrt{P_p}\tau_p}\mathbf{N}\mathbf{s}^*. \tag{6.4}$$

Note that $\mathbf{N}\mathbf{s}^* = \left[\sum_{i=1}^{\tau_p} s_i^* n_{i,1}, ..., \sum_{i=1}^{\tau_p} s_i^* n_{i,N_r}\right]^T$, then $\mathbf{N}\mathbf{s}^* \sim \mathcal{CN}(\mathbf{0}, \tau_p\sigma^2\mathbf{I}_{N_r})$.

By considering $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, it follows that the estimated channel $\hat{\mathbf{h}}_{LS}$ is a circular symmetric complex normal distributed vector $\hat{\mathbf{h}}_{LS} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{LS})$, with

$$\mathbf{R}_{LS} = \mathbb{E}\{\hat{\mathbf{h}}_{LS}\hat{\mathbf{h}}_{LS}^H\} = \mathbf{C} + \frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r}. \tag{6.5}$$

The channel estimation error is defined as $\tilde{\mathbf{h}}_{LS} = \mathbf{h} - \hat{\mathbf{h}}_{LS}$, so that $\tilde{\mathbf{h}}_{LS} \sim \mathcal{CN}(\mathbf{0}, \mathbf{W}_{LS})$ with

$$\mathbf{W}_{LS} = \frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r}$$

and the estimation mean square error (MSE) is derived as

$$\varepsilon_{LS} = \mathbb{E}\{||\tilde{\mathbf{h}}_{LS}||_F^2\} = \mathrm{tr}\{\mathbf{W}_{LS}\} = \frac{N_r\sigma^2}{\alpha^2 P_p \tau_p}, \tag{6.6}$$

where $||\cdot||_F^2$ is the Frobenius norm.

6.3.2 Minimum Mean Squared Error Estimation

I define a training matrix $\mathbf{S} = \mathbf{s} \otimes \mathbf{I}_{N_r}$ (of size $\tau_p N_r \times N_r$), so that $\mathbf{S}^H\mathbf{S} = \tau_p \mathbf{I}_{N_r}$. The $\tau_p N_r \times 1$ vector of received signal (6.3) can be conveniently rewritten as

$$\tilde{\mathbf{Y}}^p = \alpha\sqrt{P_p}\mathbf{S}\mathbf{h} + \tilde{\mathbf{N}}.$$

where $\tilde{\mathbf{Y}}^p, \tilde{\mathbf{N}} \in \mathbb{C}^{\tau_p N_r \times 1}$.
The MMSE equalizer aims at minimizing the MSE between the estimate $\hat{\mathbf{h}}_{MMSE} = \mathbf{H}\tilde{\mathbf{Y}}^p$ and the channel $\mathbf{h}$. More precisely,

$$\begin{aligned}\mathbf{H} =& \arg\min_{\mathbf{H}} \mathbb{E}\{||\mathbf{H}\tilde{\mathbf{Y}}^p - \mathbf{h}||_F^2\} \\ =& \alpha\sqrt{P_p}(\sigma^2\mathbf{I}_{N_r} + \alpha^2 P_p \mathbf{C}\mathbf{S}^H\mathbf{S})^{-1}\mathbf{C}\mathbf{S}^H; \quad \mathbf{H} \in \mathbb{C}^{N_r \times \tau_p N_r}.\end{aligned}$$

The MMSE estimate is then expressed as

$$\begin{aligned}\hat{\mathbf{h}}_{MMSE} &= \alpha\sqrt{P_p}(\sigma^2\mathbf{I}_{N_r} + \alpha^2 P_p \tau_p \mathbf{C})^{-1}\mathbf{C}\mathbf{S}^H(\alpha\sqrt{P_p}\mathbf{S}\mathbf{h} + \tilde{\mathbf{N}}) \\ &= \left(\frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r} + \mathbf{C}\right)^{-1}\mathbf{C}\left(\mathbf{h} + \frac{1}{\alpha\sqrt{P_p}\tau_p}\mathbf{S}^H\tilde{\mathbf{N}}\right). \end{aligned} \tag{6.7}$$

Notice that $\mathbf{S}^H\mathbf{N} \sim \mathcal{CN}(\mathbf{0}, \tau_p\sigma^2\mathbf{I}_{N_r})$, therefore the estimated channel $\hat{\mathbf{h}}_{MMSE}$ is also a circular symmetric complex normal distributed vector $\hat{\mathbf{h}}_{MMSE} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{MMSE})$, that is

$$\hat{\mathbf{h}}_{MMSE} = \mathbf{h} + \tilde{\mathbf{h}}_{MMSE},$$

and

$$\mathbf{R}_{MMSE} = \mathbf{C}^2\left(\frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r} + \mathbf{C}\right)^{-1}, \tag{6.8}$$

where I considered $\mathbf{C} = \mathbf{C}^H$ and applied the commutativity of $\mathbf{C}$ and $\mathbf{I}_{N_r}$ to substitute

$$\left(\frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r} + \mathbf{C}\right)^{-1}\mathbf{C} = \mathbf{C}\left(\frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r} + \mathbf{C}\right)^{-1}.$$

The channel estimation error is $\tilde{\mathbf{h}}_{MMSE} = \mathbf{h} - \hat{\mathbf{h}}_{MMSE}$ so that $\tilde{\mathbf{h}}_{MMSE} \sim \mathcal{CN}(\mathbf{0}, \mathbf{W}_{MMSE})$ with

$$\mathbf{W}_{MMSE} = \mathbf{C}\left(\mathbf{I}_{N_r} + \frac{\alpha^2 P_p \tau_p}{\sigma^2}\mathbf{C}\right)^{-1}$$

and the estimation MSE simply follows as

$$\varepsilon_{MMSE} = \mathrm{tr}\left\{\mathbf{C}\left(\mathbf{I}_{N_r} + \frac{\alpha^2 P_p \tau_p}{\sigma^2}\mathbf{C}\right)^{-1}\right\}. \tag{6.9}$$

Notice that for both LS and MMSE channel estimation, the estimation MSE is a monotonically decreasing function of the pilot energy per antenna $P_p\tau_p$.

In the next section, I characterize the receiver and the uplink signal MSE and spectral efficiency based on $\tilde{\mathbf{h}}$ which is computed for LS ($\hat{\mathbf{h}}_{LS}$) and MMSE estimation ($\hat{\mathbf{h}}_{MMSE}$) in this section.

6.4 Linear Minimum Mean Squared Error Receiver

I consider a MMSE receiver at the BS. For each transmitted data symbol x, the data signal received by the BS can be written as

$$\mathbf{y} = \alpha\sqrt{P}\mathbf{h}x + \mathbf{n},$$

where $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_{N_r})$.

Using a linear detection matrix $\mathbf{G}$ of size $1 \times N_r$, the mean-square error between the estimate $\mathbf{Gy}$ and the transmitted symbol x is $(\mathbf{Gy} - x)^2$. The resulting $\text{MSE}(\mathbf{h}) = \mathbb{E}_{x,\mathbf{n}}\{(\mathbf{Gy} - x)^2\}$ is

$$\begin{aligned}\text{MSE}(\mathbf{h}) =& \mathbf{G}(\alpha^2 P\mathbf{h}\mathbf{h}^H + \sigma^2\mathbf{I})_{N_r}\mathbf{G}^H \\ &- \alpha\sqrt{P}(\mathbf{G}\mathbf{h} + \mathbf{h}^H\mathbf{G}^H) + 1. \end{aligned} \tag{6.10}$$

According to [54], the instantaneous signal-to-noise-ratio (SNR) of the MMSE receiver is given by:

$$\gamma(\mathbf{h}) = \frac{1}{\text{MSE}(\mathbf{h})} - 1.$$

Using this relationship, the achievable spectral efficiency can be expressed as:

$$S(\mathbf{h}) = \frac{\left(\tau - \tau_p\right)}{\tau}\left[\log\left(\frac{1}{\text{MSE}(\mathbf{h})}\right)\right]. \tag{6.11}$$

Recall that the performance of the MMSE receiver depends on the availability of channel state information (CSI) at the receiver. In the considered scenario, the channel $\mathbf{h}$ is not available at the BS. The BS implements the MMSE receiver by using the estimated channel $\hat{\mathbf{h}}$ as if it was the actual channel, [31], i.e., the detection matrix, which is a function of $\hat{\mathbf{h}}$ (more precisely, $\hat{\mathbf{h}}_{LS}$ or $\hat{\mathbf{h}}_{MMSE}$) is calculated as

$$\mathbf{G} = \mathbf{G}(\hat{\mathbf{h}}) = \frac{\alpha\sqrt{P}\hat{\mathbf{h}}^H}{\alpha^2 P||\hat{\mathbf{h}}||^2 + \sigma^2}. \tag{6.12}$$

In the next section, I derive a closed form analytical expression for the uplink MSE and the spectral efficiency for both LS and MMSE channel estimation.

6.5 Analytical Derivation of the Spectral Efficiency

In this section, I propose an analytical model to study the performance of the single-cell multiple antenna system illustrated in the previous section, in order to derive the optimal resource (slot/frequency/power) allocation for pilot and data. I first introduce a useful lemma and then I derive (6.10) and (6.11) analytically, by considering the MMSE receiver in (6.12).

6.5.1 Conditional Distribution of the Channel

When implementing the MMSE receiver in (6.12), the expression of the uplink MSE in (6.10) contains both the actual channel $\mathbf{h}$ and the estimated channel $\hat{\mathbf{h}}$. However, since the square error is averaged over $\mathbf{n}$, the MSE depends on the conditional distribution $(\hat{\mathbf{h}}|\mathbf{h})$ or, equivalently, $(\mathbf{h}|\hat{\mathbf{h}})$. Therefore, in order to compute the unconditional MSE analytically, I introduce the following lemma.

Lemma 6.1. *The conditional distribution of the channel given its estimation* $\hat{\mathbf{h}}$ *is*

$$(\mathbf{h}|\hat{\mathbf{h}}) \sim \mathcal{CN}(\mathbf{D}\hat{\mathbf{h}}, \mathbf{Q}), \tag{6.13}$$

where

$$\mathbf{D} = \begin{cases} \mathbf{C}\mathbf{R}_{LS}^{-1} & \textit{for LS estimation} \\ \mathbf{I}_{N_r} & \textit{for MMSE estimation} \end{cases}$$

$$\mathbf{Q} = \begin{cases} \mathbf{C} - \mathbf{C}\mathbf{R}_{LS}^{-1}\mathbf{C} & \textit{for LS estimation} \\ \mathbf{C} - \mathbf{R}_{MMSE} & \textit{for MMSE estimation} \end{cases}$$

with $\mathbf{R}_{LS}$ *and* $\mathbf{R}_{MMSE}$ *given in* (6.5) *and* (6.7)*, respectively.*

Proof. The proof is reported in the Appendix.

6.5.2 Calculating the Uplink Mean Squared Error

By using Lemma 6.1, the MSE in (6.10) can be conveniently expressed as a function of only the estimated channel $\hat{\mathbf{h}}$ as

$$\begin{aligned} \mathrm{MSE}(\hat{\mathbf{h}}) =& \mathbf{G}(\hat{\mathbf{h}}) \left(\alpha^2 P (\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H \mathbf{D}^H + \mathbf{Q}) + \sigma^2 \mathbf{I}_{N_r} \right) \mathbf{G}(\hat{\mathbf{h}})^H \\ & - \alpha\sqrt{P} \left(\mathbf{G}(\hat{\mathbf{h}})\mathbf{D}\hat{\mathbf{h}} + \hat{\mathbf{h}}^H \mathbf{D}^H \mathbf{G}(\hat{\mathbf{h}})^H \right) + 1. \end{aligned} \tag{6.14}$$

In the special case of independent channel distributions with identical variances (i.e., $\mathbf{C} = c\mathbf{I}_{N_r}$), I exploit the following lemma.

Lemma 6.2. *Assume* $\mathbf{C} = c\mathbf{I}_{N_r}$*, where* $c \in \mathbb{R}^+$*, then the matrices* $\mathbf{D}$ *and* $\mathbf{Q}$ *are diagonal with* $\mathbf{D} = d\mathbf{I}_{N_r}$ *and* $\mathbf{Q} = q\mathbf{I}_{N_r}$ *and the MSE is given by*

$$MSE(\hat{\mathbf{h}}) = \frac{p^2 ||\hat{\mathbf{h}}||^4 (d-1)^2 + p||\hat{\mathbf{h}}||^2 (2\sigma^2 - 2d\sigma^2 + b) + \sigma^4}{\left(p||\hat{\mathbf{h}}||^2 + \sigma^2 \right)^2}, \tag{6.15}$$

where $p = \alpha^2 P$ *and* $b = qp + \sigma^2$*.*

Proof. The proof is reported in the appendix.

It is important to notice that (6.15) depends on the channel only though the norms $\|\hat{\mathbf{h}}\|^2$ and $||\hat{\mathbf{h}}||^4$.

6.5.3 Calculating the Spectral Efficiency

The spectral efficiency expression in (6.11) can be also conveniently rewritten as a function of $\hat{\mathbf{h}}$, by using the results from Lemma 6.1 and 6.2. The average spectral efficiency is then

$$\bar{S} = \frac{(\tau - \tau_p)}{\tau} \mathbb{E}_{\hat{\mathbf{h}}} \left\{ \left[\log \left(\frac{1}{\mathrm{MSE}(\hat{\mathbf{h}})} \right) \right] \right\}. \tag{6.16}$$

which leads to the following results.

Theorem 6.5.1 (Spectral efficiency with LS estimation) *Assume* $\mathbf{C} = c\mathbf{I}_{N_r}$*, where* $c \in \mathbb{R}^+$*, then the average spectral efficiency with LS channel estimation and MMSE receiver is*

$$\bar{S}_{LS} = \frac{(\tau - \tau_p)}{\tau}\left(\frac{2\mathcal{G}(x_0) - \mathcal{G}(x_1) - \mathcal{G}(x_2)}{(N_r - 1)!} - \log(d-1)^2\right) \tag{6.17}$$

with $x_{1,2} = \frac{1}{2}\left(-\frac{2\sigma^2 - 2d\sigma^2 + b}{p(d-1)^2} \pm \sqrt{\left(\frac{2\sigma^2 - 2d\sigma^2 + b}{p(d-1)^2}\right)^2 - \frac{4\sigma^4}{p^2(d-1)^2}}\right)$, $x_0 = \frac{\sigma^2}{p}$, $p = \alpha^2 P$, $b = qp + \sigma^2$, $q = c(1 - c/r$, $r = c + \frac{\sigma^2}{\alpha^2 P_p \tau_p}$,
and where

$$\mathcal{G}(x) = \mathbf{MeijerG}^{1,3}_{2,3}\left(\begin{matrix} 0, 1 \\ 0, 0, N_r \end{matrix} \middle| \frac{x}{r}\right), \tag{6.18}$$

is the Meijer G-function.

Proof. The proof is reported in the Appendix of this chapter.

Theorem 6.5.2 (Spectral efficiency with MMSE estimation) *Assume* $\mathbf{C} = c\mathbf{I}_{N_r}$*, where* $c \in \mathbb{R}^+$*, then the average spectral efficiency with MMSE channel estimation and MMSE receiver is*

$$\bar{S}_{MMSE} = \frac{(\tau - \tau_p)}{\tau}\left(\log(pb) + \frac{2\mathcal{G}(x_3) - \mathcal{G}(x_4)}{(N_r - 1)!}\right) \tag{6.19}$$

with $x_3 = \frac{\sigma^2}{p}$, $x_4 = \frac{\sigma^2}{pb}$, $b = qp + \sigma^2$, $q = \frac{\sigma^2 c}{\sigma^2 + \alpha^2 c P_p \tau_p}$*, and* $\mathcal{G}(x)$ *defined in* (6.18)*.*

Proof. The proof is reported in the Appendix of this chapter.

Notice that for a fixed pilot energy, the average spectral efficiency decreases with τ_p. Intuitively, this means that for a given pilot energy ($\tau_p P_p$), this energy should be concentrated to as few time slots as possible with the maximum allowed pilot power P_p. This is because the estimation MSE in Eqs. (6.6) and (6.9) only depend on the τ_p and P_p through the product (that is the pilot energy). The hypothesis of independent and identical channel distributions (i.e., $\mathbf{C} = c\mathbf{I}_{N_r}$) is necessary to obtain a tractable analytical expression and it is widely adopted in the related literature (i.e., [16]). Although correlation among antennas may have an effect on the value of achievable spectral efficiency, the optimal pilot resource allocation is not expected to change significantly with the antenna correlation.

6.5.4 Summary

In this section, I derived the spectral efficiency for a single cell multi-antenna system with pilot-based channel estimation that implements a linear MMSE receiver. Moreover, I provided a closed form analytical expression of the spectral efficiency that makes use of the Meijer G-function, for both LS and MMSE channel estimation methods.

I started by deriving the conditional distribution ($\mathbf{h}|\hat{\mathbf{h}}$), which allows for computing the uplink MSE as a function of the norm of the estimated channel $\|\hat{\mathbf{h}}\|^2$ through Lemma 6.2. By averaging the spectral efficiency over the distribution of $\|\hat{\mathbf{h}}\|^2$, I derived the closed form expression of the spectral efficiency for LS estimation in Result 6.5.1 and for MMSE estimation in Result 6.5.2.

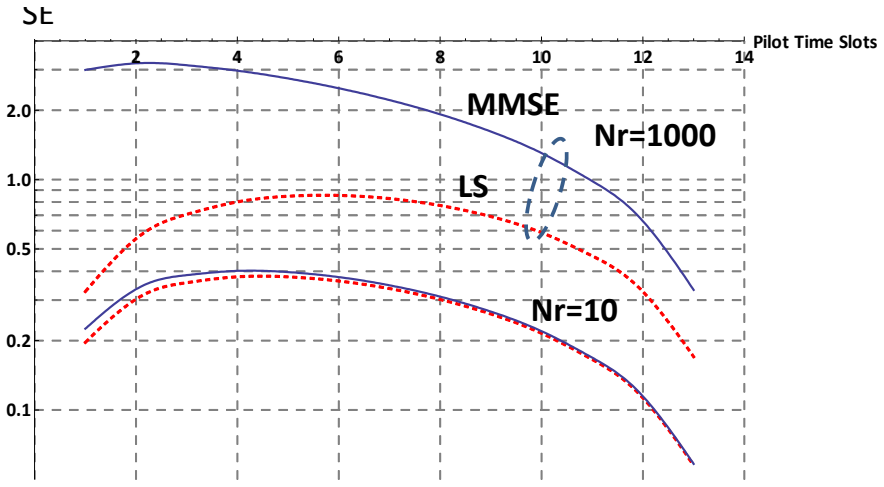

Fig. 6.1 Spectral efficiency (SE) in bps/Hz in log scale of block type channel estimation as a function of the number of pilot time slots with $N_r = 10$ and $N_r = 1000$ antennas at the BS. With block type arrangement and EPA, all $F = 12$ subcarriers in each of the $T = 14$ time slots are dedicated to either pilot or data transmission with $P_{tot} = 250$ mW total transmit power shared equally in the frequency domain. The T_p that maximizes spectral efficiency is clearly different with LS and MMSE estimations.

Under the constraints (6.1) and (6.2) for symbols and power in the time-frequency domain, this model reproduces the behavior of the block type and comb type pilot arrangement, respectively, as I will show in the following section.

6.6 Numerical Results

6.6.1 Equal Power Density for Each Symbol Allocation

Recall that equal power allocation (EPA) implies that all symbols within a slot are transmitted with P_{tot}/F transmit power. In this case, the pilot-data resource allocation trade-off consists of the number of time slots (T_p out of T) in the case of block type arrangement or the number of frequency channels (F_p out of F) in the case of comb type arrangement used for pilot transmission.

Figure 6.1 shows the spectral efficiency in bps/Hz as the function of time slots T_p when using LS and MMSE channel estimation with $N_r = 10$ and $N_r = 1000$ BS antennas for a specific value of the large scale fading (50 dB). As expected, the maximum spectral efficiency that can be reached with MMSE estimation (3 bps/Hz) is much higher than that of the optimum spectral efficiency value with LS estimation (1 bps/Hz). Also, the optimum spectral efficiency is reached with different T_p settings: while with MMSE $T_p = 2$ maximizes the spectral efficiency, with LS more slots need to be spent on CSI acquisition ($T_p = 5$).

Figure 6.2 shows a similar tendency in terms of the necessary frequency channels (subcarriers) that optimize the spectral efficiency with comb type pilot arrangement. The similar (almost identical) behavior that can be seen in Figure 6.1 and 6.2 can be explained by noticing that under the assumption of channel coherence in the time and frequency domains of a resource block of $F = 12$ frequency channel and $T = 14$ time slots, block and comb type arrangements with equal power allocation among the $F \times T = 168$ resource elements result in the same total pilot energy. That is under the coherence

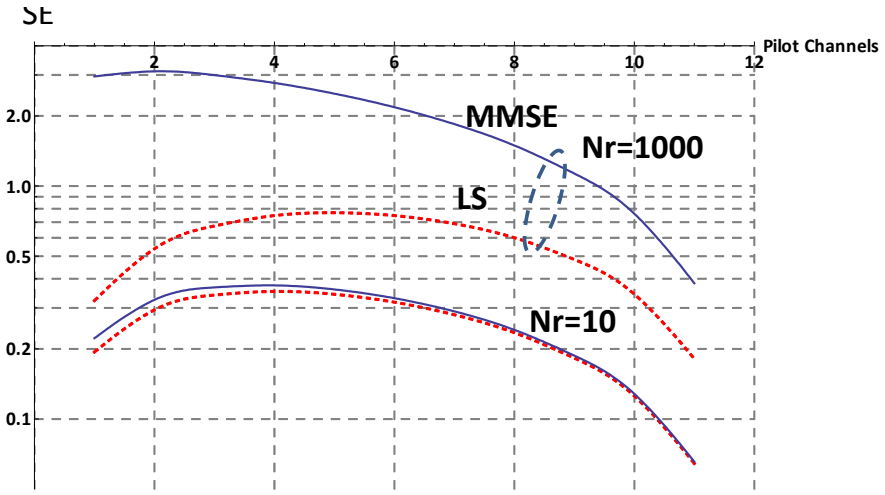

Fig. 6.2 Spectral efficiency (SE) in log scale of comb type channel estimation as a function of the number of pilot channels in the frequency domain with $N_r = 10$ and $N_r = 1000$ antennas at the BS. With comb type arrangement and EPA, F_p subcarriers (each with P_{tot}/F transmit power) are used to transmit pilot symbols in each of the $T = 14$ time slots. The F_p that maximizes spectral efficiency is clearly different with LS and MMSE estimations.

assumption, the roles of the time and frequency domains from the perspective of pilot energy are the same.

6.6.2 Optimum Power Allocation

When using the comb pilot symbol pattern, it is possible to use F_p subcarriers for transmitting pilot symbols and the remaining $F - F_p$ symbols for data transmission in each time slot. In this case, it is also possible to use unequal total power for pilot (P_p) and data symbol transmission as long as the sum over the F symbols in each time slot does not exceed the P_{tot} power budget. In this case, each of the F_p pilot subcarriers are transmitted with F_p/P_p transmit power and both F_p and P_p are design parameters.

Figure 6.3 shows the value of the pilot power P_p in mW that maximizes the spectral efficiency when using the comb arrangement and employing LS (upper) and MMSE (lower) channel estimation as the number of antennas grows from $N_r = 2$ to $N_r = 1000$. With LS, the optimal pilot power grows from about 40% to around 80% of the P_{tot} total power budget, whereas with MMSE estimation, the optimal pilot power remains the same.

Figure 6.4 shows the achieved spectral efficiency as the function of the antennas with LS and MMSE estimation with equal (EPA) and optimal pilot power allocation. Optimizing the pilot power is clearly beneficial with both LS and MMSE estimations. With LS estimation, optimizing the pilot power is particularly important as the number of antennas grows large, but even with MMSE estimation, the spectral efficiency increases by 20%. Recall that unequal power allocation over the subcarriers in each time slot requires comb type arrangement, implying that with block type pilot pattern, MMSE estimation gives large gains over LS estimation when the number of antennas is large.

Figures 6.5 and 6.6 compare the achievable spectral efficiency as the function of the pilot power and the number of pilot frequency channels with $N_r = 10$ and $N_r = 1000$ receive antennas. With $N_r = 10$, the

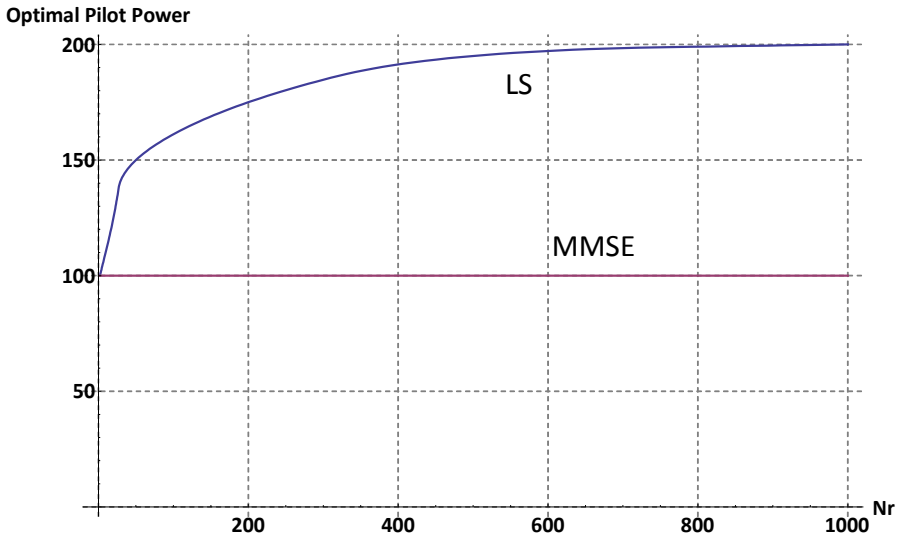

Fig. 6.3 Optimum pilot power in mW as the function of the number of receive antennas at the BS when using LS (upper curve) or MMSE (lower curve) channel estimation. With LS estimation, the optimum pilot power increases with the number of antennas, whereas with MMSE estimation, the optimum pilot power is constant (staying at 40% of the total power budget P_{tot} in each time slot).

achievable spectral efficiency is similar with LS and MMSE channel estimation, whereas with $N_r = 1000$ antennas, the optimal spectral efficiency is roughly twice as high with MMSE as with LS.

6.7 Conclusions

This chapter considered the trade-off between the time, frequency and power resources allocated to the transmission of pilot and data symbols and its impact on the MSE and spectral efficiency of the uplink of a single cell system, in which the number of receive antennas grows large. I made the point that the joint allocation of frequency, time and power resources is subject to constraints that depend on the specific pilot pattern, such as the pattern used by the block and comb type pilot arrangements. In this rather general setting, I provided an analytical method to calculate the MSE and the uplink spectral efficiency that enabled me to derive exact numerical results when the receiver at the base station employs LS or MMSE channel estimation and MMSE equalizer for uplink data reception. I found that with a large number of antennas, exploiting the engineering freedom of tuning *both* the number of pilot symbols and the pilot transmit power levels become increasingly important, especially if the relatively simple LS estimator is used at the base station. Also, the gain of using MMSE estimation (preferably with optimized pilot power allocation) increases over LS estimation. Interestingly, the optimal PPDR is different when using MMSE and LS estimators and the gain in terms of spectral efficiency when optimizing *both* the number of pilot symbols and the transmit power levels increases as the number of antennas increases. I believe that the proposed methodology as well as the obtained insights are new and provide useful guidelines for designing practical large antenna systems.

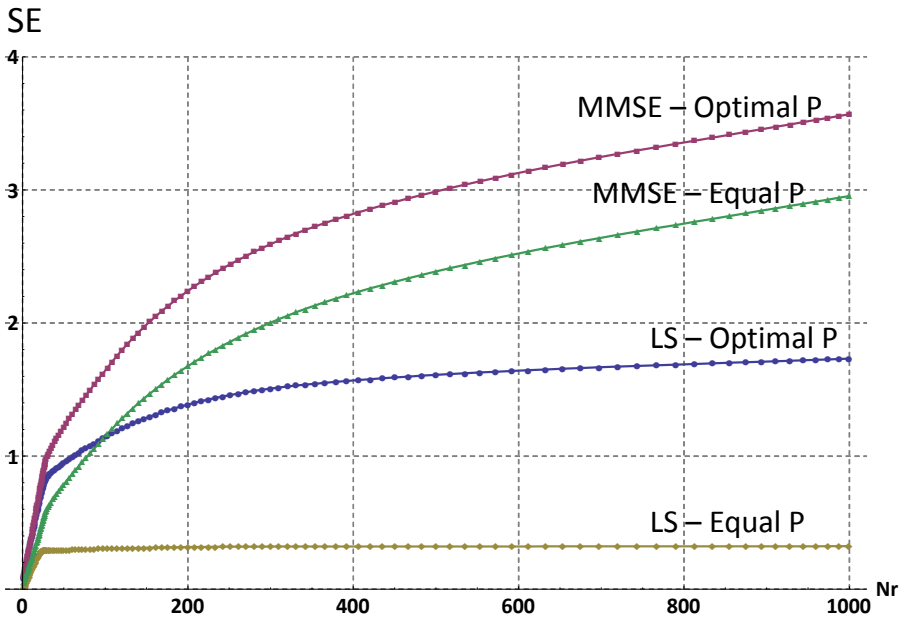

Fig. 6.4 The spectral efficiency (SE) as the function of the number of receive antennas at the BS, when employing LS (lower 2 curves) and MMSE (upper 2 curves) channel estimation. In both cases, optimum pilot power allocation is compared with equal power allocation between pilot and data transmission. With LS estimation, optimum pilot power allocation gives large gains, whereas with MMSE estimation, this spectral efficiency gain obtained by optimum pilot power allocation is less, although still significant.

Appendix of Chapter 6

Proof of Lemma 6.1

Given $\mathbf{h} \sim \mathcal{CN}(0, \mathbf{C})$, and $\hat{\mathbf{h}} \sim \mathcal{CN}(0, \mathbf{R})$, from (10.24)–(10.27) of [34], it follows that $\mathbf{h}|\hat{\mathbf{h}}$ is a complex normal distributed random vector with the following mean and covariance

$$\begin{aligned}\mathbb{E}(\mathbf{h}|\hat{\mathbf{h}}) &= \mathbb{E}(\mathbf{h}\hat{\mathbf{h}}^H)\mathbf{R}^{-1}\hat{\mathbf{h}},\\ \mathbf{C}_{\mathbf{h}|\hat{\mathbf{h}}} &= \mathbf{C} - \mathbb{E}(\mathbf{h}\hat{\mathbf{h}}^H)\mathbf{R}^{-1}\mathbb{E}(\hat{\mathbf{h}}\mathbf{h}^H).\end{aligned}$$

For the LS channel estimation model in (6.4), I derive

$$\mathbb{E}(\mathbf{h}\hat{\mathbf{h}}^H) = \mathbf{C} + \frac{1}{\alpha\sqrt{P_p}\tau_p}\mathbb{E}(\mathbf{h}\mathbf{s}^T\mathbf{N}^H) = \mathbf{C}.$$

For the MMSE channel estimation model in (6.7), I derive

$$\begin{aligned}\mathbb{E}(\mathbf{h}\hat{\mathbf{h}}^H) &= \left(\frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r} + \mathbf{C}\right)^{-1}\mathbf{C}\left(\mathbf{C} + \frac{1}{\alpha\sqrt{P_p}\tau_p}\mathbb{E}\left(\mathbf{h}\tilde{\mathbf{N}}^H\mathbf{S}\right)\right)\\ &= \left(\frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r} + \mathbf{C}\right)^{-1}\mathbf{C}^2 = \mathbf{C}^2\left(\frac{\sigma^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r} + \mathbf{C}\right)^{-1}.\end{aligned}$$

The expressions of $\mathbf{D}$ and $\mathbf{Q}$ in (6.13) are obtained by substitution.

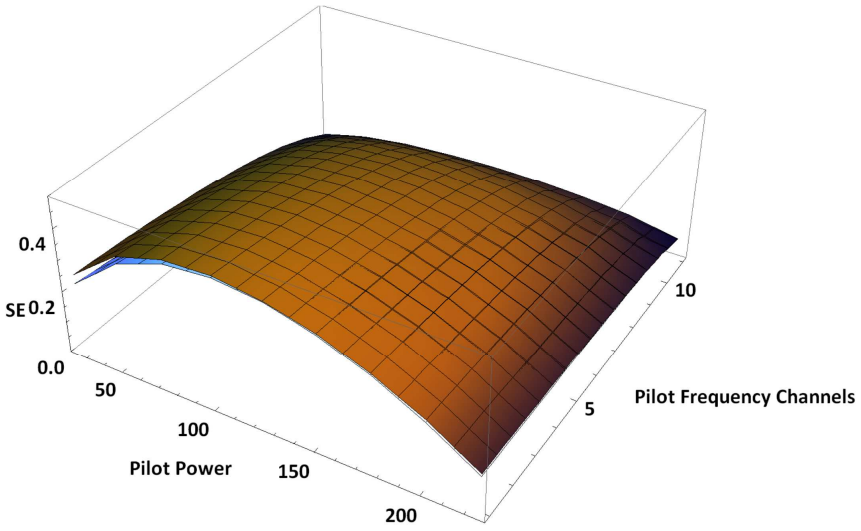

Fig. 6.5 Spectral efficiency with comb pilot arrangement and LS (lower) and MMSE (upper) channel estimation as a function of the number of frequency channels and the total pilot power (out of the P_{tot}) with $N_r = 10$ receive antennas. The pilot power that maximizes spectral efficiency is around $P_p^{opt} = 100$ mW with both LS and MMSE.

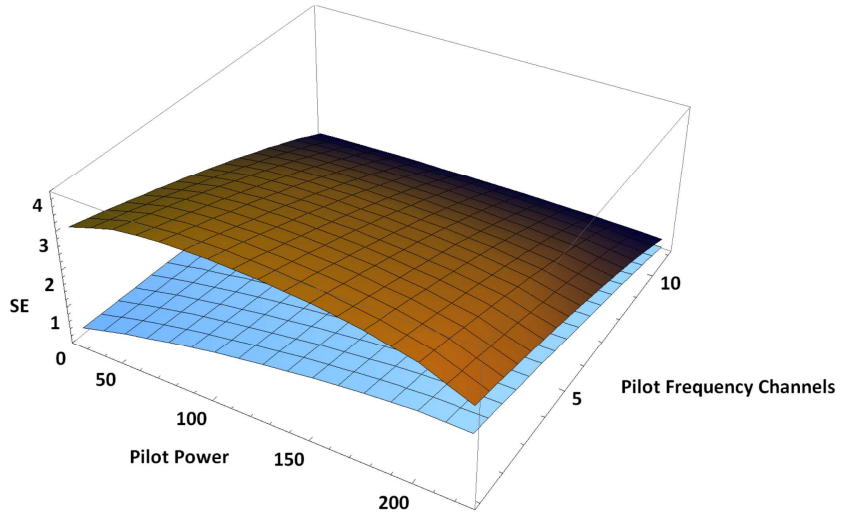

Fig. 6.6 Spectral efficiency with comb pilot arrangement and LS (lower) and MMSE (upper) channel estimation as a function of the number of frequency channels and the total pilot power with $N_r = 1000$ receive antennas. The pilot power that maximizes spectral efficiency is around $P_p^{opt} = 200$ mW with LS and 100 mW with MMSE estimation.

Proof of Lemma 6.2

Let us assume $\mathbf{C} = c\mathbf{I}_{N_r}$. In the LS estimation case, from (6.5) I have

$$\mathbf{R}_{LS} = \left(c + \frac{\sigma^2}{\alpha^2 P_p \tau_p}\right)\mathbf{I}_{N_r}$$

and therefore by using Lemma 6.1, $\mathbf{D} = d\mathbf{I}_{N_r}$ and $\mathbf{Q} = q\mathbf{I}_{N_r}$, where

$$d = c\left(c + \frac{\sigma^2}{\alpha^2 P_p \tau_p}\right)^{-1} \text{ and } \quad q = c - c^2\left(c + \frac{\sigma^2}{\alpha^2 P_p \tau_p}\right)^{-1}.$$

In the MMSE estimation case, from (6.8) I have

$$\mathbf{R}_{MMSE} = c^2\left(c + \frac{\sigma^2}{\alpha^2 P_p \tau_p}\right)^{-1}\mathbf{I}_{N_r},$$

and therefore by using Lemma 6.1, $\mathbf{D} = d\mathbf{I}_{N_r}$ and $\mathbf{Q} = q\mathbf{I}_{N_r}$, where

$$d = 1 \quad \text{and} \quad q = c - c^2\left(c + \frac{\sigma^2}{\alpha^2 P_p \tau_p}\right)^{-1}.$$

By replacing $\mathbf{D} = d\mathbf{I}_{N_r}$, $\mathbf{Q} = q\mathbf{I}_{N_r}$ and (6.12) in (6.14), I obtain

$$\begin{aligned}\text{MSE}(\hat{\mathbf{h}}) =& 1 - \frac{2\|\hat{\mathbf{h}}\|^2 d\alpha^2 P}{\|\hat{\mathbf{h}}\|^2\alpha^2 P + \sigma^2} + \frac{\|\hat{\mathbf{h}}\|^2\alpha^2 P}{\left(\|\hat{\mathbf{h}}\|^2\alpha^2 P + \sigma^2\right)^2} \\ & \times\left[\|\hat{\mathbf{h}}\|^2 d^2\alpha^2 P + q\alpha^2 P + \sigma^2\right] = \\ =& \frac{\left(\|\hat{\mathbf{h}}\|^2 p + \sigma^2\right)^2}{\left(\|\hat{\mathbf{h}}\|^2 p + \sigma^2\right)^2} - \frac{2\|\hat{\mathbf{h}}\|^2 dp(\|\hat{\mathbf{h}}\|^2 p + \sigma^2)}{\left(\|\hat{\mathbf{h}}\|^2 p + \sigma^2\right)^2} + \\ & + \frac{\|\hat{\mathbf{h}}\|^2 p(\|\hat{\mathbf{h}}\|^2 d^2 p + \overbrace{qp + \sigma^2}^{b})}{\left(\|\hat{\mathbf{h}}\|^2 p + \sigma^2\right)^2}.\end{aligned}$$

$$\text{MSE}(\hat{\mathbf{h}}) = \frac{p^2\|\hat{\mathbf{h}}\|^4(d-1)^2 + p\|\hat{\mathbf{h}}\|^2(2\sigma^2 - 2d\sigma^2 + b) + \sigma^4}{\left(\|\hat{\mathbf{h}}\|^2 p + \sigma^2\right)^2},$$

where $p = \alpha^2 P$, $b = qp + \sigma^2$.

Proof of Result 1

The key step to prove Result 1 is to derive the expectation of the log term in Eq. (6.16).

By considering the MSE in (6.15), the log term in Eq. (6.16) can be written as:

$$\begin{aligned}\log\left(\frac{1}{\mathrm{MSE}(\hat{\mathbf{h}})}\right) &= -\log\left(\mathrm{MSE}(\hat{\mathbf{h}})\right)\\ &= -\log\left(p^2Y^2(d-1)^2 + pY(2\sigma^2 - 2d\sigma^2 + b) + \sigma^4\right)\\ &\quad + 2\log\left(Yp + \sigma^2\right)\end{aligned}$$

where $Y = \|\hat{\mathbf{h}}\|^2$.

I therefore need to calculate the following expectation:

$$\begin{aligned}\mathbb{E}_{\hat{\mathbf{h}}}\left[-\log\left(\mathrm{MSE}(\hat{\mathbf{h}})\right)\right] &= 2\int_{x=0}^{\infty}\log\left(xp+\sigma^2\right)f_Y(x)dx \\ -\int_{x=0}^{\infty}\log&\left(p^2x^2(d-1)^2 + px(2\sigma^2 - 2d\sigma^2 + b) + \sigma^4\right)f_Y(x)dx,\end{aligned} \tag{6.20}$$

where the density function of Y – being the sum of exponentially distributed random variables of parameter r, where r is the diagonal element of $\mathbf{R}$ –, is given by [31]:

$$f_Y(x) = \frac{r^{-N_r}x^{N_r-1}e^{-x/r}}{(N_r-1)!} \quad x > 0.$$

Notice that for LS estimation, $d = c\left(c + \frac{\sigma^2}{\alpha^2 P_p \tau_p}\right)^{-1} \neq 1$, therefore the terms of the integral can be rearranged as follows

$$\begin{aligned}\mathbb{E}_{\hat{\mathbf{h}}}\left[-\log\left(\mathrm{MSE}(\hat{\mathbf{h}})\right)\right] &= 2\int_{x=0}^{\infty}\log\left(p\left(x+x_0\right)\right)f_Y(x)dx\\ &\quad -\int_{x=0}^{\infty}\log\left(p^2(d-1)^2(x^2+a_1x+a_0)\right)f_Y(x)dx\\ &= 2\int_{x=0}^{\infty}\log(x+x_0)f_Y(x)dx - \log(d-1)^2\\ &\quad -\int_{x=0}^{\infty}\log(x^2+a_1x+a_0)f_Y(x)dx\end{aligned} \tag{6.21}$$

where $x_0 = \frac{\sigma^2}{p}$, $a_0 = \frac{\sigma^4}{p^2(d-1)^2}$, and $a_1 = \frac{2\sigma^2-2d\sigma^2+b}{p(d-1)^2}$.

The last integral can be further simplified by considering

$$\log(x^2+a_1x+a_0) = \log(x-x_1) - \log(x-x_2),$$

where $x_{1,2} = 0.5\left(-a_1 \pm \sqrt{a_1^2 - 4a_0}\right)$.

In conclusion, I have to compute integrals of the form $\int_{x=0}^{\infty}\log\left(x+A\right)f_Y(x)dx$, which can be solved in Mathematica [55] via the Meijer G-function.

Specifically:

$$\begin{aligned}&\int_{x=0}^{\infty}\log(x+A)f_Y(x)dx = \\ &\frac{A^{N_r}}{(N_r-1)!r^{N_r}}\mathbf{MeijerG}_{2,3}^{1,3}\left(\begin{matrix}-N_r, -N_r+1\\ -N_r, -N_r, 0\end{matrix}\middle|\frac{A}{r}\right)\end{aligned} \tag{6.22}$$

where $A, N_r > 0, r > 1$ and $\mathbf{MeijerG}_{p,q}^{m,n}\left(\begin{matrix}a_1,\ldots,a_p\\ b_1,\ldots,b_q\end{matrix}\middle|z\right)$ is the Meijer G-function with parameters p, q, m, n.

Recognizing that

$$z^{N_r}\,\mathbf{MeijerG}_{2,3}^{1,3}\left(\begin{matrix}-N_r,-N_r+1\\-N_r,-N_r,0\end{matrix}\middle|z\right)$$
$$=\mathbf{MeijerG}_{2,3}^{1,3}\left(\begin{matrix}0,1\\0,0,N_r\end{matrix}\middle|z\right),$$

(6.22) is equivalent with:

$$\int_{x=0}^{\infty}\log(x+A)f_Y(x)dx=\frac{\mathbf{MeijerG}_{2,3}^{1,3}\left(\begin{matrix}0,1\\0,0,N_r\end{matrix}\middle|z\right)}{(N_r-1)!} \tag{6.23}$$

where $z=\frac{A}{r}$.
By substituting the result of (6.23) in (6.21), (6.17) follows.

Proof of Result 2

The proof follows with similar steps as for Result 1. However, in the MMSE case, I have $d=1$, and the expectation in (6.20) can be conveniently rewritten as

$$\begin{aligned}&\mathbb{E}_{\hat{\mathbf{h}}}\left[-\log\left(\mathrm{MSE}(\hat{\mathbf{h}})\right)\right]=\\&=2\int_{x=0}^{\infty}\log\left(xp+\sigma^2\right)f_Y(x)dx-\int_{x=0}^{\infty}\log\left(pxb+\sigma^4\right)f_Y(x)dx\\&=\log(pb)+2\int_{x=0}^{\infty}\log\left(x+x_3\right)f_Y(x)dx-\int_{x=0}^{\infty}\log\left(x+x_4\right)f_Y(x)dx,\end{aligned}$$

with $x_3=\frac{\sigma^2}{p}$, $x_4=\frac{\sigma^2}{pb}$, $b=qp+\sigma^2$, $q=\frac{\sigma^2 c}{\sigma^2+\alpha^2 c P_p\tau_p}$,

which is solved in Mathematica by using the Meijer G-function defined in (6.18).

Chapter 7
The Pilot-to-Data Power Ratio in Multiuser Systems

7.1 Introduction

The previous chapters suggest that for the purpose of determining the optimal pilot power setting, it is important to take into account the operation of practical channel estimation and receiver algorithms. To the best of my knowledge, exact expressions for the achieved MSE and spectral efficiency (SE) when using practical channel estimation such as LS and receiver algorithms such as MMSE, and accounting for the PDPR and antenna correlation, are not available. In this chapter, I address this problem and derive closed form expressions for the uplink of a MU-MIMO system, in which the BS uses LS or MMSE channel estimation and MMSE receiver. Throughout, I assume that the output of the MMSE detector, the residual signal plus interference from other spatial streams as well as the estimation error of the received data symbols can be approximated as Gaussian [41]. Because in practice the CSI estimation error is likely to be bounded, my design can be regarded as a worst-case design approach. Thereby, my contributions (detailed in Sections 7.3-7.6 and the Appendices) to the lines of works above can be summarized as follows:

- I derive closed form exact expressions for both the MSE and the SE taking into account the CSI errors that are specific to the employed channel estimation technique;
- I explicitly take into account the impact of antenna correlation on these performance measures.

These formulas are then used to compare the performance of MU-MIMO systems employing the naïve and MMSE receivers. An interesting insight is that when the system uses the MMSE receiver, the PDPR minimizing the MSE does not depend on the number of receive antennas at the BS but rather is dependent on the large-scale fading. This is in contrast to a system that employs the naïve receiver, for which the pilot power minimizing the MSE depends on the number of receive antennas. This insight can help set the pilot power almost optimally in practical systems in which the number of BS antennas can depend on the actual deployment scenario [56], [57]. In particular, my results show that when the optimal pilot power setting is employed at the terminal side, and the true MMSE receiver is used at the base station side, the system's performance is close to that of a hypothetical system that would have access to the perfect CSI.

This chapter is structured as follows. Section 7.2 describes the system model and summarizes preliminaries needed for development of the contributions of this chapter. Sections 7.3 and 7.4 analyze the MSE in the case of uncorrelated and correlated antennas at the receiver, respectively. Section 7.5 derives closed form expressions for the MSE and SE when the receiver uses the MMSE receiver. Section 7.6 presents numerical results on the MSE and SE, and Section 7.7 concludes the chapter.

7.2 Channel Estimation and Receiver Model

7.2.1 Channel Estimation Model

I consider the uplink of a MU-MIMO system, in which the MSs transmit orthogonal pilot sequences of length τ_p: $\mathbf{s} = \left[s_1, ..., s_{\tau_p}\right]^T \in \mathbb{C}^{\tau_p \times 1}$, in which each pilot symbol is scaled as $|s_i|^2 = 1$, for $i = 1, .., \tau_p$. The pilot sequences are constructed such that they remain orthogonal as long as the number of spatially

multiplexed users is maximum τ_p [33]. In practice, such pilot sequences can be defined using the popular Zadoff-Chu sequences [53],[58]. Specifically, without loss of generality, assume that the number MU-MIMO users is $K \leq \tau_p$. In practice, $K \ll N_r$, where N_r is the number of antennas at the BS [27].

As emphasized in [27], MU MIMO differs from point-to-point MIMO in two respects: first, the terminals are typically separated by many wavelengths, and second, the terminals cannot collaborate among themselves, either to transmit or to receive data. That is, in MU MIMO systems, the terminals are autonomous so that I can assume that the transmit array is uncorrelated. However, it is important to capture the correlation structure at the receiver side so that I can evaluate the impact of CSIR errors on the optimal pilot power and the achieved MSE.

In this chapter I assume a comb type arrangement of the pilot symbols. Given F subcarriers in the coherence bandwidth, a fraction of τ_p subcarriers are allocated to the pilot and $\tau_d = F - \tau_p$ subcarriers are allocated to the data symbols. Each MS transmits at a constant power P_{tot}, however, this transmission power can be distributed unequally among the subcarriers. In particular, considering User-ℓ with a transmitted power $P_{p,\ell}$ for each pilot symbol and P_ℓ for each data symbol transmission, the sum constraint of:

$$\tau_p P_{p,\ell} + \tau_d P_\ell = P_{tot} \tag{7.1}$$

is enforced. In practice, this type of arrangement is suitable for time varying channels, so that channel estimation is facilitated at the same time instant that is used for data transmission. Thus, the $N_r \times \tau_p$ matrix of the received pilot signal from User-ℓ at the BS can be conveniently written as:

$$\mathbf{Y}_\ell^p = \alpha_\ell \sqrt{P_{p,\ell}} \mathbf{h}_\ell \mathbf{s}^T + \mathbf{N}, \tag{7.2}$$

where I assume that $\mathbf{h}_\ell \in \mathbb{C}^{N_r \times 1}$ is a circular symmetric complex normal distributed column vector with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{C}_\ell$ (of size N_r), denoted as $\mathbf{h}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_\ell)$, α_ℓ accounts for the large scale fading, $\mathbf{N} \in \mathbb{C}^{N_r \times \tau_p}$ is the spatially and temporally AWGN with element-wise variance σ_p^2, where the index p refers to the noise power on the received pilot signal.

In this chapter I assume that the BS uses the popular LS estimator that relies on correlating the received signal with the known pilot sequence. Note that the proposed methodology to determine the MSE of the received data is not confined to the LS estimator, but is directly applicable to an MMSE or other linear channel estimation techniques as well. For each MS, the BS utilizes pilot sequence orthogonality and estimates the channel based on (7.2) assuming:

$$\begin{aligned} \hat{\mathbf{h}}_\ell = \mathbf{h}_\ell + \mathbf{w}_\ell &= \frac{1}{\alpha_\ell \sqrt{P_{p,\ell}}} \mathbf{Y}_\ell^p \mathbf{s}^* (\mathbf{s}^T \mathbf{s}^*)^{-1} \\ &= \mathbf{h}_\ell + \frac{1}{\alpha_\ell \sqrt{P_{p,\ell} \tau_p}} \mathbf{N} \mathbf{s}^*, \end{aligned} \tag{7.3}$$

where $\mathbf{s}^* = [s_1^*, ..., s_{\tau_p}^*]^T \in \mathbb{C}^{\tau_p \times 1}$ denotes the vector of pilot symbols and $(\mathbf{s}^T \mathbf{s}^*) = \tau_p$. By considering $\mathbf{h}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_\ell)$, it follows that the estimated channel $\hat{\mathbf{h}}_\ell$ is a circular symmetric complex normal distributed vector $\hat{\mathbf{h}}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_\ell)$, with

$$\mathbf{R}_\ell \triangleq \mathbb{E}\{\hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H\} = \mathbf{C}_\ell + \frac{\sigma_p^2}{\alpha_\ell^2 P_{p,\ell} \tau_p} \mathbf{I}_{N_r}. \tag{7.4}$$

By recognizing that $\mathbf{h}$ and $\hat{\mathbf{h}}$ are jointly circular symmetric complex Gaussian (multivariate normal) distributed random variables, the distribution of the channel realization $\mathbf{h}_\ell$ conditioned on the estimate $\hat{\mathbf{h}}_\ell$ is normally distributed as follows [34], [31]:

$$(\mathbf{h}_\ell \mid \hat{\mathbf{h}}_\ell) \sim \mathcal{CN}\left(\mathbf{D}_\ell \hat{\mathbf{h}}_\ell, \mathbf{Q}_\ell\right), \tag{7.5}$$

where $\mathbf{D}_\ell \triangleq \mathbf{C}_\ell \mathbf{R}_\ell^{-1}$ and $\mathbf{Q}_\ell \triangleq \mathbf{C}_\ell - \mathbf{C}_\ell \mathbf{R}_\ell^{-1} \mathbf{C}_\ell$.

7.2.2 Received Data Signal Model

The MU-MIMO received data signal at the BS can be written as:

$$\mathbf{y} = \underbrace{\alpha_\ell \mathbf{h}_\ell \sqrt{P_\ell} x_\ell}_{\text{User-}\ell} + \underbrace{\sum_{k \neq \ell}^{K} \alpha_k \mathbf{h}_k \sqrt{P_k} x_k}_{\text{Other users}} + \mathbf{n}_d, \tag{7.6}$$

where α_k $\mathbf{h}_k$ is the $M \times 1$ vector channel including large and small scale fading between User-k and the BS, P_k is the data transmit power of User-k, x_k is the transmitted data symbol by User-k and $\mathbf{n}_d$ denotes the Gaussian noise on the received data signal.

7.2.3 Employing a Minimum Mean Squared Error Receiver at the Base Station

In this chapter the BS employs an MMSE receiver $\mathbf{G}_\ell \in \mathbb{C}^{1 \times N_r}$ to estimate the data symbol transmitted by User-ℓ. As it was shown in [59], in the case of a linear receiver $\mathbf{G}_\ell$ that requires the estimated channel of only User-ℓ as its input, the MSE of the estimated data symbols of User-ℓ can be conveniently expressed in the following quadratic form:

$$\begin{aligned} \text{MSE}\left(\mathbf{G}_\ell, \hat{\mathbf{h}}_\ell\right) = \mathbf{G}_\ell \Bigg(\alpha_\ell^2 P_\ell \left(\mathbf{D}_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H + \mathbf{Q}_\ell \right) + \sum_{k \neq \ell}^{K} \alpha_k^2 P_k \mathbf{C}_k + \sigma_d^2 \mathbf{I} \Bigg) \mathbf{G}_\ell^H - \\ - \alpha_\ell \sqrt{P_\ell} (\mathbf{G}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell + \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{G}_\ell^H) + 1. \end{aligned} \tag{7.7}$$

As we shall see later, my analysis allows for an arbitrary channel covariance matrix at the receiver side ($\mathbf{C}_\ell$) in (7.7) that allows me to analyze the impact of CSI errors on the MSE performance with arbitrary correlation structure of the base station antennas. Recall that the MMSE receiver aims at minimizing the MSE between the estimate $\mathbf{G}_\ell \mathbf{y}$ and the transmitted symbol x_ℓ:

$$\mathbf{G}_\ell^\star \triangleq \arg\min_{\mathbf{G}} \mathbb{E}\{\text{MSE}\} = \arg\min_{\mathbf{G}} \mathbb{E}\{|\mathbf{G}_\ell \mathbf{y} - x_\ell|^2\}. \tag{7.8}$$

When the BS employs a naïve receiver, the estimated channel is taken as if it was the actual channel:

$$\mathbf{G}_\ell^{\text{naïve}} = \alpha_\ell \sqrt{P_\ell} \hat{\mathbf{h}}_\ell^H (\alpha_\ell^2 P_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H + \sigma_d^2 \mathbf{I})^{-1}. \tag{7.9}$$

As it was shown in [59], this receiver does not minimize the MSE. Using the quadratic form in (7.7), it can be shown that the receiver that minimizes the MSE of the received data symbols, is constructed as:

$$\begin{aligned} \mathbf{G}_\ell^\star = \alpha_\ell \sqrt{P_\ell} \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \cdot \\ \cdot \left(\alpha_\ell^2 P_\ell \left(\mathbf{D}_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H + \mathbf{Q}_\ell \right) + \sum_{k \neq \ell}^{K} \alpha_k^2 P_k \mathbf{C}_k + \sigma_d^2 \mathbf{I} \right)^{-1}. \end{aligned} \tag{7.10}$$

7.2.4 Calculating the Mean Squared Error When Employing the Minimum Mean Squared Error Receiver

In [59] it was shown that for the special case when the channel covariance matrices $\mathbf{C}_\ell$ and consequently the matrices $\mathbf{R}_\ell$, $\mathbf{D}_\ell$ and $\mathbf{Q}_\ell$ are proportional to the identity matrix $\mathbf{I}_{N_r}$ with diagonal elements c_ℓ, r_ℓ, d_ℓ and q_ℓ respectively, the unconditional MSE_ℓ of the uplink estimated data symbols of User-ℓ when employing the $\mathbf{G}_\ell^\star$ receiver can be calculated as follows.

Theorem 7.2.1 *The unconditional* MSE_ℓ *of the received data symbols of User-*ℓ *when the BS uses the optimal* $\mathbf{G}_\ell^\star$ *receiver is as follows:*

$$\begin{aligned}\mathrm{MSE}_\ell = & \frac{b_\ell\left(e^{\frac{b_\ell}{s_\ell r_\ell}}\left(b_\ell + N_r s_\ell r_\ell\right)E_{in}\left(N_r, \frac{b_\ell}{s_\ell r_\ell}\right) - s_\ell r_\ell\right)}{s_\ell^2 r_\ell^2} \\ & + \frac{N_r\left(e^{\frac{b_\ell}{s_\ell r_\ell}}\left(b_\ell + (1+N_r)s_\ell r_\ell\right)E_{in}\left(1+N_r, \frac{b_\ell}{s_\ell r_\ell}\right) - s_\ell r_\ell\right)}{s_\ell r_\ell} \\ & - 2\cdot e^{\frac{b_\ell}{s_\ell r_\ell}} N_r E_{in}\left(1+N_r, \frac{b_\ell}{s_\ell r_\ell}\right) + 1, \end{aligned} \tag{7.11}$$

where $E_{in}(n,z) \triangleq \int_1^\infty e^{-zt}/t^n\, dt$ *is a standard exponential integral function,* $s_\ell \triangleq d_\ell^2 p_\ell$, $b_\ell \triangleq q_\ell p_\ell + \sigma_d^2$ *with* $p_\ell \triangleq \alpha_\ell^2 P_\ell$.

The proof is in the Appendix of this chapter.

Notice that specifically in the case of LS channel estimation and when $\mathbf{C}_\ell$ is of the form of $c_\ell \mathbf{I}_{N_r}$, from (7.4)-(7.5) I have:

$$r_\ell = c_\ell + \frac{\sigma_p^2}{\alpha_\ell^2 P_{p,\ell}\tau_p}; \quad d_\ell = \frac{c_\ell}{r_\ell}; \quad q_\ell = c_\ell - c_\ell d_\ell. \tag{7.12}$$

7.3 Analysis of the Mean Squared Error in the Case of Uncorrelated Antennas

This section presents the optimal pilot power setting for the case when $\mathbf{C}_\ell$ is proportional to the identity matrix that is $\mathbf{C}_\ell = c_\ell \mathbf{I}_{N_r}$. I start with a further simplified version of Proposition 7.2.1.

Lemma 7.3.1 *When the BS uses the optimal* $\mathbf{G}_\ell^\star$ *receiver and the channel can be assumed* $\mathbf{C}_\ell = c_\ell \mathbf{I}_{N_r}$, *the* MSE *of the estimated data symbols of each user can be calculated as follows:*

$$MSE(\mu_\ell) = \mu_\ell e^{\mu_\ell} E_{in}(N_r, \mu_\ell), \tag{7.13}$$

where $\mu_\ell = \mu(P_{p,\ell})$ *is defined by*

$$\mu_\ell \triangleq \frac{\sigma_d^2\sigma_p^2\tau_d + c_\ell\alpha_\ell^2\left(\sigma_p^2 P_{tot} + \tau_p P_{p,\ell}(\sigma_d^2\tau_d - \sigma_p^2)\right)}{c_\ell^2\alpha_\ell^4 P_{p,\ell}\tau_p(P_{tot} - \tau_p P_{p,\ell})}. \tag{7.14}$$

The proof is in the Appendix II.

As it was underscored by [1] and [4], there is a gap in spectral efficiency between coherent and noncoherent communications and channel learning plays an important role in bridging this gap. Lemma 7.3.1 captures the training cost ($\tau_p P_{p,\ell}$) of communicating over an unknown channel specifically in the case of an uplink of MU MIMO system employing an MMSE receiver.

For the naïve ($\mathbf{G}_\ell^{\text{naïve}}$) as well as for the optimal MMSE receiver ($\mathbf{G}_\ell^\star$), it is important to find the pilot power that minimizes the MSE. For the naïve receiver I obtain the optimal pilot power by numerical optimization, whereas for the optimal MMSE receiver the pilot power that minimizes the MSE has a closed form expression. The following proposition presents the optimal PDPR as a function of the total power and coherence budget and the large scale fading between the MS and the BS.

Theorem 7.3.2 *When employing the MMSE receiver $\mathbf{G}_\ell^\star$, in the case of $\mathbf{C}_\ell = c_\ell \mathbf{I}_{N_r}$, the pilot power that minimizes the MSE is independent of the number of receive antennas N_r and is given by:*

$$P_{p,\ell}^\star = \frac{\sigma_d \sigma_p \sqrt{(c_\ell P_{tot}\alpha_\ell^2 + \sigma_p^2)(c_\ell P_{tot}\alpha_\ell^2 + \sigma_d^2\tau_d)\tau_d}}{c_\ell \alpha_\ell^2 \tau_p (\sigma_d^2\tau_d - \sigma_p^2)} - \frac{\sigma_p^2(c_\ell P_{tot}\alpha_\ell^2 + \sigma_d^2\tau_d)}{c_\ell \alpha_\ell^2 \tau_p (\sigma_d^2\tau_d - \sigma_p^2)}. \tag{7.15}$$

The proof is in the Appendix of this chapter.

Remark 7.1. In the case of $\sigma_d = \sigma_p = \sigma$, expression (7.15) can be further simplified as

$$P_{p,\ell}^* = P_{tot}\left(\frac{\sqrt{\left(1 + \frac{\sigma^2}{c_\ell P_{tot}\alpha_\ell^2}\tau_d\right)\left(1 + \frac{\sigma^2}{c_\ell P_{tot}\alpha_\ell^2}\right)\tau_d}}{\tau_p(\tau_d - 1)} - \frac{\left(1 + \frac{\sigma^2}{c_\ell P_{tot}\alpha_\ell^2}\tau_d\right)}{\tau_p(\tau_d - 1)} \right).$$

The optimal pilot power is a fraction of the power budget P_{tot} that depends on the number of pilot symbols τ_p and data symbols τ_d. It is also easy to verify that in the case of perfect channel knowledge (i.e., assuming $\sigma_p = 0$), expression (7.15) returns $P_{p,\ell}^* = 0$.

7.4 Analysis of the Mean Squared Error in the Case of Correlated Antennas

7.4.1 Determining $\mathbf{G}^\star$

I now consider the general case when the channel covariance matrices ($\mathbf{C}_k$) are not diagonal, that is when allowing for an arbitrary correlation structure between the BS antennas. I assume that the BS employs the optimal MMSE equalizer according to (7.10) and write

$$\mathbf{G}_\ell^\star = \alpha_\ell \sqrt{P_\ell}\hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \left(\mathbf{\Psi}_\ell + \alpha_\ell^2 P_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H\right)^{-1}, \tag{7.16}$$

where

$$\mathbf{\Psi}_\ell \triangleq \alpha_\ell^2 P_\ell \mathbf{Q}_\ell + \sum_{k\neq\ell}^{K} \alpha_k^2 P_k \mathbf{C}_k + \sigma_d^2 \mathbf{I}_{N_r}, \tag{7.17}$$

is a positive definite matrix which contains the covariance from all intra- and intercell interference sources that cause interference to the signal of User-ℓ and the self covariance term related with $\mathbf{Q}_\ell$.

For an explicit inversion in (7.16), I introduce the SVD of $\mathbf{\Psi}_\ell$, that is $\mathbf{\Psi}_\ell = \mathbf{\Theta}_\ell^H \mathbf{S}_\ell \mathbf{\Theta}_\ell$. Since $\mathbf{\Psi}_\ell$ is positive definite, it is non singular and I can therefore define:

$$\boldsymbol{\nu}_\ell \triangleq \mathbf{S}_\ell^{-1/2} \mathbf{\Theta}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell, \tag{7.18}$$

which is a linear transformed version of $\hat{\mathbf{h}}_\ell$. It will be useful to notice that:

$$\hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{\Psi}_\ell^{-1} = \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{\Theta}_\ell^H \mathbf{S}_\ell^{-1} \mathbf{\Theta}_\ell = \boldsymbol{\nu}_\ell^H \mathbf{S}_\ell^{-1/2} \mathbf{\Theta}_\ell, \tag{7.19}$$

and

$$\hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell^H \mathbf{\Psi}_\ell^{-1} \mathbf{D}_\ell \hat{\mathbf{h}}_\ell = ||\boldsymbol{\nu}_\ell||^2, \tag{7.20}$$

and note that from (7.18) I have

$$\mathbf{D}_\ell \hat{\mathbf{h}}_\ell = \mathbf{\Theta}_\ell^H \mathbf{S}_\ell^{1/2} \boldsymbol{\nu}_\ell. \tag{7.21}$$

With these notations, it is straightforward to prove the following useful lemma.

Lemma 7.4.1 *Given a channel estimate instance $\hat{\mathbf{h}}_\ell$, the MMSE weight matrix $\mathbf{G}_\ell$, as a function of the number of receive antennas at the BS (N_r) can be expressed as follows:*

$$\mathbf{G}_\ell^\star = \frac{\alpha_\ell \sqrt{P_\ell}}{\alpha_\ell^2 P_\ell ||\boldsymbol{\nu}_\ell||^2 + 1} \boldsymbol{\nu}_\ell^H \mathbf{S}_\ell^{-1/2} \mathbf{\Theta}_\ell, \tag{7.22}$$

where $||\boldsymbol{\nu}_\ell||^2 = \boldsymbol{\nu}_\ell^H \boldsymbol{\nu}_\ell = \sum_{i=1}^{N_r} |\nu_{\ell_i}|^2$.

The proof is in Appendix IV.
To simplify the discussion, I introduce

$$g_\ell \triangleq \frac{\alpha_\ell \sqrt{P_\ell}}{\alpha_\ell^2 P_\ell ||\boldsymbol{\nu}_\ell||^2 + 1}. \tag{7.23}$$

7.4.2 Determining the Mean Squared Error When Using $\mathbf{G}^\star$

To determine the MSE, I first need to find the distribution of $\boldsymbol{\nu}_\ell$. The distribution of $\boldsymbol{\nu}_\ell$ is readable from (7.18) (notice that $\mathbf{\Psi}_\ell$ and thereby $\mathbf{S}_\ell$ are not random variables), and recall that $\hat{\mathbf{h}}_\ell$ is complex normal distributed with, $\hat{\mathbf{h}}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_\ell)$. Therefore, for $\boldsymbol{\nu}_\ell$ I have

$$\boldsymbol{\nu}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_\ell), \tag{7.24}$$

where

$$\begin{aligned}
\mathbf{\Omega}_\ell &\triangleq \mathbb{E}(\boldsymbol{\nu}_\ell \boldsymbol{\nu}_\ell^H) = \mathbb{E}\left((\mathbf{S}_\ell^{-1/2} \mathbf{\Theta}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell)(\mathbf{S}_\ell^{-1/2} \mathbf{\Theta}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell)^H \right) \\
&= \mathbf{S}_\ell^{-1/2} \mathbf{\Theta}_\ell \mathbf{D}_\ell \mathbb{E}\left(\hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H \right) \mathbf{D}_\ell^H \mathbf{\Theta}_\ell^H \mathbf{S}_\ell^{-1/2} \\
&= \mathbf{S}_\ell^{-1/2} \mathbf{\Theta}_\ell \mathbf{D}_\ell \mathbf{R}_\ell \mathbf{D}_\ell^H \mathbf{\Theta}_\ell^H \mathbf{S}_\ell^{-1/2}.
\end{aligned}$$

I will need the SVD of $\mathbf{\Omega}_\ell$:

$$\mathbf{\Omega}_\ell = \mathbf{\Theta}_{\mathbf{\Omega}_\ell}^H \mathbf{S}_{\mathbf{\Omega}\ell} \mathbf{\Theta}_{\mathbf{\Omega}\ell}, \tag{7.25}$$

where $\mathbf{\Theta}_{\mathbf{\Omega}\ell}$ is an orthogonal matrix ($\mathbf{\Theta}_{\mathbf{\Omega}_\ell}^H \mathbf{\Theta}_{\mathbf{\Omega}_\ell} = \mathbf{I}_{N_r}$).

Furthermore, I will need the linear transform of $\boldsymbol{\nu}_\ell$, which I denote with $\boldsymbol{\omega}_\ell$ whose covariance matrix is diagonal:

$$\boldsymbol{\omega}_\ell \triangleq \boldsymbol{\Theta}_{\boldsymbol{\Omega}\ell}\boldsymbol{\nu}_\ell \ . \tag{7.26}$$

Notice that (for ease of notation dropping the index ℓ):

$$||\boldsymbol{\omega}||^2 = \boldsymbol{\omega}^H\boldsymbol{\omega} = \boldsymbol{\nu}^H\boldsymbol{\Theta}_{\boldsymbol{\Omega}}^H\boldsymbol{\Theta}_{\boldsymbol{\Omega}}\boldsymbol{\nu} = \boldsymbol{\nu}^H\boldsymbol{\nu} = ||\boldsymbol{\nu}||^2 \tag{7.27}$$

and

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\omega}}\left(\boldsymbol{\omega}\boldsymbol{\omega}^H\right) = \mathbb{E}_{\boldsymbol{\nu}}\left(\boldsymbol{\Theta}_{\boldsymbol{\Omega}}\boldsymbol{\nu}\boldsymbol{\nu}^H\boldsymbol{\Theta}_{\boldsymbol{\Omega}}^H\right) = \boldsymbol{\Theta}_{\boldsymbol{\Omega}}\mathbb{E}_{\boldsymbol{\nu}}\left(\boldsymbol{\nu}\boldsymbol{\nu}^H\right)\boldsymbol{\Theta}_{\boldsymbol{\Omega}}^H = \\ = \boldsymbol{\Theta}_{\boldsymbol{\Omega}}\boldsymbol{\Theta}_{\boldsymbol{\Omega}}^H\mathbf{S}_{\boldsymbol{\Omega}}\boldsymbol{\Theta}_{\boldsymbol{\Omega}}\boldsymbol{\Theta}_{\boldsymbol{\Omega}}^H = \mathbf{S}_{\boldsymbol{\Omega}}.\end{aligned} \tag{7.28}$$

It is now straightforward to prove the following lemma.

Lemma 7.4.2 *The MSE of User-ℓ as a function of $\boldsymbol{\nu}_\ell$, as defined in* (7.18)*, is as follows.*

$$MSE\left(\boldsymbol{\nu}_\ell\right) = \frac{1}{\alpha_\ell^2 P_\ell ||\boldsymbol{\nu}_\ell||^2 + 1} \tag{7.29}$$

The proof is in Appendix V.

Remark 7.2. It is insightful to compare (7.29) with the MSE of a system with uncorrelated antennas and perfect channel estimation, that is when $\mathbf{D} = \mathbf{I}$, $\mathbf{Q} = \mathbf{0}$ and $\hat{\mathbf{h}} = \mathbf{h}$. In this case, from (7.50) I get:

$$\mathrm{MSE}\left(\mathbf{h}\right) = \frac{1}{\alpha_\ell^2 P_\ell \frac{||\mathbf{h}_\ell||^2}{\sigma_d^2} + 1}, \tag{7.30}$$

which indicates that ν_ℓ can be seen as an "equivalent channel" in the system of correlated antennas and partial CSI information, that is ν_ℓ captures the impact of both antenna correlation and CSI estimation errors on the MSE.

7.5 Calculating the Unconditional Mean Squared Error and the Spectral Efficiency

Recall from Lemma 7.4.1 that $||\boldsymbol{\nu}_\ell||^2 = \boldsymbol{\nu}_\ell^H\boldsymbol{\nu}_\ell = \sum_{i=1}^{N_r}|\nu_{\ell_i}|^2$, where the ν_{ℓ_i}-s ($i = 1, \ldots, N_r$) are, in general, not independent random variables. However, according to (7.28), the covariance matrix of $\boldsymbol{\omega}_\ell$ – that is $\mathbf{S}_{\boldsymbol{\Omega}}$ – is diagonal, with not necessarily equal diagonal elements. Therefore, each $|\omega_{\ell_i}|^2$ (denoted by $|\omega_i|^2$ in the sequel) is exponentially distributed. Building on this observation, I can prove the following theorem, which is useful for determining the MSE and the spectral efficiency.

Theorem 7.5.1 *The mean squared error of the received data symbols and the spectral efficiency can be calculated as:*

$$\mathrm{MSE} = \int_x \frac{1}{\alpha_\ell^2 P_\ell x + 1} f(x)dx, \tag{7.31}$$

and

$$\eta = -\int_x \log\left[\frac{1}{\alpha_\ell^2 P_\ell x + 1}\right] f(x)dx, \tag{7.32}$$

respectively.

Proof. Denote the variance of ω_i with ξ_i^2, and note that $|\omega_i|^2$ is exponentially distributed with parameter $\lambda_i = 1/\xi_i^2$. Therefore, $\sum_{i=1}^{N_r} |\omega_i|^2$ is the sum of N_r independent exponentially distributed random variables. The set of distributions composed by independent exponentially distributed phases are referred to as phase type distributions [60] and has a closed form description with matrix exponential functions. That is, the density function of $\sum_{i=1}^{N_r} |\omega_i|^2$ is

$$f(x) = e_1^T e^{\mathbf{A}x} e_{N_r} \lambda_{N_r}, \tag{7.33}$$

where e_i is the i-th unit vector (whose only nonzero element is 1 at position i) and the matrix $\mathbf{A}$ is:

$$\mathbf{A} = \begin{pmatrix} -\lambda_1 & \lambda_1 & & & \\ & -\lambda_2 & \lambda_2 & & \\ & & \ddots & \ddots & \\ & & & & -\lambda_{N_r} \end{pmatrix}. \tag{7.34}$$

Using $f(x)$ and (7.29), I have:

$$\begin{aligned} \text{MSE} &= \mathbb{E}_{\boldsymbol{\nu}} \left(\text{MSE}(\boldsymbol{\nu}) \right) = \mathbb{E}_{\boldsymbol{\omega}} \left(\text{MSE}(\boldsymbol{\omega}) \right) = \\ &= \int_x \frac{1}{\alpha_\ell^2 P_\ell x + 1} f(x) dx, \end{aligned} \tag{7.35}$$

and:

$$\begin{aligned} \eta &= -\mathbb{E}_{\boldsymbol{\nu}} \left(\log \left[\text{MSE}(\boldsymbol{\nu}) \right] \right) = -\mathbb{E}_{\boldsymbol{\omega}} \left(\log \left[\text{MSE}(\boldsymbol{\omega}) \right] \right) \\ &= -\int_x \log \left[\frac{1}{\alpha_\ell^2 P_\ell x + 1} \right] f(x) dx, \end{aligned} \tag{7.36}$$

which completes the proof.

This general case simplifies to the following two special ones.

7.5.1 Case 1: Distinct Variances

I will now assume that N is the number of non-zero singular values in $\mathbf{S}_{\boldsymbol{\Omega}}$ and all non-zero ξ_i (and λ_i) are distinct (different). In this special but in practice important case, I can state the following proposition.

Proposition 7.5.2 *If the non-zero variances $\xi_i^2 \neq 0$ are distinct, the* MSE *and the spectral efficiency can be calculated as*

$$\text{MSE} = \sum_{i=1}^{N} \frac{-\lambda_i^{\frac{-N}{2}} e^{\frac{\lambda_i}{p}} E_{in}\left(1, \frac{-\lambda_i}{p}\right)}{p \prod_{j=1, j\neq i}^{N} \left(1 - \frac{\lambda_i}{\lambda_j}\right)}, \tag{7.37}$$

and

$$\eta = \sum_{i=1}^{N} \frac{-\lambda_i^{\frac{2-N}{2}} e^{\frac{\lambda_i}{p}} E_{\text{in}}\left(1, \frac{-\lambda_i}{p}\right)}{\prod_{j=1, j\neq i}^{N} \left(1 - \frac{\lambda_i}{\lambda_j}\right)}, \tag{7.38}$$

respectively.

Proof. Notice that when all non-zero ξ_i, and consequently all λ_i are distinct (different), $f(x)$ simplifies to:

$$f(x) = \sum_{i=1}^{N} \frac{\lambda_i e^{-\lambda_i x}}{\prod_{j=1, j\neq i}^{N} \left(1 - \frac{\lambda_i}{\lambda_j}\right)}. \tag{7.39}$$

Therefore, using (7.31), for the MSE I get:

$$\text{MSE} = \sum_{i=1}^{N} \frac{-\lambda_i^{\frac{-N}{2}} e^{\frac{\lambda_i}{p}} E_{\text{in}}\left(1, \frac{-\lambda_i}{p}\right)}{p \prod_{j=1, j\neq i}^{N} \left(1 - \frac{\lambda_i}{\lambda_j}\right)}, \tag{7.40}$$

where recall from Proposition 7.2.1 that $p = \alpha^2 P_\ell$.

Similarly, using (7.32), for the SE I get:

$$\eta = \sum_{i=1}^{N} \frac{-\lambda_i^{\frac{2-N}{2}} e^{\frac{\lambda_i}{p}} E_{\text{in}}\left(1, \frac{-\lambda_i}{p}\right)}{\prod_{j=1, j\neq i}^{N} \left(1 - \frac{\lambda_i}{\lambda_j}\right)}, \tag{7.41}$$

which completes the proof.

7.5.2 Case 2: All Variances of ω are Equal

In the special, but in practice important case, when all variances of $\boldsymbol{\omega}$ are equal, that is when $\xi_i = \xi = \lambda^{-1/2}$, $\forall i \leq N$, I can state the following proposition.

Proposition 7.5.3 *If all variances $\xi_i^2 \neq 0$ are equal, the* MSE *and the spectral efficiency can be calculated as*

$$\text{MSE} = \frac{\lambda}{p} e^{\frac{\lambda}{p}} E_{in}\left(N, \frac{\lambda}{p}\right) \tag{7.42}$$

and

$$\eta = \frac{\mathcal{G}(\frac{\lambda}{p})}{a^N (N-1)!}, \tag{7.43}$$

respectively, where

$$\mathcal{G}(x) \triangleq \mathbf{MeijerG}_{1,0}^{3,1}\left(\begin{matrix} -N_r; -(N_r-1) \\ -N_r, -N_r, 0; . \end{matrix} \middle| x \right), \tag{7.44}$$

is the Meijer G function.

Proof. For the proof, notice that when all variances $\xi_i^2 \neq 0$ are equal, $f(x)$ becomes:

$$f(x, N, \lambda) = \frac{\lambda^N x^{N-1} e^{-\lambda x}}{(N-1)!}, \tag{7.45}$$

and for the MSE I get:

$$\text{MSE} = \frac{\lambda}{p} e^{\frac{\lambda}{p}} E_{\text{in}}\left(N, \frac{\lambda}{p}\right). \tag{7.46}$$

Similarly, using $f(x)$, for the SE, I get:

$$\eta = \frac{\mathcal{G}(\frac{\lambda}{p})}{a^N (N-1)!}, \tag{7.47}$$

which completes the proof.

In case of identical variances in ω, (7.42) gives the same expression as (7.13) in accordance with the fact that $\mathbf{S_\Omega}$ is proportional to the identity matrix.

7.6 Numerical Analysis of the Mean Squared Error

7.6.1 Channel Model and Covariance Matrix

Table 7.1 System Parameters

Parameter	**Value**
Number of antennas	$N_r = 4, 16, 20, 64, 100, 500$
Path Loss	$\alpha_\ell = 40, 45, 50$ dB
Power budget	$\tau_p P_{p,\ell} + \tau_d P_\ell = P_{tot} = 250$ mW, as in Eq. (7.1).
Total number of symbols (per time slot)	$F = 12$
Antenna spacing	$D/\lambda = 0.15, \ldots, 1.5$
Mean Angle of Arrival (AoA)	$\bar{\theta} = 70°$
Angular spread	$2 \cdot \theta_\Delta = 5, \ldots, 45°$

In this section I consider a single cell system, in which MSs use orthogonal pilots to facilitate the estimation of the uplink channel by the BS. Recall from Section 7.2 that the channel estimation process is independent for each MS and I can therefore focus on a single user. The covariance matrix $\mathbf{C}_\ell$ of the channel $\mathbf{h}_\ell$ as the function of the antenna spacing, mean angle of arrival and angular spread is modeled as by the well known spatial channel model, which is known to be accurate in non-line-of-sight environment with rich scattering and all antenna elements identically polarized, see [46]. For uniformly distributed angle of arrivals, the (m,n) $(m,n \in \{1,\ldots,N_r\})$ element of the covariance matrix of User-ℓ $\mathbf{C}_\ell$ is given by

$$\mathbf{C}_{m,n} = \frac{1}{2\theta_\Delta} \int_{-\theta_\Delta}^{\theta_\Delta} e^{j \cdot 2\pi \cdot \frac{D}{\lambda}(n-m)\cos(\bar{\theta}+x)} dx, \tag{7.48}$$

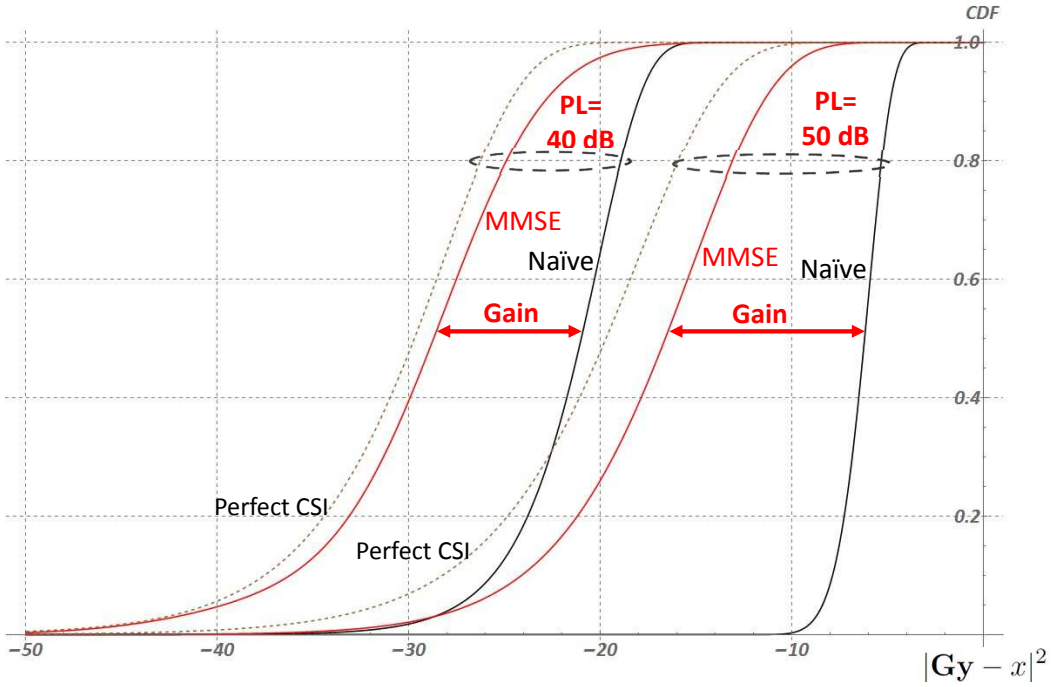

Fig. 7.1 Cumulative distribution function of the squared error in a single user MIMO scenario when the path loss between the UE and the BS is set to 40 and 50 dB when using the naïve receiver, the MMSE receiver and the receiver which has access to the perfect CSI with $N_r = 500$ antennas.

where the system parameters are given in Table 7.1. The covariance matrix $\mathbf{C}_\ell$ becomes practically diagonal as the antenna spacing and the angular spread grows beyond $D\lambda > 1$ and $\theta_\Delta > 30°$. In contrast, with critically spaced antennas $D\lambda = 0.5$ and $\theta_\Delta < 10°$, the antenna correlation in terms of the off-diagonal elements of $\mathbf{C}_\ell$ can be considered strong. Note that modeling the correlation matrices at the receiver side according to (7.48) corresponds to using the one-sided narrowband Kronecker model with receiver-side correlation, which is an appropriate model for the uplink of MU MIMO systems [27].

7.6.2 Numerical Results

In this section I consider a single cell single user MIMO system, in which the mobile terminal is equipped with a single transmit antenna, whereas the BS employs N_r receive antennas. Note that the performance characteristics of the proposed MMSE receiver as compared with the naïve receiver are similar in the multi-user MIMO case from the perspective of the tagged user, since the proposed receiver treats the multi-user interference as noise according to (7.10). The key input parameters to this system that are necessary to obtain numerical results using the MSE derivation in this chapter are listed in Table 7.1.

Figure 7.1 shows the cumulative distribution function (CDF) of the squared error of the estimated data symbols at the BS, i.e. the CDF of $\|\mathbf{Gy} - x\|^2$ using the naïve and the MMSE receiver when the number of antennas at the BS is $N_r = 500$. In all three cases in terms of path loss ($\alpha_\ell = 40$, $\alpha_\ell = 45$ and $\alpha_\ell = 50$ dB), the gain of the MMSE receiver is large in the entire region of the CDF. For example, at $\alpha_\ell = 40$, the median of the CDF is -21 dB with the naïve receiver and -29 dB with the MMSE receiver. This result indicates that using the MMSE receiver is advantageous not only in the average sense, but in virtually all channel states.

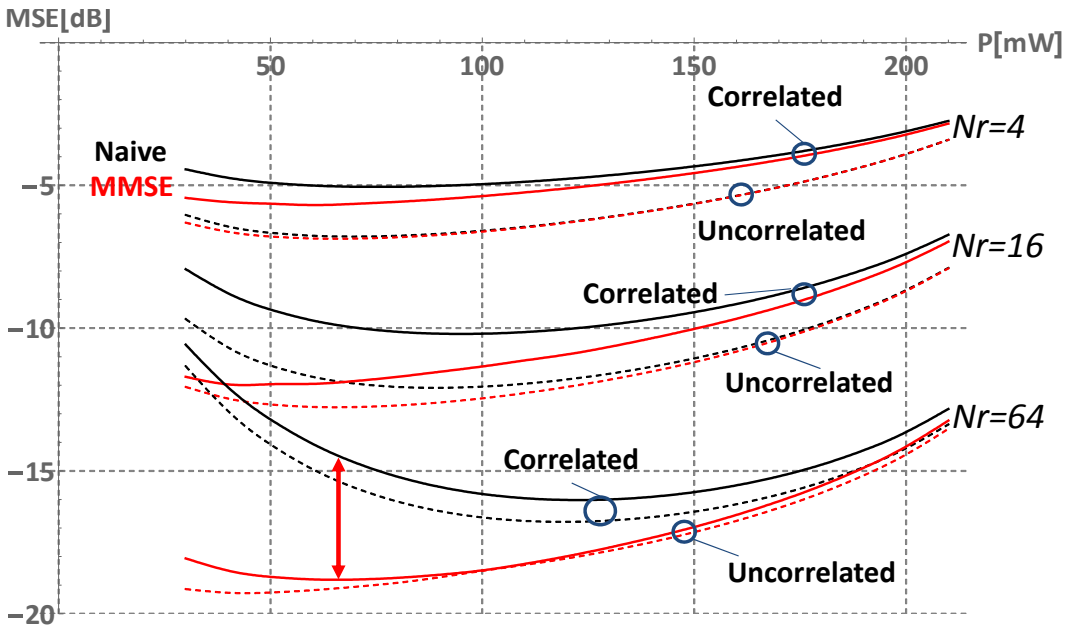

Fig. 7.2 Comparing the performance of the naïve and the MMSE receiver in the case of correlated (solid lines) and uncorrelated (dashed lines) antennas (with $N_r = 4, 16, 64$).

Figure 7.2 examines the impact of antenna correlation on the MSE with the naïve and the MMSE receivers. The impact of antenna correlation in terms of the achievable MSE decreases as the number of antennas increases from $N_r = 4$ to $N_r = 64$. An intuitive explanation of this insight is that the impact of correlation can be thought of as a factor that decreases the effective number of antennas, that is the number of antennas which contribute to the estimation of the transmitted data symbol. As the number of antennas grows large, antenna correlation decreases the effective number of antennas, but the loss due to this is not as significant as this loss when the number antennas is low. Instead, as the figure shows, at large number of antennas tuning the pilot power plays a more important role in minimizing the MSE than the effect of antenna correlation.

Figure 7.3 compares the performance of the naïve and the MMSE receivers with that of a receiver that has access to the perfect CSI, that is assuming that $\hat{\mathbf{h}}_\ell = \mathbf{h}_\ell$. This situation corresponds to $\mathbf{D}_\ell = \mathbf{I}$ and $\mathbf{Q}_\ell = \mathbf{0}$ and the structure of the naïve and the MMSE receivers coincide. Indeed, recall that the naïve receiver does minimize the MSE in the case of perfect CSI. The key aspect to observe in Figure 7.3 is that the gap between the MMSE receiver and the receiver operating with perfect CSI does not depend on N_r. This is in sharp contrast with the gap between the naïve receiver and the receiver with perfect CSI, which largely increases as the number of antennas gets large.

Figure 7.4 shows the SE of a MU MIMO system, in which the number of spatially multiplexed users is equal to the length of the employed pilot sequence τ_p. Figure 7.4 illustrates the trade-off between increasing the number of MU MIMO users and the necessary number of pilot symbols used to create orthogonal pilot sequences. A greater number of users increases the SE of the system at the expense of spending more symbols on the pilot signals. Therefore, I can see that around $\tau_p = 6$ the SE reaches its highest value. The gain in terms of SE of using the MMSE receiver is around 25% when the number of antennas is large.

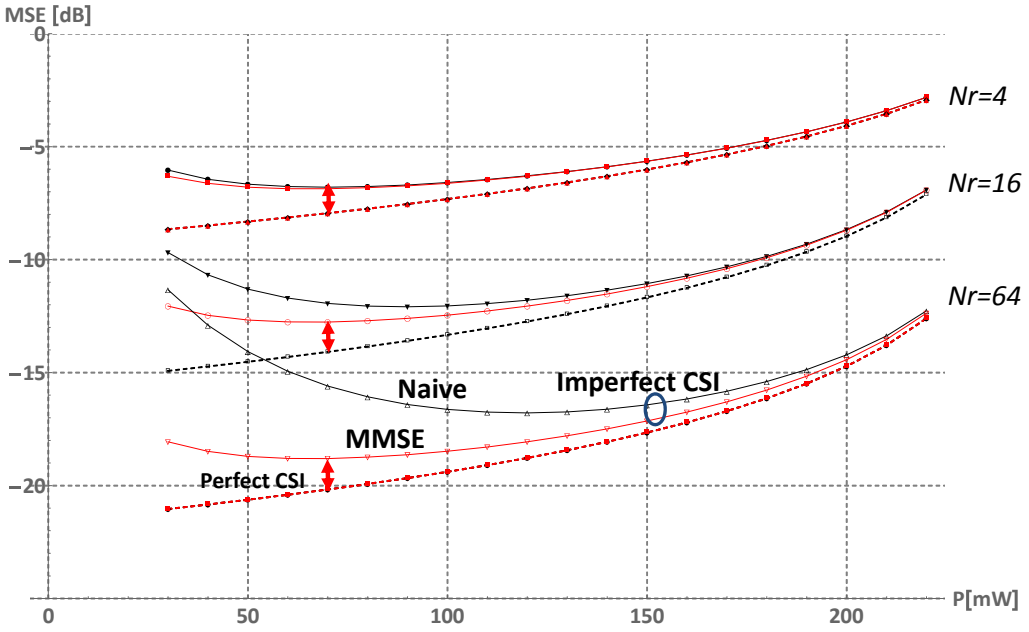

Fig. 7.3 Comparing the MSE performance of the naïve and MMSE receivers with that of a receiver that uses perfect CSI. As the pilot power increases, the MSE achieved by the receiver that uses perfect CSI increases, because due to the sum power constraint the transmit power available for the data symbols decreases.

7.7 Conclusions

In this chapter, I first derived an analytical expression ($\mathbf{G}_\ell^\star$) of a linear receiver structure that minimizes the MSE of the uplink estimated data symbols when the receive antennas possess a known correlation structure. I then derived closed form expressions for the MSE and the achievable SE when employing this MMSE receiver as a function of the pilot and data power, number of antennas, and path loss. I used Monte Carlo simulations to verify the analytical results and to gain insight into the system behavior when using $\mathbf{G}_\ell^\star$.

From the analysis I conclude that when employing the true MMSE receiver ($\mathbf{G}^\star$) at the BS in a MU MIMO system, the pilot power that minimizes the MSE is independent of the number of receive antennas. This implies that the optimal training does not need to be adjusted for sites with different numbers of antennas or when upgrading existing antenna sites to a larger number of antennas. In the special, (but in practice, typical) case when the thermal noise power levels on the data and pilot signals are equal, setting the pilot power by the terminal is easy, because the terminals continuously measure the path loss to the serving BS.

The simulation results provide the following insights:

- The performance difference between the naïve and the MMSE receiver increases with an increasing number of antennas. However, the performance gap between the MMSE receiver and the receiver that has access to a perfect CSI does not increase with the number of antennas.
- When the number of antennas is large, the impact of antenna correlation on the MSE is relatively small as compared with the impact of appropriately tuning the pilot power. When using the MMSE receiver ($\mathbf{G}_\ell^\star$), the optimal pilot power does not depend on the number of receive antennas, but is quite sensitive to large-scale fading.

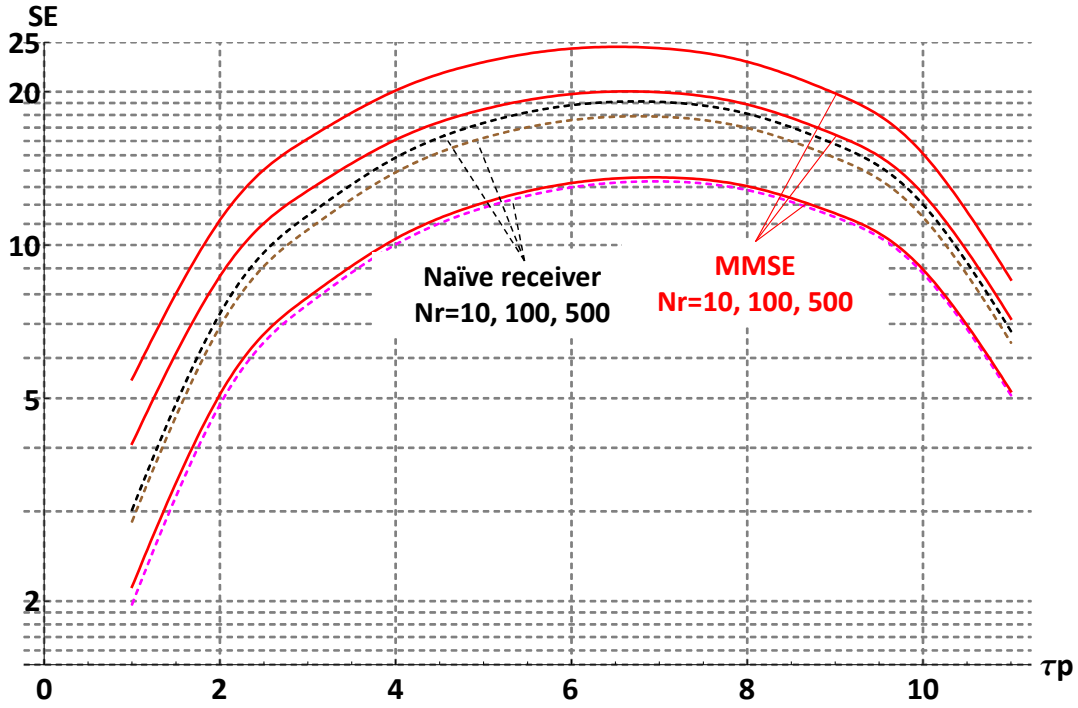

Fig. 7.4 Spectral efficiency as a function of the employed pilot symbols τ_p. In this example, the number of users in MU MIMO system is set equally to τ_p, that is I assume that the number of users that can be spatially multiplexed equals the pilot sequence length.

- When the number of antennas is large, the gain of using the MMSE receiver over using the naïve receiver is large, not only in terms of MSE, but also in the entire CDF of the squared error of the estimated data symbols.

I also showed that the well known relation between the MSE and the SE that holds for the case when perfect CSI at the MMSE receiver is available is valid also for the case of imperfect CSI at the regularized MMSE receiver ($\mathbf{G}^\star$). The deeper analysis of the impact of CSI errors in the case of non-separable channel models is an important topic for future research.

Appendix of Chapter 7

Proof of Proposition 7.2.1

Proof. If $\mathbf{C}_\ell = c_\ell \mathbf{I}$, implying $\mathbf{D}_\ell = d_\ell \mathbf{I}$, $\mathbf{Q}_\ell = q_\ell \mathbf{I}$ and the optimal $\mathbf{G}_\ell^\star$ of (7.10) can be written as:

$$\begin{aligned}\mathbf{G}_\ell^\star &= \frac{\alpha_\ell \sqrt{P_\ell} d_\ell}{\alpha_\ell^2 P_\ell \left(d_\ell^2 ||\hat{\mathbf{h}}_\ell||^2 + q_\ell\right) + \sum_{k\neq\ell}^K \alpha_k^2 P_k c_k + \sigma_d^2} \ \hat{\mathbf{h}}_\ell^H \\ &\triangleq g_\ell \cdot \hat{\mathbf{h}}_\ell^H . \end{aligned} \tag{7.49}$$

Substituting $\mathbf{G}_\ell^\star$ into (7.7) I get:

$$\begin{aligned}&\mathrm{MSE}\left(\hat{\mathbf{h}}_\ell\right) = -2\alpha_\ell \sqrt{P_\ell} g_\ell d_\ell ||\hat{\mathbf{h}}_\ell||^2 + 1 + \\ &g_\ell^2 \cdot \left(\alpha_\ell^2 P_\ell d_\ell^2 ||\hat{\mathbf{h}}_\ell||^4 + \left(\alpha_\ell^2 P_\ell q_\ell + \sum_{k\neq\ell}^K \alpha_k^2 P_k c_k + \sigma_d^2\right) ||\hat{\mathbf{h}}_\ell||^2\right).\end{aligned} \tag{7.50}$$

Recognizing that $\varphi_\ell \triangleq ||\hat{\mathbf{h}}_\ell||^2$ is Gamma distributed, the density function of φ_ℓ $\forall \ell$ is given by (dropping the index ℓ for convenience):

$$f_\varphi(x) = \frac{r^{-N_r} x^{N_r - 1} e^{-x/r}}{(N_r - 1)!} \quad x > 0. \tag{7.51}$$

Proposition 7.2.1 follows from Lemma (7.50) by taking the average of $\mathrm{MSE}\left(\hat{\mathbf{h}}_\ell\right)$ using the the following integrals:

$$\begin{aligned}&\int_{x=0}^\infty T_1 f_\varphi(x) dx = \\ &s_\ell \cdot \frac{N_r \left(-s_\ell r_\ell + e^{\frac{b_\ell}{s_\ell r_\ell}} \left(b_\ell + (1+N_r) s_\ell r_\ell\right) E_{in}\left(1+N_r, \frac{b_\ell}{s_\ell r_\ell}\right)\right)}{s_\ell^2 r_\ell};\end{aligned} \tag{7.52}$$

$$\begin{aligned}&\int_{x=0}^\infty T_2 f_\varphi(x) dx = \\ &b_\ell \cdot \frac{-s_\ell r_\ell + e^{\frac{b_\ell}{s_\ell r_\ell}} \left(b_\ell + N_r s_\ell r_\ell\right) E_{in}\left(N_r, \frac{b_\ell}{s_\ell r_\ell}\right)}{s_\ell^2 r_\ell^2};\end{aligned} \tag{7.53}$$

$$\int_{x=0}^\infty T_3 f_\varphi(x) dx = 2 \cdot e^{\frac{b_\ell}{s_\ell r_\ell}} N_r E_{in}\left(1+N_r, \frac{b_\ell}{s_\ell r_\ell}\right), \tag{7.54}$$

where $E_{in}(n,z) \triangleq \int_1^\infty e^{-zt}/t^n \, dt$ is a standard exponential integral function.

Proof of Lemma 7.3.1

Proof. I rewrite the MSE expression in (7.11), by making use of the following recursive relation, from [61] (also available at [62, 8.19.12]):

$$\mu_\ell E_{\mathrm{in}}(N_r, \mu_\ell) + N_r E_{\mathrm{in}}(N_r + 1, \mu_\ell) = e^{-\mu_\ell}. \tag{7.55}$$

Substituting $\mu_\ell = \frac{b}{rs}$ in this relation, using the terms of the MSE in (7.11) and rearranging, I obtain:

$$\text{MSE} = \frac{b_\ell}{r_\ell s_\ell} e^{\frac{b_\ell}{r_\ell s_\ell}} E_{\text{in}}\Big(N_r, \frac{b_\ell}{r_\ell s_\ell}\Big), \tag{7.56}$$

where, similarly to the notation used in Proposition 1, $b_\ell \triangleq q_\ell p_\ell + \sigma_d^2$ with $p_\ell \triangleq \alpha_\ell^2 P_\ell$ and $s_\ell \triangleq d_\ell^2 p_\ell$ and r_ℓ, d_ℓ and q_ℓ are defined in (7.12).

Finally, recognizing that:

$$\begin{aligned}\mu_\ell = \frac{b_\ell}{r_\ell s_\ell} = \frac{q_\ell \alpha_\ell^2 P_\ell + \sigma_d^2}{d_\ell^2 \alpha_\ell^2 P_\ell r_\ell} = \\ = \frac{\sigma_d^2 \sigma_p^2 \tau_d + c_\ell \alpha_\ell^2 (\sigma_p^2 P_{tot} + \tau_p P_{p,\ell}(\sigma_d^2 \tau_d - \sigma_p^2))}{c_\ell^2 \alpha_\ell^4 P_{p,\ell} \tau_p (P_{tot} - P_{p,\ell} \tau_p)}\end{aligned} \tag{7.57}$$

and substituting (7.1) into P_ℓ the lemma follows.

Proof of Proposition 7.3.2

I first prove the following lemma that will be useful for the proof of Proposition 7.3.2.

Lemma 7.7.1 *For $n > 0$ and $x \geq 0$, the following limit holds:*

$$\lim_{x \to \infty} x^2 (1 - e^x (n+x) E_{in}(n,x)) = -n. \tag{7.58}$$

Proof. Recalling the basic relationship between the incomplete Gamma function and the exponential integral function:

$$E_{\text{in}}(n,x) = x^{n-1} \Gamma(1-n,x), \tag{7.59}$$

and using the following expansion formula that is valid for large values of x (see [63]):

$$\begin{aligned}&\Gamma(1-n,x) \sim \\ &\sim x^{-n} e^{-x} \Big(1 + \frac{-n}{x} + \frac{-n(-n-1)}{x^2} + \\ &\quad + \frac{-n(-n-1)(-n-2)}{x^3} + \dots\Big),\end{aligned} \tag{7.60}$$

I have:

$$\begin{aligned}x^2 (1 - e^x (n+x) E_{\text{in}}(n,x)) \sim x^2 \Big(1 - e^x (n+x) x^{n-1} x^{-n} e^{-x} \cdot \\ \cdot (1 + \frac{-n}{x} + \frac{-n(-n-1)}{x^2} + \frac{-n(-n-1)(-n-2)}{x^3} + \dots)\Big).\end{aligned} \tag{7.61}$$

Rearranging terms, I finally get, for large x:

$$\begin{aligned}&x^2 (1 - e^x (n+x) E_{\text{in}}(n,x)) \sim \\ &-n + \frac{2n(1+n)}{x} - \frac{3n(1+n)(2+n)}{x^2} + \\ &+ \frac{4n(1+n)(2+n)(3+n)}{x^3} \mp \dots\end{aligned} \tag{7.62}$$

from which it follows:

$$\lim_{x\to\infty} x^2(1-e^x(n+x)E_{\text{in}}(n,x)) = -n. \tag{7.63}$$

I can now prove Proposition 7.3.2.

Proof (Proof of Proposition 7.3.2). I begin by taking the first derivative of MSE as a function of $P_{p,\ell}$. To this end, I use (7.11) and take the derivative of the MSE with respect to μ_ℓ:

$$\begin{aligned}\text{MSE}'(\mu_\ell) = &-\mu_\ell e^{\mu_\ell} E_{\text{in}}(N_r-1,\mu_\ell) + \\ &+ e^{\mu_\ell} E_{\text{in}}(N_r,\mu_\ell) + \mu_\ell e^{\mu_\ell} E_{\text{in}}(N_r).\end{aligned} \tag{7.64}$$

After some algebraic manipulation based on (7.55), I obtain:

$$\text{MSE}'(\mu_\ell) = e^{\mu_\ell}(N_r+\mu_\ell)E_{\text{in}}(N_r,\mu_\ell) - 1. \tag{7.65}$$

From [61] (also available at [62, 8.19.21]) I have

$$1 < (x+n)e^x E_{\text{in}}(n,x) < \frac{x+n}{x+n-1}.$$

Substituting $x = \mu_\ell$ and $n = N_r$ shows that $\text{MSE}'(\mu_\ell) \neq 0$ if $0 < \mu_\ell$.

Next, I consider the first derivative of $\mu(P_{p,\ell})$ as defined in (7.14) with respect to $P_{p,\ell}$:

$$\begin{aligned}\mu'(P_{p,\ell}) = &\frac{\sigma_d^2\sigma_p^2\tau_d(2P_{p,\ell}\tau_p - P_{tot})}{c_\ell^2 P_{p,\ell}^2\alpha_\ell^4\tau_p(P_{tot}-\tau_p P_{p,\ell})^2} + \\ &+ \frac{c_\ell\alpha_\ell^2\left(P_{p,\ell}^2\tau_p^2(\sigma_d^2\tau_d-\sigma_p^2) + 2P_{p,\ell}P_{tot}\sigma_p^2\tau_p - \sigma_p^2P_{tot}^2\right)}{c_\ell^2 P_{p,\ell}^2\alpha_\ell^4\tau_p(P_{tot}-\tau_p P_{p,\ell})^2}.\end{aligned} \tag{7.66}$$

The numerator of (7.66) is a second order polynomial of $P_{p,\ell}$ with the following coefficients $a_0 = -P_{tot}z$, $a_1 = 2\tau_p z$, $a_2 = c_\ell\alpha_\ell^2\tau_p^2(\sigma_d^2\tau_d - \sigma_p^2)$, where $z = (c_\ell P_{tot}\alpha_\ell^2 + \sigma_d^2\tau_d)\sigma_p^2$. a_0 is negative, a_1 is positive and the sign of a_2 depends on the sign of $\sigma_d^2\tau_d - \sigma_p^2$. In reasonable cases a_2 is positive as well, because $\tau_d > 1$ and $\sigma_d \approx \sigma_p$. When a_2 is positive the numerator of (7.66) has one positive and one negative root, because $a_1 < \sqrt{a_1^2 - a_0a_2}$ and the positive root is

$$P_{p,\ell}^* = \frac{-a_1 + \sqrt{a_1^2 - 4a_0a_2}}{2a_2}. \tag{7.67}$$

Finally, the first derivative of the MSE with respect to $P_{p,\ell}$ is:

$$\frac{d}{dP_{p,\ell}}\text{MSE} = \text{MSE}'(\mu_\ell)\cdot\mu'(P_{p,\ell}). \tag{7.68}$$

Recall that $\text{MSE}'(\mu_\ell) \neq 0$, the roots of $\frac{d}{dP_{p,\ell}}\text{MSE}$ are identical with the roots of the numerator of (7.66) and the positive root of $\frac{d}{dP_{p,\ell}}\text{MSE}$ is $P_{p,\ell}^*$.

I still need to show that $P_{p,\ell}^\star$ corresponds to a local minimum. To this end, I study the sign of $\lim_{P_{p,\ell}\to 0^+}\frac{d}{dP_{p,\ell}}\text{MSE}$. If the limit is negative then $P_{p,\ell}^\star$ corresponds to a local minimum. Unfortunately, (7.68) is not directly applicable because $\lim_{P_{p,\ell}\to 0^+}\mu'(P_{p,\ell}) = 0$ and $\lim_{\mu_\ell\to\infty}\text{MSE}'(\mu_\ell) = \infty$. Instead, according to (7.56) I introduce $F(n,x) = xe^x E_{in}(n,x)$ and rewrite (7.57) as

$$\mu_\ell = \frac{b_1 + b_2 P_{p,\ell}}{P_{p,\ell}(b_3 - P_{p,\ell})},$$

where $b_1 = \frac{\sigma_d^2\sigma_p^2\tau_d + c_\ell\alpha_\ell^2 P_{tot}\sigma_p^2}{c_\ell^2\alpha_\ell^4\tau_p^2}$, $b_2 = \frac{c_\ell\alpha_\ell^2\tau_p(\sigma_d^2\tau_d - \sigma_p^2)}{c_\ell^2\alpha_\ell^4\tau_p^2}$, $b_3 = \frac{P_{tot}}{\tau_p}$ and note that b_1 and b_3 are positive. This way $\text{MSE} = F\left(N_r, \frac{b_1 + b_2 P_{p,\ell}}{P_{p,\ell}(b_3 - P_{p,\ell})}\right)$. Introducing $\tilde{P}_\ell = \frac{b_1 P_{p,\ell}(b_3 - P_{p,\ell})}{b_3(b_1 + b_2 P_{p,\ell})}$ I also have $\text{MSE} = F\left(N_r, \frac{b_1}{\tilde{P}_\ell b_3}\right)$ and can rewrite the limit as

$$\lim_{P_{p,\ell}\to 0^+} \frac{d}{dP_{p,\ell}}\text{MSE} = \lim_{P_{p,\ell}\to 0^+} \frac{d}{dP_{p,\ell}} F\left(N_r, \frac{b_1 + b_2 P_{p,\ell}}{P_{p,\ell}(b_3 - P_{p,\ell})}\right) =$$
$$= \lim_{P_{p,\ell}\to 0^+} \frac{d}{dP_{p,\ell}}\tilde{P}_\ell \quad \lim_{\tilde{P}_\ell\to 0^+} \frac{d}{d\tilde{P}_\ell} F\left(N_r, \frac{b_1}{\tilde{P}_\ell b_3}\right).$$

The first term converges to 1, while the second term converges to $-\frac{N_r b_3}{b_1}$ based on (7.58).

Proof of Lemma 7.4.1

Proof. According to the matrix inversion lemma for matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ of size $n\times n$, $n\times m$, $m\times m$, $m\times n$, respectively, I have

$$(\mathbf{A} + \mathbf{BCD})^{-1} =$$
$$\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1}\right)^{-1}\mathbf{DA}^{-1}.$$

Substituting $\mathbf{A} = \mathbf{\Psi}_\ell$, $\mathbf{B} = \alpha_\ell\sqrt{P_\ell}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell$, $\mathbf{C} = 1$, $\mathbf{D} = \alpha_\ell\sqrt{P_\ell}\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H$ I have:

$$\begin{aligned}\left(\mathbf{\Psi}_\ell + \alpha_\ell^2 P_\ell\mathbf{D}_\ell\hat{\mathbf{h}}_\ell\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\right)^{-1} &= \mathbf{\Psi}_\ell^{-1} - \mathbf{\Psi}_\ell^{-1}\alpha_\ell\sqrt{P_\ell}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell\cdot\left(\alpha_\ell\sqrt{P_\ell}\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\mathbf{\Psi}_\ell^{-1}\alpha_\ell\sqrt{P_\ell}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell + 1\right)^{-1}\cdot \\ &\cdot\alpha_\ell\sqrt{P_\ell}\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\mathbf{\Psi}_\ell^{-1} = \\ &= \mathbf{\Psi}_\ell^{-1} - \frac{\alpha_\ell^2 P_\ell}{\alpha_\ell^2 P_\ell\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\mathbf{\Psi}_\ell^{-1}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell + 1}\cdot\mathbf{\Psi}_\ell^{-1}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\mathbf{\Psi}_\ell^{-1}. \end{aligned} \tag{7.69}$$

Substituting (7.69) into (7.16) gives:

$$\mathbf{G}_\ell^\star = \alpha_\ell\sqrt{P_\ell}\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\cdot\left(\mathbf{\Psi}_\ell^{-1} - \frac{\alpha_\ell^2 P_\ell}{\alpha_\ell^2 P_\ell\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell\mathbf{\Psi}_\ell^{-1}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell + 1}\mathbf{\Psi}_\ell^{-1}\mathbf{D}_\ell\hat{\mathbf{h}}_\ell\hat{\mathbf{h}}_\ell^H\mathbf{D}_\ell^H\mathbf{\Psi}_\ell^{-1}\right). \tag{7.70}$$

Recall that $\mathbf{\Psi}_\ell = \mathbf{\Theta}_\ell^H\mathbf{S}_\ell\mathbf{\Theta}_\ell$ is the SVD of $\mathbf{\Psi}_\ell$. Substituting (7.19) - (7.20), and this SVD into (7.70) I get:

$$\begin{aligned}\mathbf{G}_\ell^\star &= \alpha_\ell\sqrt{P_\ell}\left(\boldsymbol{\nu}_\ell^H\mathbf{S}_\ell^{-1/2}\mathbf{\Theta}_\ell - \frac{\alpha_\ell^2 P_\ell}{\alpha_\ell^2 P_\ell||\boldsymbol{\nu}_\ell||^2 + 1}||\boldsymbol{\nu}_\ell||^2\boldsymbol{\nu}_\ell^H\mathbf{S}_\ell^{-1/2}\mathbf{\Theta}_\ell\right) = \\ &= \alpha_\ell\sqrt{P_\ell}\left(1 - \frac{\alpha_\ell^2 P_\ell||\boldsymbol{\nu}_\ell||^2}{\alpha_\ell^2 P_\ell||\boldsymbol{\nu}_\ell||^2 + 1}\right)\boldsymbol{\nu}_\ell^H\mathbf{S}_\ell^{-1/2}\mathbf{\Theta}_\ell = \frac{\alpha_\ell\sqrt{P_\ell}}{\alpha_\ell^2 P_\ell||\boldsymbol{\nu}_\ell||^2 + 1}\boldsymbol{\nu}_\ell^H\mathbf{S}_\ell^{-1/2}\mathbf{\Theta}_\ell. \end{aligned} \tag{7.71}$$

Proof of Lemma 7.4.2

Proof. Substituting (7.17) into (7.7) with optimal MMSE receiver, I get:

$$\mathrm{MSE}\left(\hat{\mathbf{h}}_\ell\right) = \mathbf{G^\star}_\ell \left(\alpha_\ell^2 P_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell{}^H + \mathbf{\Psi}_\ell\right) \mathbf{G^\star}_\ell^H - \alpha_\ell \sqrt{P_\ell} (\mathbf{G^\star}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell + \hat{\mathbf{h}}_\ell^H \mathbf{D}_\ell{}^H \mathbf{G^\star}_\ell^H) + 1. \quad (7.72)$$

From (7.22), (7.21) and the SVD of $\mathbf{\Psi}_\ell$ I have:

$$\mathbf{G^\star}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell = g_\ell \boldsymbol{\nu}_\ell^H \boldsymbol{\nu}_\ell = g_\ell ||\boldsymbol{\nu}_\ell||^2, \quad (7.73)$$

$$\mathbf{G^\star}_\ell \mathbf{\Psi}_\ell \mathbf{G^\star}_\ell^H = g_\ell^2 \boldsymbol{\nu}_\ell^H \boldsymbol{\nu}_\ell = g_\ell^2 ||\boldsymbol{\nu}_\ell||^2, \quad (7.74)$$

substituting this into (7.72) I obtain

$$\mathrm{MSE}\left(\boldsymbol{\nu}_\ell\right) = \alpha_\ell^2 P_\ell g_\ell^2 ||\boldsymbol{\nu}_\ell||^4 + g_\ell^2 ||\boldsymbol{\nu}_\ell||^2 - 2\alpha_\ell \sqrt{P_\ell} g_\ell ||\boldsymbol{\nu}_\ell||^2 + 1, \quad (7.75)$$

where g_ℓ is also a function of $||\boldsymbol{\nu}_\ell||^2$ according to (7.23). Substituting (7.23) into (7.75) gives the lemma after some algebraic manipulations.

Chapter 8
Applications of the Results: Pilot-to-Data Power Ratio Balancing in the Massive MIMO Concept by the METIS Project

8.1 Background 1: Long Term Evolution and 5G Networks by the 3rd Generation Partnership Project

Currently, the Long Term Evolution (LTE) networks defined and developed by the 3rd Generation Partnership Project (3GPP) are the fastest growing cellular technology in history. Indeed, almost a quarter of all global mobile subscribers are using LTE, and it is expected that by 2021 this will increase to more than half of all subscribers, accounting for approximately 4.3 billion subscriptions. In certain regions, such as Korea, Japan, China and the U.S.A, LTE has already reached or exceeded 90 percent penetration [64]. The technology footprint of 3GPP technologies in general and LTE in particular, has led to unprecedented economies of scale, which, in turn resulted in a rapid growth of technology solutions that enable LTE to operate with high spectral and energy efficiency in a great number of spectrum bands.

As the demands for higher throughput and more data capacity, particularly for video, and better broadband services continue to be growing, the mobile industry is now focusing on developing a new set of technology enablers and features, collectively recognized as the 5^{th} generation (5G) mobile technology. 5G, often referred to as New Radio (NR) in 3GPP, is also targeted to meet the diverse requirements imposed by vertical markets, including massive machine type communication between sensors and actuators, ultra-reliable and/or low latency communication and a host of various Internet of Things (IoT) applications.

MIMO systems involving a number of antenna elements with an order of magnitude larger than in the early releases of wireless standards is a quickly maturing technology developed by the 3GPP. Indeed, an ongoing work item of 3GPP for the Release-14 of the Long Term Evolution (LTE) is identifying the technology enablers and performance benefits of deploying up to 64 antenna ports at wireless access points and BSs. While this is a significant increase of the number of antenna ports as compared with today's typical deployments, to fully realize the promises of scaling up MIMO to very large (referred to as full dimension or massive MIMO) arrays in practice requires further research and system development work.

8.2 Background 2: MIMO Systems for 5G Developed in the METIS Project

As the standardization of massive MIMO systems in the 3GPP progresses, the research community has started to explore the potential of very large arrays as an enabler technology for meeting the requirements of 5th generation systems. Indeed, in its final deliverable, the European 5G project Mobile and Wireless Communication Enablers for the 2020 Information Society (METIS) identifies massive MIMO as a key 5G enabler and proposes specific technology components that will allow the cost efficient deployment of cellular (and specifically 3GPP) systems taking advantage of hundreds of antennas at cellular base stations. The main findings of the METIS project are summarized in [65].

One of the technology components is CSI acquisition and handling the inherent pilot-data resource allocation trade-off in a near optimal fashion.

8.3 Application: Channel State Information Acquisition in the METIS 5G Concept

Understanding and managing the inherent trade-offs related to CSI acquisition in massive MIMO systems is fundamental to the design of such systems [4], [5], [7]. As it has been discussed in details in [4], the level of pilot contamination in terms of the number of users using non-orthogonal pilot sequences can be controlled by, among other techniques, setting the pilot sequence reuse factor across the cells of a multi-cell system.

Also, the effect of pilot contamination can be mitigated by downlink pilot contamination elimination precoding (PCEP) proposed by [5]. In my approach, the level of pilot contamination is controlled by adaptively setting the pilot reuse factor and balancing the pilot-data power ratio, which is beneficial for both the uplink and downlink performance.

When operating in limited coherence time and frequency channel conditions, the number of symbols that is available within the coherence time of the channel is limited and the inherent trade-offs between allocating resources to pilot and data symbols include the following:

- Allocating more power, time, or frequency resources improves the quality of the channel estimate, but leaves fewer resources for uplink and downlink data transmission.
- Constructing longer pilot sequences helps to avoid tight pilot reuse in multi-cell systems, which in turn helps to reduce or avoid pilot contamination. On the other hand, spending a greater number of symbols on pilots increases the pilot overhead.
- In multiuser MIMO systems, increasing the number of orthogonal pilot sequences may increase the number of spatially multiplexed users, since a larger number of orthogonal pilot sequences enable the system to distinguish a larger number of users in the code domain. However, this comes at the expense of having longer pilot sequences.

The METIS massive MIMO concept combines the coordinated allocation of resources available for pilot (reference) signals across multiple cells and the balancing of the pilot-data resources within each cell [8]. For pilot contamination mitigation the Operation and Maintenance System employs low-rate multi-cell coordination to set the pilot reuse factor (e.g. pilot reuse-1 or pilot reuse-3) depending on the cell load and the coherence time budget that can be used for creating pilot sequences.

For example, at pedestrian speed of 1.5 m/s and outdoor cell radius of around 1000 m at 2 GHz carrier frequency, the number of symbols within the coherent bandwidth and time are around $B \cdot T = 300$ KHz x 25 ms = 7500, whereas this coherent budget is only a few hundred symbols at vehicular speeds (with coherence time of $T \approx 1.25$ ms). Notice that even a greater-than-one pilot reuse scheme does not eliminate pilot contamination, since pilot interference can be caused by all surrounding cells. Therefore, within each cell, each mobile station employs pilot-data power ratio balancing at a fine time scale to maximize spectral efficiency [8].

Chapter 9
Summary

9.1 Thesis I: Pilot-to-Data Power Ratio in Single User Systems

Thesis I considers a single input multiple output SIMO system in which the MS balances its PDPR, while the BS uses LS channel estimation to initialize its linear MMSE equalizer. Thesis I is concerned with calculating the MSE of the uplink estimated data symbols based on the uplink received data signal and the available channel estimate by the multiple antenna BS.

Theorem 9.1. *Let the number of receive antennas at a multiple antenna BS be denoted by N_r. The expected value of the mean square error of the equalized symbols is*

$$\mathbb{E}\{\text{MSE}\} = d^2 N_r \Big(\mathcal{G}(a, 1+N_r) + pr\mathcal{G}(1+N_r, 1+N_r) - 1 \Big) + \\ + \frac{b}{pr} \Big(\mathcal{G}(a, N_r) + pr\mathcal{G}(N_r, N_r) - 1 \Big) - 2d \cdot \Big(pr\mathcal{G}(N_r, 1+N_r) \Big) + 1;$$

where P is the transmit power employed by the MS for transmitting the data symbols, $p = \alpha^2 P$, $a = \sigma^2$, α is the large scale fading, σ^2 is the noise power at the receive antenna, $b = qp + \sigma^2$, $\mathbf{R} \triangleq \mathbb{E}\left\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\right\} = \mathbf{C} + \frac{\sigma^2}{P^p \alpha^2}\,\mathbf{I} = r\,\mathbf{I}$, and

$$\mathcal{G}(x, y) \triangleq \frac{1}{pr} e^{\frac{a}{pr}} x E_{in}\Big(y, \frac{a}{pr}\Big),$$

and $E_{in}(n, z) \triangleq \int_1^\infty e^{-zt}/t^n \, dt$ is a standard exponential integral function.

9.2 Thesis II: Minimum Mean Squared Error Receiver in the Presence of Channel State Information Errors

Thesis II is concerned with deriving an explicit formula for a receiver that minimizes the mean squared error of the estimated data symbols when the receiver has a non-perfect channel estimate. Thesis II also derives a closed form expression for the MSE when this optimum receiver is used, as a function of the number of receive antennas and the applied data and pilot power. Specifically, using the notation and terminology of Thesis I, I have the following results.

Let κ be the index of a tagged user in a MU-MIMO system of K users, $\kappa = 1 \ldots K$, and let $\mathbf{G}_\kappa^\star$ denote the receiver vector that minimizes the MSE of the estimated data symbols of the tagged user.
The optimal MU-MIMO receiver vector for User-κ is as follows:

Theorem 9.2.1 *The optimal $\mathbf{G}_\kappa^\star$ can be derived as:*

$$\mathbf{G}_\kappa^\star = \alpha_\kappa \sqrt{P_\kappa} \hat{\mathbf{h}}_\kappa^H \mathbf{D}_\kappa^H \cdot \\ \cdot \left(\alpha_\kappa^2 P_\kappa \left(\mathbf{D}_\kappa \hat{\mathbf{h}}_\kappa \hat{\mathbf{h}}_\kappa^H \mathbf{D}_\kappa{}^H + \mathbf{Q}_\kappa \right) + \sum_{k \neq \kappa}^{K} \alpha_k^2 P_k \mathbf{C}_k + \sigma_d^2 \mathbf{I} \right)^{-1}. \tag{9.1}$$

Next, it is of great interest to determine the MSE of the estimated data symbols when employing $\mathbf{G}_\kappa^\star$ as the receiver of the tagged user. To this end, the following result holds.

Theorem 9.2.2 *The unconditional* MSE *of the received data symbols of User-κ when the BS uses the optimal $\mathbf{G}_\kappa^\star$ receiver is as follows.*

$$\begin{aligned}\text{MSE} = \\ s_\kappa \cdot \frac{N_r\Big(-s_\kappa r + e^{\frac{b_\kappa}{s_\kappa r}}\big(b_\kappa + (1+N_r)s_\kappa r\big)E_{in}\big(1+N_r, \frac{b_\kappa}{s_\kappa r}\big)\Big)}{s_\kappa^2 r} + \\ + b_\kappa \cdot \frac{-s_\kappa r + e^{\frac{b_\kappa}{s_\kappa r}}\big(b_\kappa + N_r s_\kappa r\big)E_{in}\big(N_r, \frac{b_\kappa}{s_\kappa r}\big)}{s_\kappa^2 r^2} - \\ -2\cdot e^{\frac{b_\kappa}{s_\kappa r}} N_r E_{in}\Big(1+N_r, \frac{b_\kappa}{s_\kappa r}\Big) + 1,\end{aligned} \tag{9.2}$$

where $E_{in}(n,z) \triangleq \int_1^\infty e^{-zt}/t^n\, dt$ is a standard exponential integral function.

9.3 Thesis III: The Impact of Antenna Correlation on the Pilot-to-Data Power Ratio

Thesis III is concerned with determining the MSE of the estimated data symbols at multiple antenna receivers with correlated receive antenna structures. Specifically, Thesis III allows for an arbitrary correlation matrix structure at the multiple antenna receiver and derives a closed form expression for the MSE of the received data symbols as follows.

Theorem 9.3.1 *The mean square error (MSE) of the uplink received data with arbitrary covariance matrix $\mathbf{C}$ of the uplink channel can be calculated as*

$$\text{MSE} = T_1 + T_2 + T_3 + 1, \tag{9.3}$$

where

$$T_1 = \sum_k \sum_{\ell,\ell\neq k} d_k d_\ell \cdot$$

$$\cdot \int_{x=0}^{\infty} x e^{-x\sigma^2/(\alpha^2 P)} \frac{1}{x+r_k}\,\frac{1}{x+r_\ell}\prod_i \frac{r_i}{x+r_i}\, dx +$$

$$+ \sum_k d_k^2 \int_{x=0}^{\infty} x e^{-x\sigma^2/(\alpha^2 P)}\,\frac{2}{(x+r_k)^2}\prod_i \frac{r_i}{x+r_i}\, dx;$$

$$T_2 = \frac{1}{\alpha^2 P}\sum_k m_k \int_{x=0}^{\infty} x e^{-x\frac{\sigma^2}{\alpha^2 P}} \frac{1}{x+r_k}\prod_i \frac{r_i}{x+r_i} dx;$$

and

$$T_3 = 2\sum_k d_k \int_{x=0}^{\infty} e^{-x\frac{\sigma^2}{\alpha^2 P}} \frac{1}{x+r_k} \prod_i \frac{r_i}{x+r_i} dx,$$

where $\mathbf{S_M} \triangleq \alpha^2 P \mathbf{S_Q} + \sigma^2 \mathbf{I}$ *is a diagonal matrix with diagonal elements* $m_k = \mathbf{S_M}_{kk} = \alpha^2 P q_k + \sigma^2$*, where* $q_k = \mathbf{S_Q}_{kk}$.

9.4 Thesis IV: Block and Comb Type Channel Estimation

Thesis IV derives analytical results for the spectral efficiency of MU-MIMO systems, in which the overall resources must be shared between channel estimation and data transmission. Specifically, Thesis IV shows that the spectral efficiency in MU-MIMO systems is not only a function of the PDPR, but it also depends on the specific channel estimation scheme, as given by the following results.

Theorem 9.4.1 (Spectral efficiency with LS estimation) *Assume* $\mathbf{C} = c\mathbf{I}_{N_r}$*, where* $c \in \mathbb{R}^+$*, then the average spectral efficiency with LS channel estimation and MMSE receiver is*

$$\bar{S}_{LS} = \frac{(\tau - \tau_p)}{\tau} \left(\frac{2\mathcal{G}(x_0) - \mathcal{G}(x_1) - \mathcal{G}(x_2)}{(N_r - 1)!} - \log(d-1)^2 \right) \tag{9.4}$$

with $x_{1,2} = \frac{1}{2}\left(-\frac{2\sigma^2 - 2d\sigma^2 + b}{p(d-1)^2} \pm \sqrt{\left(\frac{2\sigma^2 - 2d\sigma^2 + b}{p(d-1)^2}\right)^2 - \frac{4\sigma^4}{p^2(d-1)^2}} \right)$, $x_0 = \frac{\sigma^2}{p}$, $p = \alpha^2 P$, $b = qp + \sigma^2$, $q = c(1 - c/r$, $r = c + \frac{\sigma^2}{\alpha^2 P_p \tau_p}$,
and where

$$\mathcal{G}(x) = \mathbf{MeijerG}_{2,3}^{1,3}\left(\begin{matrix} 0, 1 \\ 0, 0, N_r \end{matrix} \middle| \frac{x}{r} \right), \tag{9.5}$$

is the Meijer G-function.

Theorem 9.4.2 (Spectral efficiency with MMSE estimation) *Assume* $\mathbf{C} = c\mathbf{I}_{N_r}$*, where* $c \in \mathbb{R}^+$*, then the average spectral efficiency with MMSE channel estimation and MMSE receiver is*

$$\bar{S}_{MMSE} = \frac{(\tau - \tau_p)}{\tau} \left(\log(pb) + \frac{2\mathcal{G}(x_3) - \mathcal{G}(x_4)}{(N_r - 1)!} \right) \tag{9.6}$$

with $x_3 = \frac{\sigma^2}{p}$, $x_4 = \frac{\sigma^2}{pb}$, $b = qp + \sigma^2$, $q = \frac{\sigma^2 c}{\sigma^2 + \alpha^2 c P_p \tau_p}$*, and* $\mathcal{G}(x)$ *defined in 9.4.1.*

9.5 Thesis V: The Pilot-to-Data Power Ratio in Multiuser Systems

Thesis V derives analytical results for the MSE of the received data symbols and the overall spectral efficiency of MU-MIMO systems, in which the overall resources must be shared between channel estimation and data transmission, and the receive antennas are correlated according to an arbitrary correlation structure.

Let

$$\mathbf{\Psi}_\ell \triangleq \alpha_\ell^2 P_\ell \mathbf{Q}_\ell + \sum_{k \neq \ell}^{K} \alpha_k^2 P_k \mathbf{C}_k + \sigma_d^2 \mathbf{I}_{N_r}, \tag{9.7}$$

and

$$\mathbf{\Psi}_\ell = \mathbf{\Theta}_\ell^H \mathbf{S}_\ell \mathbf{\Theta}_\ell \tag{9.8}$$

denote the singular value decomposition (SVD) of $\mathbf{\Psi}_\ell$. Furthermore, define the linear transformed version of the estimated channel $\hat{\mathbf{h}}_\ell$ as:

$$\boldsymbol{\nu}_\ell \triangleq \mathbf{S}_\ell^{-1/2} \mathbf{\Theta}_\ell \mathbf{D}_\ell \hat{\mathbf{h}}_\ell, \tag{9.9}$$

and denote the distribution of $\boldsymbol{\nu}_\ell$ as:

$$\boldsymbol{\nu}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_\ell), \tag{9.10}$$

where

$$\mathbf{\Omega}_\ell \triangleq \mathbb{E}(\boldsymbol{\nu}_\ell \boldsymbol{\nu}_\ell^H) = \mathbf{S}_\ell^{-1/2} \mathbf{\Theta}_\ell \mathbf{D}_\ell \mathbf{R}_\ell \mathbf{D}_\ell^H \mathbf{\Theta}_\ell^H \mathbf{S}_\ell^{-1/2},$$

and denote the SVD of $\mathbf{\Omega}_\ell$:

$$\mathbf{\Omega}_\ell = \mathbf{\Theta}_{\mathbf{\Omega}_\ell}^H \mathbf{S}_{\mathbf{\Omega}\ell} \mathbf{\Theta}_{\mathbf{\Omega}\ell}, \tag{9.11}$$

where $\mathbf{\Theta}_{\mathbf{\Omega}\ell}$ is an orthogonal matrix.

Also, denote the linear transform of $\boldsymbol{\nu}_\ell$, with $\boldsymbol{\omega}_\ell$

$$\boldsymbol{\omega}_\ell \triangleq \mathbf{\Theta}_{\mathbf{\Omega}\ell} \boldsymbol{\nu}_\ell \,, \tag{9.12}$$

and its diagonal covariance matrix with $\mathbf{S}_{\mathbf{\Omega}}$.

With this notation, the MSE and the SE can be calculated as follows:

Theorem 9.5.1 *Denote the variance of ω_i with ξ_i^2. Then, $|\omega_i|^2$ is exponentially distributed with parameter $\lambda_i = 1/\xi_i^2$, and the mean squared error of the received data symbols can be calculated as:*

$$\text{MSE} = \int_x \frac{1}{\alpha_\ell^2 P_\ell x + 1} f(x) dx, \tag{9.13}$$

while the spectral efficiency can be calculated as:

$$\eta = -\int_x \log \left[\frac{1}{\alpha_\ell^2 P_\ell x + 1} \right] f(x) dx, \tag{9.14}$$

where α_ℓ represents the large scale fading (path loss) of User-ℓ, and $f(x)$ is the the density function of $\sum_{i=1}^{N_r} |\omega_i|^2$:

$$f(x) = e_1^T e^{\mathbf{A}x} e_{N_r} \lambda_{N_r}, \tag{9.15}$$

where e_i is the i-th unit vector (whose only nonzero element is 1 at position i), and the matrix $\mathbf{A}$ is:

$$\mathbf{A} = \begin{pmatrix} -\lambda_1 & \lambda_1 & & \\ & -\lambda_2 & \lambda_2 & \\ & & \ddots & \ddots \\ & & & -\lambda_{N_r} \end{pmatrix}. \tag{9.16}$$

For the special but important case, when all non-zero ξ_i (and λ_i) are distinct (different), the following result holds:

Proposition 9.5.2 *When all non-zero ξ_i (and λ_i) are distinct (different), then*

$$f(x) = \sum_{i=1}^{N} \frac{\lambda_i e^{-\lambda_i x}}{\prod_{j=1, j\neq i}^{N} \left(1 - \frac{\lambda_i}{\lambda_j}\right)}, \tag{9.17}$$

and the mean squared error can be calculated as:

$$\text{MSE} = \sum_{i=1}^{N} \frac{-\lambda_i^{\frac{-N}{2}} e^{\frac{\lambda_i}{p}} E_{in}\left(1, \frac{-\lambda_i}{p}\right)}{p \prod_{j=1, j\neq i}^{N} \left(1 - \frac{\lambda_i}{\lambda_j}\right)}, \tag{9.18}$$

where $p = \alpha^2 P_\ell$.

The SE can be calculated as follows:

$$\eta = \sum_{i=1}^{N} \frac{-\lambda_i^{\frac{2-N}{2}} e^{\frac{\lambda_i}{p}} E_{in}\left(1, \frac{-\lambda_i}{p}\right)}{\prod_{j=1, j\neq i}^{N} \left(1 - \frac{\lambda_i}{\lambda_j}\right)}. \tag{9.19}$$

In the special but important when all variances of $\boldsymbol{\omega}$ are equal, the following proposition holds:

Proposition 9.5.3 *Suppose $\xi_i = \xi = \lambda^{-1/2}, \quad \forall i \leq N$. Then, $f(x)$ follows the Erlang distribution as follows:*

$$f(x, N, \lambda) = \frac{\lambda^N x^{N-1} e^{-\lambda x}}{(N-1)!}, \tag{9.20}$$

and the MSE is given by:

$$\text{MSE} = \frac{\lambda}{p} e^{\frac{\lambda}{p}} E_{in}\left(N, \frac{\lambda}{p}\right), \tag{9.21}$$

and the spectral efficiency can be calculated as:

$$\eta = \frac{\mathcal{G}(\frac{\lambda}{p})}{a^N (N-1)!}, \tag{9.22}$$

where

$$\mathcal{G}(x) \triangleq \textbf{MeijerG}_{1,0}^{3,1}\left(\begin{matrix} -N_r; -(N_r - 1) \\ -N_r, -N_r, 0; . \end{matrix} \middle| x \right), \tag{9.23}$$

is the Meijer G function.

Bibliography

[1] L. Zheng, D. Tse, and M. Médard, "Channel Coherence in the Low-SNR Regime," *IEEE Trans. Information Theory*, vol. 53, no. 3, pp. 976–997, March 2007.

[2] S. Ray, M. Médard, and L. Zheng, "On Noncoherent MIMO Channels in the Wideband Regime: Capacity and Reliability," *IEEE Trans. Information Theory*, vol. 53, no. 6, pp. 1983–2009, June 2007.

[3] D. Porrat, "Information Theory of Wideband Communications," *IEEE Communications Surveys and Tutorials*, vol. 9, no. 2, pp. 2–16, July 2007.

[4] V. Raghavan, G. Hariharan, and A. M. Sayeed, "Capacity of Sparse Multipath Channels in the Ultra-Wideband Regime," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 3, pp. 357–371, October 2007.

[5] G. Hariharan, V. Raghavan, and A. M. Sayeed, "Capacity of Sparse Wideband Channels with Partial Channel Feedback," *European Transactions on Telecommunications*, vol. 19, no. 4, pp. 475–493, June 2008.

[6] D. Hammarwall, M. Bengtsson, and B. E. Ottersten, "Acquiring Partial CSI for Spatially Selective Transmission by Instantaneous Channel Norm Feedback," *IEEE Trans. Signal Processing*, vol. 56, no. 3, pp. 1188–1204, March 2008.

[7] ——, "Utilizing the Spatial Information Provided by Channel Norm Feedback in SDMA Systems," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 3278–3293, 2008.

[8] V. Raghavan, S. V. Hanly, and V. Veeravalli, "Statistical Beamforming on the Grassmann Manifold for the Two-User Broadcast Channel," *IEEE Trans. Information Theory*, vol. 59, no. 10, pp. 6464–6489, October 2013.

[9] J. Wang, M. Bengtsson, B. Ottersten, and D. Palomar, "Robust MIMO Precoding for Several Classes of Channel Uncertainty," *IEEE Trans. Signal Processing*, vol. 61, no. 12, pp. 3056–3070, April 2013.

[10] X. Zhang, D. P. Palomar, and B. Ottersten, "Robust MAC MIMO Transceiver Design with Partial CSIT and CSIR," in *IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC)* , Pacific Grove, CA, USA, 4-7 November 2007, pp. 324–328.

[11] M. Ding and S. D. Blostein, "Relation Between Joint Optimizations for Multiuser MIMO Uplink and Downlink with Imperfect CSI," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las vegas, NV, USA, March 31-April 4 2008, pp. 3149 – 3152.

[12] M. Médard, "The Effect upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel," *IEEE Trans. on Information Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[13] B. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links ?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[14] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the Capacity of Frequency-Selective Channels in Training-Based Transmission Schemes," *IEEE Trans. Signal Processing*, vol. 52, no. 9, pp. 2572–2583, September 2004.

[15] T. Marzetta, "How much training is needed for multiuser mimo ?" *IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC)*, pp. 359–363, June 2006.

[16] N. Jindal and A. Lozano, "A Unified Treatment of Optimum Pilot Overhear in Multipath Fading Channels," *IEEE Transactions on Communications*, vol. 58, no. 10, pp. 2939–2948, October 2010.

[17] B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode, "Adaptive antenna systems," *Proceedings of the IEEE*, vol. 55, no. 12, pp. 2143–2159, December 1967.

[18] S. R. Applebaum, "Adaptive arrays," *IEEE Transactions on Antennas and Propagation*, vol. 24, no. 5, pp. 585–598, September 1976.

[19] J. Winters, *Space-Time Wireless Systems – From Array Processing to MIMO Communications*, H. Bölcskei, D. Gesbert, C. B. Papadias, and A.-J. van der Veen, Eds. Cambridge University Press, 2006.

[20] J. H. Winters, "Optimum combining in digital mobile radio with cochannel interference," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 4, pp. 528–539, April 1984.

[21] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of mimo channels," *IEEE Journal on Selected Areas of Communications*, vol. 21, no. 5, pp. 684–702, June 2003.

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

[23] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in ofdm systems," *IEEE Transactions on Broadcasting*, vol. 48, no. 3, pp. 223–229, September 2002.

[24] Y.-H. Nam, Y. Akimoto, Y. Kim, M. i. Lee, K. Bhattad, and A. Ekpenyong, "Evolution of reference signals for lte-advanced systems," *IEEE Communications Magazine*, vol. 5, no. 2, pp. 132–138, February 2012.

[25] J. Choi, T. Kim, D. J. Love, and J.-Y. Seol, "Exploiting the preferred domain of fdd massive mimo systems with uniform planar arrays," in *IEEE International Conference on Communications*, London, UK, June 2015, pp. 3068–3073.

[26] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Trans. Wireless Comm.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[27] F. Rusek, D. Persson, B. K.Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling Up MIMO - Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, 2013.

[28] B. Gopalakrishnan and N. Jindal, "An analysis of pilot contamination on multiuser mimo cellular systems with many antennas," in *International Workshop on Signal Processing Advances in Wireless Communications*, San Francisco, June 2011, pp. 381–385.

[29] T. Kim and J. G. Andrews, "Optimal pilot-to-data power ratio for mimo-ofdm," *IEEE Globecom*, pp. 1481–1485, December 2005.

[30] ——, "Balancing pilot and data power for adaptive mimo-ofdm systems," *IEEE Globecom*, pp. 1–5, 2006.

[31] G. Fodor and M. Telek, "On the Pilot-Data Power Trade Off in Single Input Multiple Output Systems," *European Wireless '14*, vol. Barcelona, Spain, May 2014.

[32] K. Guo, Y. Guo, G. Fodor, and G. Ascheid, "Uplink Power Control with MMSE Receiver in Multi-Cell MU-Massive-MIMO Systems," in *IEEE International Conference on Communications (ICC)*, 10-14 June 2014, pp. 5184–5190.

[33] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A Coordinated Approach to Channel Estimation in Large-Scale Multiple-Antenna Systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264–273, February 2013.

[34] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory*. Prentice Hall, 1993, no. ISBN: 0133457117.

[35] R. Gallager, "Circularly-Symmetric Gaussian Complex Vectors," *http://www.rle.mit.edu/rgallager/documents/CircSymGauss.pdf*, 2008.

[36] N. Sun and J. Wu, "Maximizing spectral efficiency for high mobility systems with imperfect channel state information," *IEEE Transactions on Wireless Communication*, vol. 13, no. 3, pp. 1462–1470, March 2014.

[37] K. M. Z. Islam, T. Y. Al-Naffouri, and N. Al-Dhahir, "On Optimum Pilot Design for Comb-Type OFDM Transmission over Doubly-Selective Channels," *IEEE Trans. Comm.*, vol. 59, no. 4, pp. 930–935, April 2011.

[38] G. Fodor, P. D. Marco, and M. Telek, "On the Impact of Antenna Correlation on the Pilot-Data Balance in Multiple Antenna Systems," in *IEEE International Conference on Communications (ICC)*, London, UK, June 8-12 2015, pp. 2590–2596.

[39] A. Sezgin, E. A. Jorswieck, and E. Costa, "LDC in MIMO Ricean Channels: Optimal Transmit Strategy with MMSE Detection," *IEEE Trans. on Signal Processing*, vol. 56, no. 1, pp. 313–328, January 2011.

[40] N. Kim, Y. Lee, and H. Park, "Performance Analysis of MIMO System with Linear MMSE Receiver," *IEEE Trans. On Wireless Comm.*, vol. 7, no. 11, pp. 4474–4478, November 2008.

[41] E. Eraslan, B. Daneshrad, and C.-Y. Lou, "Performance Indicator for MIMO MMSE Receivers in the Presence of Channel Estimation Error," *IEEE Wireless Commun. Letters,*, vol. 2, no. 2, pp. 211–214, 2013.

[42] G. Fodor, P. D. Marco, and M. Telek, "Performance Analysis of Block and Comb Type Channel Estimation for Massive MIMO Systems," in *First International Conference on 5G*, Finland, November 2014.

[43] P. Soldati, M. Johansson, G. Fodor, and S. Sorrentino, "On Pilot Dimensioning in Multicell Single Input Multiple Output Systems," *IEEE Workshop on Broadband Wireless Access*, vol. Houston, TX, USA,, 2011.

[44] R. Chen, J. G. Andrews, R. W. Heath, and A. Ghosh, "Uplink Power Control in Multi-cell Spatial Multiplexing Wireless Systems," *IEEE Trans. on Wireless Communications*, vol. 6, no. 7, pp. 2700–2711, July 2007.

[45] G. Fodor, M. Telek, and P. Di Marco, "On the Impact of Antenna Correlation on the Pilot-Data Balance in Multiple Antenna Systems," *http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-143530*, March 2014.

[46] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A Coordinated Approach to Channel Estimation in Large-scale Multiple-antenna Systems," *http://arxiv.org/abs/1203.5924*, no. arXiv:1203.5924v1 [cs.IT], March 2012.

[47] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*. Academic Press, 2007, no. ISBN-13: 978-0-12-373637-6.

[48] L. Li, A. Ashikhmin, and T. Marzetta, "Pilot Contamination Precoding for Interference Reduction in Large Scale Antenna Systems," 51^{th} *Annual Allerton Conference*, vol. Allerton House, Illinois, USA, pp. 226–232, October 2013.

[49] E. Golovins and N. Ventura, "Optimal Training for the SM-MIMO-OFDM Systems with MMSE Channel Estimation," *6th Annual Communication Networks and Services Research Conference*, pp. 470–477, 2008.

[50] V. K. V. Gottumukkala and H. Minn, "Capacity Analysis and Pilot-Data Power Allocation for MIMO-OFDM With Transmitter and Receiver IQ Inbalances and Residual Carrier Frequency Offset," *IEEE Trans. Veh. Techn.*, pp. 553–565, 2012.

[51] K. T. Truong, A. Lozano, and R. W. H. Jr., "Optimal Training in Continuous Block-Fading Massive MIMO Systems," 20^{th} *European Wireless, Barcelona, Spain*, May 2014.

[52] B. Furht and S. A. Ahson, *Long Term Evolution: 3GPP LTE Radio and Cellular Technology*. Auerbach Publications, ISBN-10: 1420072102, April 2009.

[53] X. Hou, Z. Zhang, and H. Kayama, "DMRS Design and Channel Estimation for LTE-Advanced MIMO Uplink," 70^{th} *IEEE Vehicular Technology Conference (Fall)*, pp. 1–5, 2009.

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

[55] S. Wolfram, *The Mathematica Book*. Wolfram Media, ISBN 1-57955-022-3, 2003.

[56] X. Tang, M.-S. Alouini, and A. J. Goldsmith, "Effect of Channel Estimation Error on M-QAM BER Performance in Rayleigh Fading," *IEEE Transactions on Communications*, vol. 47, no. 12, pp. 1856–1864, December 1999.

[57] Y. G. Li, J. H. Winters, and N. R. Sollenberger, "MIMO-OFDM for Wireless Communications: Signal Detection With Enhanced Channel Estimation," *IEEE Transactions on Communications*, vol. 50, no. 9, pp. 1471–1477, September 2002.

[58] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*. WILEY, 2nd edition, 2011, ISBN-10: 0470660252.

[59] G. Fodor, P. D. Marco, and M. Telek, "On Minimizing the MSE in the Presence of Channel Information Errors," *IEEE Communications Letters*, vol. 19, no. 9, pp. 1604 – 1607, September 2015.

[60] M. F. Neuts, *Matrix Geometric Solutions in Stochastic Models*. Baltimore, MD, USA: The Johns Hopkins University Press, June 1981, no. ISBN-10: 0801825601.

[61] F. W. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, Eds., *NIST Handbook of Mathematical Functions*, 1st ed. New York, NY, USA: Cambridge University Press, 2010.

[62] P. D. Gallagher, "NIST Handbook of Mathematical Functions," http://dlmf.nist.gov/8.19, 2010, [Online; accessed 8-Oct-2015].

[63] N. M. Temme, "Uniform Asymptotic Expansions of the Incomplete Gamma Functions and the Incomplete Beta Function," *Mathematics of Computation*, vol. 29, no. 132, pp. 1109–1114, 1975, American Mathematical Society.

[64] 5G Americas, "Wireless technology evolution towards 5g: 3gpp release 13 to release 15 and beyond," www.5Gamericas.org, Tech. Rep., 2017.

[65] A. Osseiran, P. Marsch, and J. Monserrat, *5G Mobile and Wireless Communications Technolgy*. Cambridge University Press, 2016.