Novel Channel Estimation Methods under Pilot Contamination in Massive MIMO

By

Qingqing Cheng

A thesis submitted to Macquarie University

for the degree of Master of Research

Department of Engineering

March 2016

Examiner's Copy

Typeset in LaTeX 2_ε.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Qingqing Cheng

Acknowledgments

First and foremost, I would like to express my sincerest appreciation to my supervisors, Professor Stephen Hanly and Professor Iain Collings, who have constantly supervised and helped me during my Master of Research time. I could not finish the research work without their great insight, remarkable guidance, and encouragement during the last 9 months.

Furthermore, I would like to acknowledge Chunshan Liu and Min Li for their invaluable advice throughout my research work. I would like to express my thanks to my husband who has gave me endless support and love. I also forever indebted to my parents and other family members for their infinite love and help.

Finally, I would like to thank my friends who always helped and encouraged me during my master study.

Abstract

Massive multiple-input multiple-output (Massive MIMO) is a key feature of proposed 5G cellular systems, offering potentially many benefits. However, all benefits rest on the ability to obtain channel state information (CSI) at the base station (BS) during uplink transmission. The BS can in principle measure CSI from known user-transmitted pilot sequences, but in a multiple cell system, the use of non-orthogonal pilot sequences in different cells leads to a problem of pilot contamination.

In this thesis, we mainly focus on recently proposed covariance-aided channel estimation for time division duplex (TDD) Massive MIMO cellular systems suffering pilot contamination. The recent work assumes that the covariance matrices of users are known and do not change with time. In this thesis, we address two new scenarios: 1) the covariance matrix in one cell changes due to a new user arriving in that cell; 2) the covariance matrices of all users change due to mobility of the users. In these scenarios, we develop novel algorithms that estimate the new covariance matrices and then use these estimated covariance matrices to obtain high quality channel estimates. The proposed algorithms are compared with other estimation methods mentioned in this thesis to show the benefits of the proposed algorithms.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter comprises four sections. Section 1.1 presents the background on massive multiple-input multiple-output (Massive MIMO). Section 1.2 then provides motivation of this thesis. The objectives of this thesis are summarized in Section 1.3. Finally, in Section 1.4 the organization of this thesis is described.

1.1 Background

Massive MIMO system is a significant breakthrough, going beyond the conventional MIMO system. It was first proposed by Thomas L. Marzetta in 2007 [1] and was presented in more detail in 2010 [2, 3]. In current wireless systems, Massive MIMO is also known as large-scale antenna system, very large MIMO, hyper MIMO and full-dimension MIMO [2]. The most notable characteristic for this new technique is that the number of the antennas deployed at each base station (BS) in every cell is usually several tens or even more, and simultaneously serving tens of users. Moreover, for the same cell, the number of antennas at each BS greatly exceeds the number of active users in the same time-frequency resource [4].

The large number of antennas at each BS brings many advantages to Massive MIMO systems [5]. Compared with conventional MIMO systems, Massive MIMO systems have the ability to get much more system capacity and energy efficiency [2]. Furthermore, Massive MIMO systems offer robustness to interference such as intentional jamming and unintended man-made interference. Therefore, Massive MIMO systems become a

significant part of future 5G wireless communication systems [6]. More details about benefits of Massive MIMO systems will be provided in Section 2.1.2. However, as a developing technique, Massive MIMO potentially suffers from some problems, including pilot contamination and architectural challenges (further information will be described in Section 2.1.3). The most serious problem is pilot contamination which will be discussed in Section 2.1.4. This challenge has the potential to reduce the performance of Massive MIMO dramatically, and is an urgent problem that must be addressed in research on Massive MIMO.

1.2 Motivation

As discussed in the previous section, the challenge of pilot contamination needs to be overcome if we want to achieve the potentially benefits of Massive MIMO systems. Pilot contamination is caused by the reuse of the same or at least non orthogonal pilot sequences among different cells [7]. During uplink training period, users in each cell transmit known pilot sequences to the BS in their own cells and the BS estimates uplink channel state information (CSI) by these pilot information. For the time division duplex (TDD) Massive MIMO systems, the estimated CSI from uplink training can be used as downlink CSI due to the channel reciprocity that holds in TDD systems [3]. If pilot sequences in all cells are orthogonal with each other, then the BS is able to obtain good estimates of the desired CSI. However, the number of orthogonal pilot sequences is limited. Therefore, the same pilot sequences (or non orthogonal pilot sequences) are reused among different cells (the pilot sequences in the same cell are assumed orthogonal with each other) [8]. The re-use of pilot sequences leads to interference between different cells during uplink training, which is called pilot contamination. The focus of this thesis is on the issue of pilot contamination in Massive MIMO systems.

Among current methods proposed to deal with pilot contamination, channel estimation methods are more and more popular, especially the covariance-aided channel estimation methods [9]. These covariance-aided channel estimation methods make full use of assumed known channel statistics (channel covariance matrices). Good channel estimates can be obtained by these covariance-aided estimate methods under some

conditions [10]. However, these covariance-aided channel estimation methods require knowledge of each cell's channel covariance matrix, which is difficult to obtain due to the very same limitation of pilot contamination that occurs during uplink training.

The covariance matrix of each channel is usually assumed known[11]. However, in this thesis we consider some new scenarios in which at least some covariance matrices are unknown. In one scenario, a new user comes into a cell with an unknown covariance matrix. In another scenario, all the users' covariance matrices change due to the mobility of users. Under these scenarios, however, the method of estimating covariance matrices just mentioned may no longer be suitable (the reason will be discussed in Section 4.1). When channel covariance matrices are not available, some other channel estimation methods have been proposed [12]. For example, the authors in [13] proposed a blind channel estimation approach which estimates channel state information by decomposing the singular vectors of received signal's matrices. This blind method obtains good channel estimates, but the complexity is high due to the need for a singular value decomposition of received signal's matrices. A similar subspace based semi-blind channel estimation method is presented in [14]. This method exploits eigenvalue decomposition on the received signal's matrices to estimate the channel information. The estimated performance of this method is relatively good, but computational complexity is also high. Therefore, in this thesis we will estimate new covariance matrices under new scenarios and then use them in covariance-aided channel estimation methods to estimate CSI.

1.3 Research Objectives

The objectives of this thesis are to research covariance aided channel estimation methods under pilot contamination in Massive MIMO systems. In previous research on pilot contamination in Massive MIMO systems, estimation methods using covariance matrices have been shown to provide good estimation results under some conditions [15]. However, those methods are all based on the assumption that the covariance matrices can be estimated correctly and separately. But these methods will not be sufficient when covariance matrices change due to mobility of users or arrivals of new

users, which are the scenarios investigated in this thesis. More detailed aims are the following.

1) A clear background about the area of Massive MIMO is firstly presented, such as the development, the benefits, and the challenges of Massive MIMO. We focus on the particular problem of pilot contamination and discuss some measures which can tackle this problem;

2) Before researching new scenarios, this thesis studies channel estimation methods under situations in which the channel covariance matrices keep constant. Under this situation, this thesis analyzes the effect of pilot contamination on system performance and discusses several existing channel estimation methods: Least Squares (LS), Minimum Mean Square Error (MMSE) and Maximum A Posteriori (MAP) estimation methods. The conclusion from comparing these channel estimation methods is that covariance-aided channel estimation methods are much superior to conventional methods without covariance matrices.

3) Then this thesis concentrates on channel estimation methods under new scenarios in which covariance matrices are not directly available due to the mobility of users or arrivals of new users. For these two cases we propose corresponding novel algorithms to obtain the estimate of new covariance matrices and estimate the CSI based on estimated covariance matrices. We provide both analytical work and numerical experiments to test the performance of our proposed algorithms.

1.4 Organization of the Thesis

The rest of this thesis is organized as follows.

Chapter 2 provides a review on Massive MIMO systems. In Section 2.1, the historic development of Massive MIMO systems is presented, then the benefits and challenges of Massive MIMO systems are shown. Some existing methods for dealing with the problem of pilot contamination are described in Section 2.2, which includes three main aspects: the precoding process, pilot operations and channel estimation methods.

Chapter 3 analyzes the problem of pilot contamination in Massive MIMO systems and some existing channel estimation methods. Section 3.1 discuss the system model used in Chapter 3. Based on the system model, Section 3.2 introduces some existing channel estimation methods, and gives the advantages and disadvantages of each method. Section 3.3 analyze the problem of pilot contamination in Massive MIMO systems.

Chapter 4 studies channel estimation methods in new scenarios. Section 4.2 provides the system model for the new scenarios. Two new scenarios are described in Section 4.3 and corresponding novel estimation methods for these cases are proposed: Section 4.3.1 studies the situation that new covariance matrices are not directly available due to arrivals of new users; Section 4.3.2 studies the situation in where all covariance matrices change due to mobility of users. The performance of proposed methods is presented in Section 4.4.

Finally, Chapter 5 is the conclusion and future work. Section 5.1 summarizes the research of this thesis and Section 5.2 recommends related future work.

Chapter 2

Literature Review

Since the focus of this thesis is the problem of pilot contamination in Massive MIMO systems, in this chapter, we will give an introduction to Massive MIMO systems in Section 2.1, then pilot decontamination methods are reviewed in Section 2.2.

2.1 Overview of Massive MIMO

Massive MIMO is a key component of proposed 5G cellular systems, which has the potential to provide capacity to meet ever-growing data requirements of wireless devices. However, as an emerging technology, Massive MIMO also suffers some challenges. The most serious is pilot contamination, which can greatly reduce the prospective gains of Massive MIMO.

In this section, the historic development of Massive MIMO is presented in Section 2.1.1. The benefits and challenges of Massive MIMO are discussed in Section 2.1.2 and Section 2.1.3, respectively. The problem of pilot contamination is given in Section 2.1.4.

2.1.1 The Historic Development of Massive MIMO

The unceasing requirements for high data rates and high service quality is a challenge for current wireless communication industry. One way to overcome this challenge is to increase capacity of system several times[16].

It has been well known since 1995 that under complex Gaussian fading model with independent and identically distributed (i.i.d.) channels between each transmitter and receiver pair, the capacity of point-to-point multiple-input multiple-output (point-to-point MIMO) system increases linearly with the minimum number of transmit antennas and receive antennas [17, 18]. Researchers later understood that similar results would hold even if each receiver had only one antenna, provided that the number of users were large. This gave rise to the area of Multiple-user MIMO in which each BS serves several users [19].

Multi-user MIMO can provide a large number of advantages compared with point to point MIMO. Since every user (active terminal) can utilizes all the time-frequency bins, the resource allocation can be simplified. It also does not need the assumption of rich scattering environment. Moreover, if the BS knows the CSI, it is possible to get arbitrarily high capacity (sum-rate) in a single cell, even with only single antenna users, provided that each BS has sufficiently large antenna array. This is true even with simple linear beamforming techniques such as Zero Forcing beamforming (ZF beamforming). If ZF beamforming is used: the sum-rate of system from ZF beamforming increases linearly with the number of antennas at the BS, provided that the number of users also scales up with the number of antennas at the BS (in constant ratio, and the ratio is less than unity so that Zero Forcing is possible). In this case, the data rate each user gets does not grow large, but the sum-rate grows large because of the large number of users [19]. Fig. 2.1 describes the idea of multi-user MIMO systems. In this figure the BS serves a lot of users using the same time and frequency resources. The BS has a antenna array which displays with several antennas, and each user has one antenna.

Recent work has extended these results to multiple cell setting, including multi-cell MIMO cooperation [20, 21]. It can be seen from [20] that if the number of users is held fixed, the sum-rate of the system increases with the increase of the number of antennas at the BS, provided that the users themselves have antenna arrays which grow large and the number of data stream per user then needs to grow large (the per-user data rate grows large in this case). In addition, a lot of work was done on multi-user MIMO in the early 2000s, and good summaries can be found in [22, 23]. [23] showed that

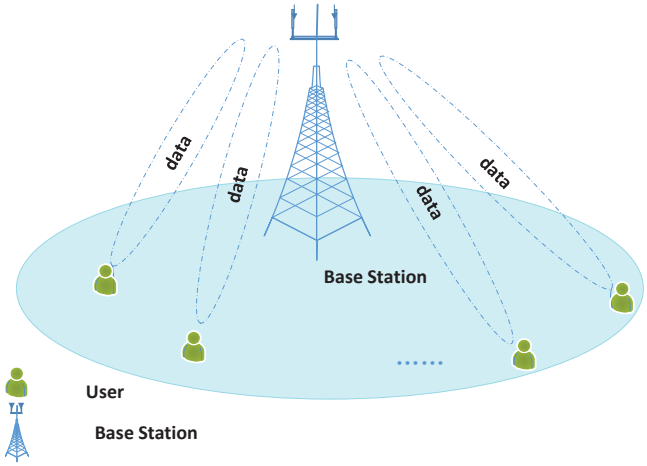

Figure 2.1: A multi-user MIMO system.

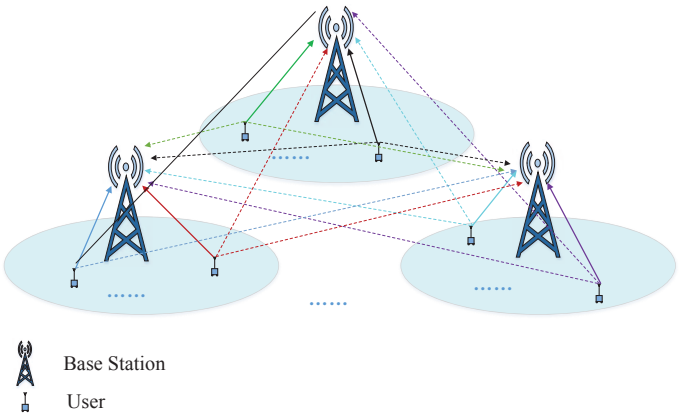

Figure 2.2: A multi-cell multi-user MIMO system.

at high signal to noise ratio (SNR) and if the CSI is known perfectly at the BS, then it is optimal to do linear Zero Forcing beamforming to carefully selected users who's channels are then approximately orthogonal. Fig. 2.2 illustrates the operation of a multi-cell multi-user MIMO system. There are many cells in the system. In each cell, the BS serves many single-antenna users simultaneously (using the same time and frequency resources), the BS has a array with several antennas. Each BS can receive signal from all users in all cells.

MIMO systems equipped with multiple antennas at both BSs and users have much higher sum-rate than corresponding single-antenna systems. Generally speaking, the sum-rate of MIMO systems increases linearly with the minimum number of antennas

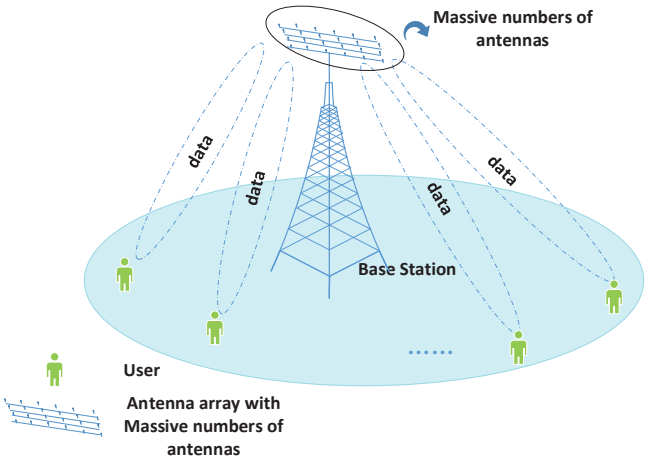

Figure 2.3: A single-cell Massive MIMO system.

both at each BS and users, as long as the channel knowledge can be available for both communication sides [24, 25]. However, for conventional MIMO systems, the number of antennas is limited. For example, according to the Long Term Evolution (LTE) standard, the data rates during uplink communication is 75 Mbps using 20 MHz of spectrum with 4×4 antennas [26]. Unfortunately, this spectral efficiency is far from enough to meet the requirement of future 5G wireless communication system. Therefore, it is time that a new technique be developed which is able to supply much higher data rate than a traditional MIMO system. The Massive MIMO system, which has much stronger ability to increase the data rate of system, becomes an excellent candidate.

One of the most remarkable features of Massive MIMO systems is that the antenna array at each BS consists of hundreds antennas instead of several antennas at each BS in conventional MIMO systems [27]. Figure 2.3 shows a single-cell Massive MIMO system, where the antenna array consists of a massive number of antennas.

There are two typical Massive MIMO system models: the frequency division duplex (FDD) Massive MIMO system and the time division duplex (TDD) Massive MIMO system [28]. For FDD Massive MIMO systems, the uplink and downlink transmission are in different frequency bands. Moreover, if terminals (users) support full-duplex, both uplink and downlink transmission can happen simultaneously. For TDD Massive MIMO systems, uplink and downlink transmission share the same frequency band but

occur at different times.

For FDD Massive MIMO systems, obtaining the CSI for a downlink channel requires a two-stage procedure. More precisely, the BS transmits pilot sequences to all users in the cell, and then all users need to feedback the downlink estimated CSI to that BS. The time for transmitting pilot sequences in FDD systems increases with the number of antennas at the BS. Since the quantity of antennas at each BS is extremely large in Massive MIMO systems, the time for channel estimation is corresponding long. Therefore, it is practically impossible to obtain global real-time CSI in any reasonable delay, because the channel will change before the adequate CSI is estimated and signaling latency is also un-avoidable over the downlink transmission [28]. Moreover, for FDD systems, the quantity of independent pilot sequences, which are needed for estimating CSI, are ever-growing with the number of antennas at each BS [29]. Since each BS has a large number of antennas, the required number of independent pilot sequences is correspondingly very large [30, 31].

In contrast, TDD systems can exploit channel reciprocity: the CSI on the downlink is the same as the uplink, so the CSI on the downlink can be measured at the BS using uplink pilot sequences sent by users [32]. Therefore, the BS can estimate the uplink CSI from the uplink transmissions, and this estimated CSI from the uplink is also an estimate of the downlink CSI which can be used by the linear pre-coders during the downlink transmissions. Moreover, for TDD systems, the required number of pilot sequences increases linearly with the number of users in every cell instead of antennas at each BS. Therefore, when the quantity of users in each cell is not so large, the required number of pilot sequences is also quite limited. Therefore, TDD Massive MIMO systems can improve system performance by increasing the number of antennas at the BS with limited overheads [29].

2.1.2 The Benefits of Massive MIMO

Thanks to the large number of antennas at each BS, Massive MIMO systems have more features compared with traditional MIMO systems, which are listed as following.

1) It is possible for Massive MIMO systems to increase the capacity of system for 10

times or more compared with traditional MIMO systems[3];

2) Massive MIMO systems are able to robust to multi-user interference, channel fading and hardware imperfection [2];

3) Massive MIMO systems have a large number of antennas at the BS, which enable to reduce or even average out the effect of channel fading [33];

4) The channels between different users become approximately orthogonal thanks to enormous numbers of antennas at the BS [3];

5) Massive MIMO systems can achieve good system performance just based on simple linear precoders (downlink)[3];

6) A large number of antennas at the BS provides robustness against individual antenna failure;

7) Massive numbers of antennas at each BS provide a large surplus of degrees of freedom, which is helpful to system hardware-friendly signal shaping [2];

8) Massive MIMO systems can be built with cheap and low-power elements instead of expensive and high-power items (such as high-power amplifiers and large coaxial cables) in traditional MIMO systems [2];

9) Massive MIMO systems can simplify the multiple access layer[2];

10) Massive MIMO systems are able to raise energy efficiency for several times [2].This benefit results from the large number of antennas at the BS, because these antennas can focus energy on small space area if coherently combined (uplink) or precoded (downlink) using beamforming techniques. The more antennas installed at each BS, the less transmitted power is needed per user, as long as CSI can be estimated reasonably well. According to [34], each BS is equipped with M antennas (M increases without limit), the transmitted power of per user can be reduced proportionally to $1/\sqrt{M}$ with estimated CSI using MMSE estimation method;

2.1.3 Challenges of Massive MIMO

Although Massive MIMO can offer many benefits, as a new technique, it also suffers some challenges [2].

In Massive MIMO systems, each BS has an antenna array equipped with perhaps hundreds of antennas. These numerous antennas can bring great deal of benefits to wireless communication system. However, it is very difficult to install a large number of antennas due to the quite limited available space at the antenna array. If the number of antennas is not large enough, we can not assume orthogonality of channels between different users, and it is then impossible to obtain good system performance based on simple transmitters or receivers[29, 35]. The authors of [36] showed that if each BS has a large number of antennas, the performance of the low-complexity matched filter (MF) is close to the high-complexity minimum mean-square error (MMSE) filter. However, if the number of the antennas at the BS is not large enough, then the performance of the MF is dramatically worse than the MMSE filter.

Besides, there are so many radio frequency (RF) elements such as mixers and amplifiers in each antenna array. These elements add to the energy consumption in total, and can even negate the theoretical benefits of power saving from having a massive number of antennas [34].

Almost all benefits of Massive MIMO systems discussed in section 2.1.2 are based on the basic assumption that the BS can obtain good estimates of the CSI from uplink training by sending known pilot sequences from users. To obtain good estimates of CSI, pilot sequences of different users are required to be orthogonal with each other. However, it is difficult to meet this requirement for multi-cell multi-user Massive MIMO systems, which will be discussed in section 2.1.4. Therefore, the same pilot sequences are likely reused in different cells, or the pilot sequences between different cells are partially correlated. Unfortunately, this then reduces the accuracy of channel estimation, which in turn decreases system performance such as capacity and energy efficiency. The effect of non-orthogonal pilot sequences used in different cells is termed pilot contamination (more details will be discussed in Section 2.1.4).

Besides architectural challenges and pilot contamination, Massive MIMO systems

also suffer some other challenges, such as hardware impairments, cost of reciprocity calibration and power consumption [37]. These challenges depend on the hardware implementation, so they are not really fundamental issues. Among these challenges, pilot contamination is the most serious, therefore, this thesis concentrates on the issue of pilot contamination in Massive MIMO systems.

2.1.4 Pilot Contamination

In wireless communication systems, the BS needs to estimate the CSI of each user using pilot sequences, and then use the estimated CSI to detect uplink data and to make precoders for downlink transmissions. To get good estimates of CSI, the number of orthogonal pilot sequences should equal to the number of users. In other words, the number of orthogonal pilot sequences in the system should ideally be $U \times L$, where U is the number of users in each cell, L is the number of cells in the system. In order to maintain orthogonality of each pilot sequence, the length of the sequence must increase, and the time for transmitting the pilot sequence also increases. However, in a practical wireless system, the coherence interval (the CSI keeps the same in the coherence interval) is limited, and only a fraction of this interval can be devoted to pilot symbols, the rest being reserved for data. So the length of pilot sequence can not be too long, which implies that the maximum number of orthogonal pilot sequences is limited [4]. However, for multi-cell multi-user Massive MIMO systems, since the number of users in the system is very large (all users send pilot sequences to BSs simultaneously), the required number of orthogonal pilot sequences are also large. Therefore, the practical number of available orthogonal pilot sequences are much less than the requirement. Moreover, we want to obtain the maximum sum data-rate of the cell , so we assign orthogonal pilot sequences to users in the same cell. Therefore, the same or partially correlated pilot sequences are reused in different cells.

Since the pilot sequences are transmitted at the same time and frequency, the BS receives the linear combination of all pilot sequences. If these pilot sequences are orthogonal with each other, then the BS can obtain good estimates of the CSI based on these unique pilot sequences. However, if two or several cells share the same

pilot sequence or different cells use non-orthogonal pilot sequences, then the BS can not distinguish the pilot sequence of the desired cell. Therefore, the estimate of the desired CSI at the BS will be contaminated by these same or non-orthogonal pilot sequences from adjacent cells. These incorrect estimates of CSI will then have negative effect on downlink beamforming and overall system performance, which termed as pilot contamination [7]. What is worse, this interference caused by pilot contamination will not diminish even when the number of antennas at the BS increases [2]. It has been proven that in Massive MIMO systems, with the help of numerous antennas, other detrimental factors such as noise and fading can be mitigated by increasing the number of antennas at the BS [2]. Thus the main factor which limits Massive MIMO system performance is the pilot contamination. Fig. 2.4 describes the pilot contamination in Massive MIMO systems. In Fig.2.4 every BS can receive the combination of all pilot sequences. Take the 1st BS as an example, it receives the combination of all pilot sequences from all users. If there is one user sends P_1 pilot sequence in each cell, then the 1st BS can received the combination of P_1 pilot sequence. The channel estimate of user A at the 1st BS is contaminated by those users sharing the same pilot sequence located in adjacent cells.

Pilot contamination can negate the theoretical benefits of increased system capacity, improved energy efficiency and other benefits from having a large number of antennas at the BS. In summary, pilot contamination seriously affects the performance of Massive MIMO systems [38]. Therefore, the problem of pilot contamination must be dealt with.

2.2 Methods of Pilot Decontamination

Recently, pilot contamination has attracted substantial attention and different methods of pilot decontamination have been proposed. Among various existing methods of handling pilot contamination, pilot assignment, channel estimation methods and precoding process are three key aspects. In this section, we introduce the pilot assignment in section 2.2.1, followed by channel estimation methods in section 2.2.2, and the precoding process is discussed in section 2.2.3.

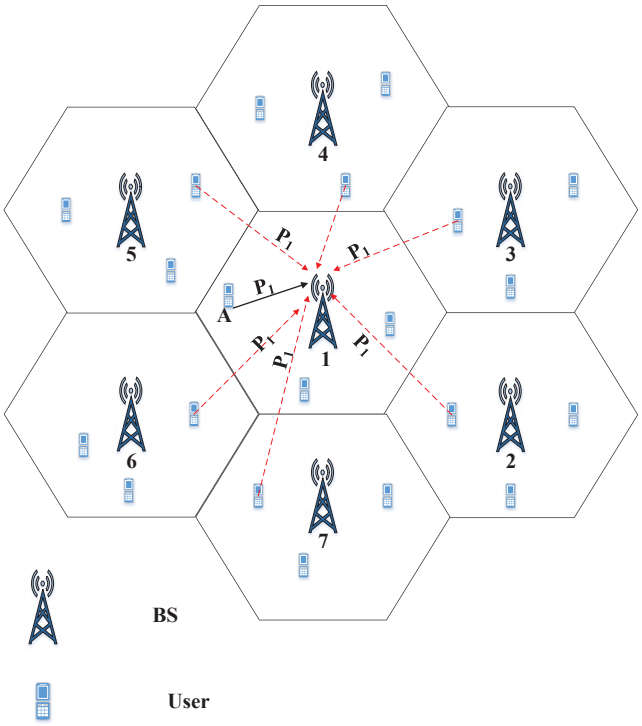

Figure 2.4: Pilot contamination in Massive MIMO systems.

2.2.1 Pilot Assignment

The pilot assignment is an effective mechanism to reduce the negative effect of pilot contamination in Massive MIMO systems [39]. The idea of this method is to co-ordinate the allocation of pilot sequences across all cells to reduce the correlation between the pilots as observed at each BS. A coordinated approach of pilot assignment was proposed in [11]. The pilot assignment methods in this paper made use of the mean square error of channel estimate which related to channel covariance matrices. This pilot assignment method can offer a powerful way to reduce the negative influence of pilot contamination on Massive MIMO systems. However, this pilot assignment method in [11] was effective only when there is no overlap of angle of arrival (AOA) between different multipaths and the good estimates of covariance matrices can be obtained (more detail will be given in Section 3.2.3).

The authors in [40] proposed a pilot reuse scheme which used the relationship between channel spectrum correlations and channel power angle spectrum. This pilot reuse scheme can reduce pilot contamination. But this method requires that the AOA

of different paths from different users do not overlap. The authors in [41] proposed a scheme in which antennas selection and user scheduling are done jointly. This scheme had the ability to obtain a good system performance, but the computation complexity was very high which was a serious issue for this approach.

2.2.2 The Channel Estimation Methods

Drawing support from channel estimation methods to reduce the negative effect of pilot contamination on the system performance has become more and more popular. In this section, we introduce some estimation approaches which try to tackle the problem of pilot contamination.

When pilot sequences are not orthogonal, some methods have been proposed to estimate CSI. The research reported in [13] and [42] studied blind pilot decontamination methods, which estimate CSI using receive signal covariance matrices. These methods were based on random matrix theory to predict eigenvalue spectra of receive signal covariance matrices. They can obtain better estimation results than conventional channel estimation methods (such as LS methods), but the computational complexity of these methods are very high.

The covariance-aided channel estimation methods have got more and more attention. In [11] and [43], estimation approaches were studied exploiting channel covariance matrices. The authors of [14] also made full use of covariance matrices of channels to mitigate the negative influence of pilot contamination on system performance. The authors indicated that, with the help of covariance matrices, the system can obtain good estimates of CSI and get much better system performance than those methods which estimate CSI just based on the pilot sequences (such as LS estimation methods).

Although these covariance-aided estimation methods can obtain good estimates of CSI, they are difficult to obtain the channel covariance matrices. For current covariance-aided channel estimation approaches, the usual focus is on the scenario that covariance matrices of channels remain the same. For this situation, the methods of estimating covariance matrices have been studied in [11]. However, for the other important scenarios: covariance matrices change due the mobility of users or arrivals of

new users, the method of tracking changing covariance matrices needs to be researched further.

2.2.3 The Precoding Process

Reducing the negative influence of pilot contamination on system performance by some suitable precoding methods has received more and more attention [44]. A precoding method of multi-cell TDD system was discussed in [45]. The authors proposed a multi-cell minimum mean-squared error based (MMSE-based) precoding method. The authors accounted for the inter-cell interference (pilot contamination) and pilot allocation when designing the MMSE-based precoder, so that this method can obtain better performance than conventional zero-forcing (ZF) precoding method under pilot contamination. But the complexity of this MMSE-based precoding method is very high. The authors in [28] considered FDD multi-cell Massive MIMO system and proposed a two-tier precoding strategy to deal with the inter-cell interference (pilot contamination) and the intra-cell interference. However, the precoding method in [28] needs a huge number of independent pilot sequences to get estimates of real-time global CSI at the BS to mitigate the pilot contamination. This requirement is practically infeasible for Massive MIMO systems due to the limited channel coherence time. A two-stage subspace constrained precoding method was proposed in [46]. The author made clear that the precoding method in [46] was able to obtain good performance. However, the optimization of subspace dimension partitioning was much complicated and was a serious challenge for this method.

2.3 Conclusion

In this chapter, basic background of Massive MIMO was firstly presented. Then we described a serious problem: pilot contamination. After that we discussed pilot decontamination methods. We can conclude that if covariance matrices of channels are fixed, then covariance-aided channel methods can obtain good channel estimates. However, for new scenarios: covariance matrices change due to the mobility of users or new arrivals of users, covariance-aided channel estimation methods need to be studied further.

Chapter 3

Channel Estimation Methods in Massive MIMO

As described in chapter 2, a Massive MIMO system is developed based on a conventional MIMO system but where the BS is equipped with a very large number of antennas [47, 48]. Massive numbers of antennas at each BS provide Massive MIMO with several advantages. It should be noted that these benefits rely on the BS obtaining good estimates of the CSI of each user. However, for Massive MIMO systems it is very difficult to accurately estimate CSI due to pilot contamination. In this chapter we review techniques for estimating CSI under pilot contamination.

In Section 3.1, we introduce system model of Massive MIMO that we will use in this thesis. Some channel estimation methods are described in Section 3.2. Section 3.3 presents the reason and effect of pilot contamination.

3.1 System Model of Massive MIMO

In this section, a TDD Massive MIMO system which consists of L time-synchronized cells is considered. The synchronization between uplink pilots provides a worst case scenario from a pilot contamination point of view, since any lack of synchronization will tend to statistically de-correlate the pilots. The channel in this chapter is block-fading and we focus on channel estimation in one block in this chapter [49]. Therefore, the statistics of channels (covariance matrices of channels) are assumed fixed in this

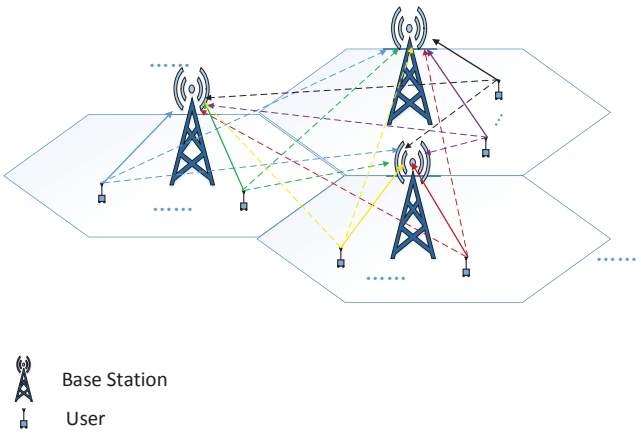

Figure 3.1: The system model of Massive MIMO.

chapter. There is one BS in each cell serving U single-antenna users. Each BS is equipped with M-element uniform linear array (ULA) of antennas. All users in all cells share the same time-frequency resource. Furthermore, in this chapter, pilot sequences used in the same cell are mutually orthogonal and the length for each pilot sequence is τ. However, pilot sequences used in different cells are not orthogonal. In other words, the intra-cell interference is negligible but the inter-cell interference can not be ignored during the channel estimation phase.

During uplink transmission (including uplink training and uplink data transmission), every single-antenna user transmits a pilot sequence or data signal to the BS in its own cell and every BS can receive pilot sequences from all users in all cells. The BS can estimate downlink CSI by uplink training thanks to the great feature of TDD system: channel reciprocity. The system model is shown in Figure 3.1.

During uplink training phase, each BS receives all pilot sequences from all users. Each pilot sequence lasts for τ samples. Let the τ samples received at the mth antenna of the jth BS be represented by the $1 \times \tau$ row vector $\mathbf{y}_{jm}$. Then

$$\mathbf{y}_{jm} = \sum_{l=1}^{L} \sum_{u=1}^{U} h_{jlum} \mathbf{s}_{lu}^{T} + \mathbf{n}_{jm}, \tag{3.1}$$

where h_{jlum} is the channel complex value between the mth antenna of the jth cell BS and the uth user in the lth cell. $\mathbf{n}_{jm}$ denotes the noise at the mth antenna of the jth

cell BS, which is the vector of independent and identically distribution (i.i.d.) zero-mean, circularly-symmetric complex Gaussian random variables. The noise is a $1 \times \tau$ vector and unknown for both BS and users. $\mathbf{s}_{lu}$ is a $\tau \times 1$ real-valued column vector of pilot symbols which is transmitted from the uth user in the lth cell. Here, we stress that the length of $\mathbf{s}_{lu}$ is τ, and the energy of $\mathbf{s}_{lu}$ is

$$|s_{lu1}|^2 + |s_{lu2}|^2 + \ldots + |s_{lu\tau}|^2 = \tau. \tag{3.2}$$

The equation (3.1) can be denoted as

$$\mathbf{y}_j = \sum_{l=1}^{L} \mathbf{h}_{jl}\mathbf{s}_l^T + \mathbf{n}_j, \tag{3.3}$$

where $\mathbf{y}_j = [\mathbf{y}_{j1}^T \quad \mathbf{y}_{j2}^T \ldots \mathbf{y}_{jM}^T]^T$ is the $M \times \tau$ matrix, $\mathbf{n}_j = [\mathbf{n}_{j1}^T \quad \mathbf{n}_{j2}^T \ldots \mathbf{n}_{jM}^T]^T$ is also a $M \times \tau$ matrix. The pilot sequences $\mathbf{s}_l = [\mathbf{s}_{l1} \quad \mathbf{s}_{l2} \ldots \mathbf{s}_{lU}]$ ($\tau \times U$ matrix), $\mathbf{h}_{jl}$ is the $M \times N$ matrix given by

$$\mathbf{h}_{jl} = \begin{pmatrix} h_{jl11} & \ldots & h_{jl1U} \\ \vdots & \ddots & \vdots \\ h_{jlM1} & \ldots & h_{jlMU} \end{pmatrix}. \tag{3.4}$$

Let $\mathbf{h}_{jlu}$ denote the uth column of $\mathbf{h}_{jl}$, then $\mathbf{h}_{jlu}$ can be written as

$$\mathbf{h}_{jlu} = \sqrt{\beta_{jlu}} * \boldsymbol{\varphi}_{jlu}, \tag{3.5}$$

where the $\boldsymbol{\varphi}_{jlu}, u = 1, 2 \ldots U$ are i.i.d. complex random variable distributed as $\mathcal{CN}(0, 1)$ with zero mean and unit variance. The β_{jlu} is the large-scale attenuation (geometric attenuation and shadowing) which can be presented as

$$\beta_{jlu} = \frac{\alpha}{d_{jlu}^{\gamma}}, \tag{3.6}$$

where the γ is gain factor exponent. The α is a constant that determined by the cell-edge signal to noise ratio (SNR). The d_{jlu} is distance between the user in the lth cell to the jth BS. For the same BS, the values of the large-scale attenuation are the same from different antennas. This assumption makes sense because the size of the antenna array is very insignificant when compared with distance between the BS and users.

3.2 Channel Estimation Methods

The $\mathbf{h}_{jl}$ ($j = 1, 2, ..L$, $l = 1, 2, ..L$) is the CSI between the BS in the jth cell and users in the lth cell. In this section, we will introduce some existing channel estimation methods that can be used to estimate the CSI: the Least Square (LS) estimation method, Minimum Mean Square Error (MMSE) estimation method and Maximum A Posteriori (MAP) estimation method. In the same cell pilot sequences are mutually orthogonal, but in other cells the same pilot sequences are reused. We focus on the performance of users that share the same pilot sequence, as they are not affected by other users. Therefore, in this section, there is one user ($U = 1$) in every cell(refer to equation 3.12). For this reason, the $\mathbf{h}_{jl}$ in equation (3.4) is $M \times 1$ column vector.

3.2.1 Least Square Estimation Method

Least Squares (LS) estimation method does not use any prior knowledge about the channel, basing the estimates entirely on the received matrices $\mathbf{y}_j$. This makes this method robust to lack of knowledge about the CSI. We now describe the LS estimation of the CSI.

According to equation (3.3), the received signal at the BS is the combination of all users' pilot sequences. Therefore, equation (3.3) can be written in another way:

$$\mathbf{y}_j = \mathbf{h}_{jj}\mathbf{s}_j^T + \sum_{\substack{f=1 \\ f \neq j}}^{L} \mathbf{h}_{jf}\mathbf{s}_f^T + \mathbf{n}_j \tag{3.7}$$

where $j = 1, 2, ..L$, $f = 1, 2, ..L$, the first term in equation (3.7) corresponds to the channels from the jth BS and users in the jth cell. The second term is from interfering users in the fth cell to the jth BS ($f \neq j$), and the third term is white, Gaussian receiver noise at the jth BS.

The conventional LS methods estimate $\mathbf{h}_{jl}$ by minimizing a LS error criterion which is the squared difference between received signal $\mathbf{y}_j$ and assumed signal $\mathbf{h}_{jl}\mathbf{s}_l^T$. The error criterion of LS methods is

$$J(\mathbf{h}_{jl}) = (\mathbf{y}_j - \mathbf{h}_{jl}\mathbf{s}_l^T)(\mathbf{y}_j - \mathbf{h}_{jl}\mathbf{s}_l^T)^H. \tag{3.8}$$

The gradient of $J(\mathbf{h}_{jl})$ is

$$\frac{\partial J(\mathbf{h}_{jl})}{\partial \mathbf{h}_{jl}^H} = -\mathbf{y}_j \mathbf{s}_l + \mathbf{h}_{jl} \mathbf{s}_l^T (\mathbf{s}_l^H)^T. \tag{3.9}$$

Setting the gradient equal to zero yields the LS estimate

$$\hat{\mathbf{h}}_{jl}^{LS} = \mathbf{y}_j \mathbf{s}_l ((\mathbf{s}_l^H \mathbf{s}_l)^T)^{-1}. \tag{3.10}$$

The $\mathbf{s}_l$ in this thesis are Walsh codes, and they are real-valued column vectors, so $\mathbf{s}_l^H \mathbf{s}_l = \tau$. Therefore, equation (3.10) can be written

$$\hat{\mathbf{h}}_{jl}^{LS} = \frac{1}{\tau} \mathbf{y}_j \mathbf{s}_l. \tag{3.11}$$

We assume all users in the system use the same pilot sequence, which can be shown as

$$\mathbf{s}_1 = \mathbf{s}_2 \ldots = \mathbf{s}_L = \mathbf{s}. \tag{3.12}$$

Under this situation, the LS estimate of $\mathbf{h}_{jl}$ is

$$\hat{\mathbf{h}}_{jl}^{LS} = \mathbf{h}_{jl} + \sum_{f \neq l}^{L} \mathbf{h}_{jf} + \mathbf{n}_j \mathbf{s} / \tau. \tag{3.13}$$

It can be seen from equation (3.13) that, the channel estimate of desired channel $\mathbf{h}_{jl}$ is directly influenced by $\sum_{f \neq l}^{L} \mathbf{h}_{jf}$ from interfering cells, which is called pilot contamination. Besides, the more serious the pilot contamination is (the more non-orthogonal pilot sequences used in the estimation process), the worse LS estimate gets. In summary, under the pilot contamination, the LS estimation method can not obtain good estimates of desired channels.

3.2.2 Minimum Mean Square Error Estimation Method

The minimum mean square error estimation (MMSE estimation) method takes the statistics of channels into consideration when estimating the CSI. The estimation process of MMSE is described as follows.

The MMSE estimate of desired channel is

$$\hat{\mathbf{h}}_{jl}^{MMSE} = \mathbf{y}_j \hat{\mathbf{\Omega}}_{jl}. \tag{3.14}$$

The $\hat{\mathbf{\Omega}}_{jl}$ is the MMSE filter, which is a $\tau \times 1$ vector of complex-valued coefficients. The principle of MMSE estimation is to get the minimum mean square error between the estimated value and the true value, which can be written as

$$\hat{\mathbf{\Omega}}_{jl} = \arg\min_{\hat{\Omega}} \mathrm{E}\| \mathbf{h}_{jl} - \hat{\mathbf{h}}_{jl}^{MMSE} \|_F^2, \tag{3.15}$$

where the $\| \cdot \|_F$ denotes the Frobenius norm.The mean square error (MSE) of the estimate is given by

$$\begin{aligned} \mathcal{M}_{jl}^{MMSE} &= \mathrm{E}\{\| \mathbf{h}_{jl} - \hat{\mathbf{h}}_{jl} \|_F^2\} \\ &= \mathrm{E}\Bigg\{ tr\Bigg\{ \Bigg(\mathbf{h}_{jl} - \Big((\mathbf{h}_{jl} + \sum_{f \neq l}^{L} \mathbf{h}_{jf}) \mathbf{s}_l^H + \mathbf{n}_j \Big) \hat{\mathbf{\Omega}}_{jl} \Bigg) \\ &\quad \times \Bigg(\mathbf{h}_{jl}^H - \hat{\mathbf{\Omega}}_{jl}^H \Big(\mathbf{s}_l (\mathbf{h}_{jl}^H + \sum_{f \neq l}^{L} \mathbf{h}_{jf}^H) + \mathbf{n}_j^H \Big) \Bigg) \Bigg\} \Bigg\}. \end{aligned} \tag{3.16}$$

Based on equation(3.16), taking partial derivative of $\mathcal{M}_{jl}^{MMSE}$ with respect to the conjugate transpose of $\mathbf{\Omega}$

$$\frac{\partial \mathcal{M}_{jl}^{MMSE}}{\partial \mathbf{\Omega}_{jl}^H} = \mathbf{s}_l \mathrm{E}\{\mathbf{h}_{jl}^H \mathbf{h}_{jl}\} - \Big(\sum_{l=1}^{L} \mathbf{s}_l \mathrm{E}\{\mathbf{h}_{jl}^H \mathbf{h}_{jl}\} \mathbf{s}_l^H + M\omega^2 \mathbf{I} \Big) \hat{\mathbf{\Omega}}_{jl}. \tag{3.17}$$

Since $\mathbf{h}_{jl}$ is a $M \times 1$ vector, $\mathrm{E}\{\mathbf{h}_{jl}^H \mathbf{h}_{jl}\}$ is a scalar which is denoted below by E_{jl}. Let $\frac{\partial \mathcal{M}_{jl}^{MMSE}}{\partial \mathbf{\Omega}_{jl}^H} = 0$ and get result, which is

$$\begin{aligned} \hat{\mathbf{\Omega}}_{jl}^{MMSE} &= \Big(\sum_{l=1}^{L} \mathbf{s}_l \mathrm{E}_{jl} \mathbf{s}_l^H + M\omega^2 \mathbf{I} \Big)^{-1} \mathbf{s}_l \mathrm{E}_{jl} \\ &= \Big(\mathbf{s}_l \mathrm{E}_{jl} \mathbf{s}_l^H + \sum_{f \neq l}^{L} \mathbf{s}_f \mathrm{E}_{jf} \mathbf{s}_f^H + M\omega^2 \mathbf{I} \Big)^{-1} \mathbf{s}_l \mathrm{E}_{jl}. \end{aligned} \tag{3.18}$$

Therefore, according to equation (3.14) and equation (3.18), the MMSE estimator is

$$\hat{\mathbf{h}}_{jl}^{MMSE} = \mathbf{y}_j \Big(\mathbf{s}_l \mathrm{E}_{jl} \mathbf{s}_l^H + \sum_{f \neq l}^{L} \mathbf{s}_f \mathrm{E}_{jf} \mathbf{s}_f^H + M\omega^2 \mathbf{I} \Big)^{-1} \mathbf{s}_l \mathrm{E}_{jl}. \tag{3.19}$$

3.2.3 Maximum A Posteriori Estimation Method

The focus of this section is Maximum A Posteriori (MAP) estimation method which is popular in many applications [50]. The estimation process of MAP estimation is explained in the following context of channel estimation.

The pilot matrix in equation (3.3) can be written as

$$\mathbf{S} \triangleq \Big[\mathbf{s_1} \otimes \mathbf{I_M} \cdots \mathbf{s_L} \otimes \mathbf{I_M}\Big], \tag{3.20}$$

where $\otimes$ is Kronecker product, and $\mathbf{I_M}$ is the $M \times M$ identity matrix. Thus

$$\mathbf{S} = \begin{pmatrix} s_{11}\mathbf{I_M} & s_{21}\mathbf{I_M} & \dots & s_{L1}\mathbf{I_M} \\ s_{12}\mathbf{I_M} & s_{22}\mathbf{I_M} & \dots & s_{L2}\mathbf{I_M} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1\tau}\mathbf{I_M} & s_{2\tau}\mathbf{I_M} & \dots & s_{L\tau}\mathbf{I_M} \end{pmatrix}, \tag{3.21}$$

where $\mathbf{S}$ is a $\tau M \times LM$ matrix.

Then the received signal in equation (3.3) can be written as

$$\mathbf{Y}_j = \mathbf{S}\mathbf{H}_j + \mathbf{N}_j, \tag{3.22}$$

where $\mathbf{H}_j = \mathbf{vec}(\mathbf{h}_j)$ is a $LM \times 1$ vector by stacking all L channels into a vector, and $\mathbf{h}_j = [\mathbf{h}_{j1} \quad \mathbf{h}_{j2} \dots \mathbf{h}_{jL}]$ is a $M \times L$ matrix. $\mathbf{Y}_j = \mathbf{vec}(\mathbf{y}_j)$ is a $\tau M \times 1$ vector. $\mathbf{N}_j = \mathbf{vec}(\mathbf{n}_j)$ is also a vector with the same size as $\mathbf{Y}_j$.

The principal rule of MAP estimation is to obtain the estimation value by maximizing the posteriori probability, which is

$$\begin{aligned} \hat{\mathbf{H}}_j &= \arg\max_{\hat{\mathbf{H}}_j} \quad p\Big(\mathbf{H}_j|\mathbf{Y}_j\Big) \\ &= \arg\max_{\hat{\mathbf{H}}_j} \quad \frac{p\Big(\mathbf{H}_j\Big)p\Big(\mathbf{Y}_j|\mathbf{H}_j\Big)}{p\Big(\mathbf{Y}_j\Big)}, \end{aligned} \tag{3.23}$$

where $\mathbf{H}_j = [\mathbf{h}_{j1}^T \mathbf{h}_{j2}^T \cdots \mathbf{h}_{jL}^T]^T$. The $\mathbf{h}_{jl}$ are multivariate complex Gaussian vectors with zero-mean and covariance matrix $\mathbf{R}_{jl} = \mathrm{E}[\mathbf{h}_{jl}\mathbf{h}_{jl}^H]$. Therefore, the probability density function (PDF) of $\mathbf{h}_{jl}$ is

$$p\Big(\mathbf{h}_{jl}\Big) = \frac{\exp\Big(-\mathbf{h}_{jl}^H \mathbf{R}_{jl}^{-1}\mathbf{h}_{jl}\Big)}{\pi^M \det\mathbf{R}_{jl}}. \tag{3.24}$$

Moreover, the $\mathbf{h}_{jl}$ vector are independent, so $\mathbf{H}_j$ is itself multivariate complex Gaussian vector. Therefore, the joint PDF of $\mathbf{H}_j$ is given as follows

$$\begin{aligned} p\Big(\mathbf{H}_j\Big) &= \frac{\exp\Big(-\sum_{l=1}^{L}\mathbf{h}_{jl}^H\mathbf{R}_{jl}^{-1}\mathbf{h}_{jl}\Big)}{\pi^{LM}\Big(\det\mathbf{R}_{j1}\cdots\det\mathbf{R}_{jL}\Big)} \\ &= \frac{\exp\Big(\mathbf{h}_j^H\mathbf{R}_j^{-1}\mathbf{h}_j\Big)}{\pi^{LM}\Big(\det\mathbf{R}_{jl}\cdots\det\mathbf{R}_{jL}\Big)}, \end{aligned} \tag{3.25}$$

where $\mathbf{R}_j$ is given by

$$\begin{aligned} \mathbf{R}_j &\triangleq \operatorname{diag}\Big(\mathbf{R}_1,\cdots,\mathbf{R}_L\Big) \\ &= \begin{pmatrix} \mathbf{R}_{j1} & 0 & \dots & 0 \\ 0 & \mathbf{R}_{j2} & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \mathbf{R}_{jL} \end{pmatrix}, \end{aligned} \tag{3.26}$$

in which $\mathbf{R}_{jl} = \mathrm{E}[\mathbf{h}_{jl}\mathbf{h}_{jl}^H]$ is the covariance matrix of $\mathbf{h}_{jl}$.

Combine equation (3.26) with equation (3.22), then $p(\mathbf{Y}_j|\mathbf{H}_j)$ in equation (3.23) is written as

$$p\Big(\mathbf{Y}_j|\mathbf{H}_j\Big) = \frac{\exp\Big(-\big(\mathbf{Y}_j-\mathbf{S}\mathbf{H}_j\big)^H\big(\mathbf{Y}_j-\mathbf{S}\mathbf{H}_j\big)/\omega^2\Big)}{(\pi\omega^2)^{\tau M}}, \tag{3.27}$$

where the ω is the variance of noise.

Since $p(\mathbf{Y}_j)$ is a constant, it does not affect the argmax in equation (3.23). Therefore, equation (3.23) can written in another way, which is based on equation (3.25) and equation (3.27), and for the reason of simplicity, we present the denominator in a simplified expression

$$\begin{aligned} p\Big(\mathbf{H}_j|\mathbf{Y_j}\Big) &= \frac{\exp\Big(\mathbf{h}_j^H\mathbf{R}_j^{-1}\mathbf{h}_j\Big)\exp\Big(-\big(\mathbf{Y}_j-\mathbf{S}\mathbf{H}_j\big)^H\big(\mathbf{Y}_j-\mathbf{S}\mathbf{H}_j\big)/\omega^2\Big)}{\pi^{LM}(\pi\omega^2)^{\tau M}\big(\det\mathbf{R}_{jl}\cdots\det\mathbf{R}_{jL}\big)} \\ &= \frac{\exp\Big(\mathbf{h}_j^H\mathbf{R}_j^{-1}\mathbf{h}_j\Big)\exp\Big(-\big(\mathbf{Y}_j-\mathbf{S}\mathbf{H}_j\big)^H\big(\mathbf{Y}_j-\mathbf{S}\mathbf{H}_j\big)/\omega^2\Big)}{\mathbf{\Psi}} \\ &= \frac{\exp\Big(\mathbf{h}_j^H\mathbf{R}_j^{-1}\mathbf{h}_j-\big(\mathbf{Y}_j-\mathbf{S}\mathbf{H}_j\big)^H\big(\mathbf{Y}_j-\mathbf{S}\mathbf{H}_j\big)/\omega^2\Big)}{\mathbf{\Psi}} \\ &= \frac{\mathbf{\Upsilon}(\mathbf{H})}{\mathbf{\Psi}}, \end{aligned} \tag{3.28}$$

$$\boldsymbol{\Upsilon}(\mathbf{H}) = \exp\Big(\mathbf{h}_j^H \mathbf{R}_j^{-1} \mathbf{h}_j - \big(\mathbf{Y}_j - \mathbf{S}\mathbf{H}_j\big)^H \big(\mathbf{Y}_j - \mathbf{S}\mathbf{H}_j\big)/\omega^2\Big), \tag{3.29}$$

$$\boldsymbol{\Psi} = \pi^{LM}(\pi\omega^2)^{\tau M}\big(\det\mathbf{R}_{jl} \cdots \det\mathbf{R}_{jL}\big). \tag{3.30}$$

From equation (3.28), the MAP estimator of $\mathbf{H}_j$ is obtained by setting the partial derivative of the numerator

$$\begin{aligned} \frac{\partial\boldsymbol{\Upsilon}(\mathbf{H})}{\partial\mathbf{H}_j^H} &= \frac{\partial\Big(\exp\Big(\mathbf{h}_j^H \mathbf{R}_j^{-1} \mathbf{h}_j - \big(\mathbf{Y}_j - \mathbf{S}\mathbf{H}_j\big)^H \big(\mathbf{Y}_j - \mathbf{S}\mathbf{H}_j\big)/\omega^2\Big)\Big)}{\partial\mathbf{H}_j^H} \\ &= \mathbf{R}_j^{-1}\mathbf{H}_j - \frac{\mathbf{S}^H\mathbf{Y}_j + \mathbf{S}^H\mathbf{S}\mathbf{H}_j}{\omega^2}, \end{aligned} \tag{3.31}$$

to zero:

$$\mathbf{R}_j^{-1}\mathbf{H}_j - \frac{\mathbf{S}^H\mathbf{Y}_j + \mathbf{S}^H\mathbf{S}\mathbf{H}_j}{\omega^2} = 0. \tag{3.32}$$

Then it is very easy to verify that

$$\mathbf{H}_j^{MAP} = \Big(\mathbf{R}_j\mathbf{S}^H\mathbf{S} + \omega^2\mathbf{I}_{LM}\Big)^{-1}\mathbf{R}_j\mathbf{S}^H\mathbf{Y}_j. \tag{3.33}$$

It should be noted that, the MMSE estimate can also be expressed in this way, which is

$$\mathbf{H}_j^{MMSE} = \mathbf{R}_j\mathbf{S}^H\Big(\mathbf{S}\mathbf{R}_j\mathbf{S}^H + \omega^2\mathbf{I}_{\tau M}\Big)^{-1}\mathbf{Y}_j. \tag{3.34}$$

Since $\mathbf{H}_j$ is jointly Gaussian distributed, the MAP estimation method is equal to the MMSE estimation method. Moreover, due to the matrix inversion identity $\mathbf{F}(\mathbf{I}+\mathbf{P}\mathbf{F})^{-1} = (\mathbf{F}\mathbf{P}+\mathbf{I})^{-1}\mathbf{F}$ (the Woodbury Identity), equation (3.33) is same as equation (3.34).

For a multipath system model, the channel $\mathbf{h}_{jl}$ be given by

$$\mathbf{h}_{jl} = \frac{1}{P}\sum_{p=1}^{P}\zeta_{jlp}\mathbf{a}(\varphi_{jlp}), \tag{3.35}$$

where P is the number of i.i.d paths. The $\zeta_{jlp} \sim \mathcal{CN}(0, \sigma_{jlp}^2)$ is the complex gain factor related to distance and SNR between users to the jth BS. $\mathbf{a}(\varphi_{jlp})$ is the $M \times 1$ signature

vector of ULA based on φ_{jlp} which is the angle of arrival (AOA) [51]. The covariance matrix is

$$\begin{aligned} \mathbf{R}_{jl} &= \mathrm{E}[\mathbf{h}_{jl}\mathbf{h}_{jl}^H] \\ &= \frac{\sigma_{jlp}^2}{P}\sum_{p=1}^{P}\mathrm{E}[\mathbf{a}(\varphi_{jlp})\mathbf{a}(\varphi_{jlp})^H], \end{aligned} \tag{3.36}$$

where $\mathbf{R}_{jl}$ in the MAP estimation method is a function of AOAs of different users (one user in each cell).

Channel covariance matrices include substantial information such as the AOA of channels. It has been proven that if there is no overlap between AOAs for desired and interference channels, then the covariance matrices differ significantly across users. Therefore, the MAP estimation method can obtain good estimates of the CSI based on the unique information in covariance matrices under pilot contamination [11]. It can be seen from Fig.3.2 that there is no overlap between user A and users B, then the BS can separate the signals of two users based on the different information lying in their covariance matrices. However, for user A and user C, their AOAs have overlap, so the information lying in their covariance matrices are very similar. The BS receives the combination of these two users' signals but can not separate them due to the very similar information lying in their covariance matrices. This observation is the basis for the pilot assignment technique proposed in [11].

MAP estimation methods also require the high quality estimated covariance matrix of each channel to estimate the CSI in the jth cell. In many papers that use the MAP estimation method, the covariance matrices are assumed a priori knowledge at the BSs. In Chapter 4 of this thesis we remove this assumption.

3.3 The Effect of Pilot Contamination

In this section, we discuss the effect of pilot contamination on the capacity of uplink data transmission. We note that it will also affect downlink precoding as well, but in this thesis we focus on uplink data transmission.

For uplink data transmission of Massive MIMO systems, the number of users in

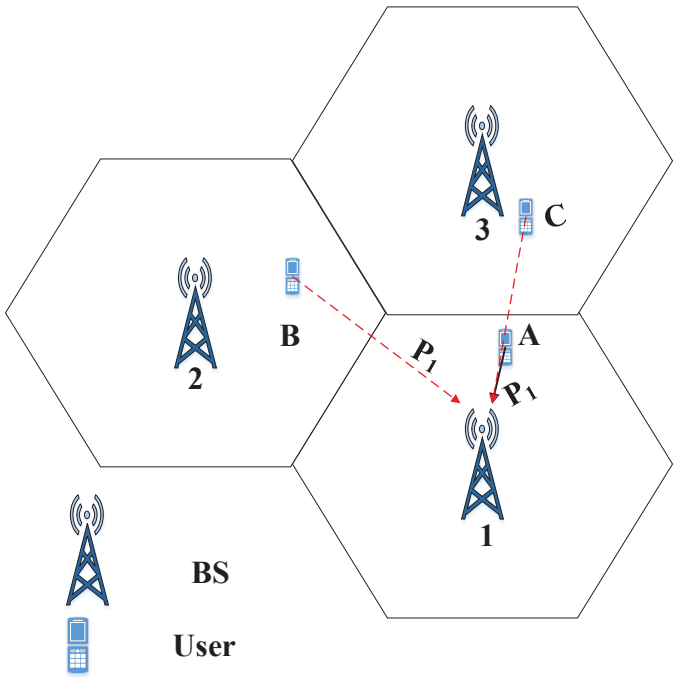

Figure 3.2: The AOA distribution of users in Massive MIMO system.

each cell is also assumed to be 1, which is the same with uplink training. Users in the fth cell send uplink data signals to the BS in their own cells, which can can be denoted as q_f. q_f is a scalar and $\mathrm{E}[q_f q_f^H] = \delta_q^2$. Therefore, the BS in jth cell receives $M \times 1$ data vector

$$\begin{aligned} \boldsymbol{x}_j &= \sum_{f=1}^{L} \mathbf{h}_{jf} q_f^T + \mathbf{w}_j \\ &= \mathbf{h}_{jj} q_j^T + \sum_{f \neq j}^{L} \mathbf{h}_{jf} q_f^T + \mathbf{w}_j, \end{aligned} \tag{3.37}$$

where the $\mathbf{w}_j$ is noise, which is a $M \times 1$ vector of i.i.d. zero-mean, circularly-symmetric complex Gaussian random variables. The BS processes this received signal by multiplying the conjugate transpose of channel estimate. The process is

$$\begin{aligned} \boldsymbol{\Upsilon}_j &= \hat{\mathbf{h}}_{jj}^H \mathbf{x}_j \\ &= \hat{\mathbf{h}}_{jj}^H (\mathbf{h}_{jj} q_j^T + \sum_{f \neq j}^{L} \mathbf{h}_{jf} q_f^T + \mathbf{w}_j), \end{aligned} \tag{3.38}$$

where $\hat{\mathbf{h}}_{jj}^H$ is the conjugate transpose of $\hat{\mathbf{h}}_{jj}$, and $\hat{\mathbf{h}}_{jj}$ can be obtained from equation (3.13). For the sake of simplicity, we choose LS estimation method, and similar conclusion can be obtained using the MMSE or MAP method. This is called conjugate

beamforming in [3]. Thus equation (3.38) can be written as

$$\boldsymbol{\Upsilon}_j = \big(\mathbf{h}_{jj} + \sum_{f\neq j}^{L} \mathbf{h}_{jf} + \mathbf{n}_j\mathbf{s}/\tau_s\big)^H \big(\mathbf{h}_{jj}q_j^T + \sum_{f\neq j}^{L} \mathbf{h}_{jf}\mathbf{q}_f^T + \mathbf{w}_j\big). \tag{3.39}$$

From equation (3.39), the power of desired signal is

$$\mathbf{P}_j^{sig} = \|(\mathbf{h}_{jj} + \sum_{f\neq j}^{L} \mathbf{h}_{jf} + \mathbf{n}_j\mathbf{s}/\tau_s)^H(\mathbf{h}_{jj}q_j^T)\|_F^2, \tag{3.40}$$

where the $\|\cdot\|_F$ denotes the Frobenius norm. The next step following equation (3.40) is

$$\begin{aligned}\mathbf{P}_j^{sig} &= \mathrm{E}\Big\{ \Big| \mathbf{h}_{jj}^H\mathbf{h}_{jj}q_j + \sum_{f\neq j}^{L} \mathbf{h}_{jf}^H\mathbf{h}_{jj}q_j + (\tau_s)^{-1}\mathbf{s}^T\mathbf{n}_j^H\mathbf{h}_{jj}q_j \Big|^2 \Big\} \\ &= \mathrm{E}\Big\{ \Big| E_{jf}q_j + \sum_{f\neq j}^{L} \boldsymbol{h}_{jf}^H\mathbf{h}_{jj}q_j + (\tau_s)^{-1}\mathbf{s}^T\mathbf{n}_j^H\mathbf{h}_{jj}q_j \Big|^2 \Big\},\end{aligned} \tag{3.41}$$

where $E_{jj} = \mathrm{E}[\mathbf{h}_{jj}^H\mathbf{h}_{jj}]$. Since the number of antennas at each user is just one, the E_{jf} is a scalar. According to equation (3.5), E_{jj} is written as

$$\begin{aligned}E_{jj} &= \mathrm{E}[\boldsymbol{h}_{jj}^H\boldsymbol{h}_{jj}] \\ &= M\beta_{jj}.\end{aligned} \tag{3.42}$$

Following the similar way, the second part in equation (3.41) is calculated as

$$\mathrm{E}\Big\{ \big| \sum_{f\neq j}^{L} \mathbf{h}_{jf}^H\mathbf{h}_{jj}q_j \big|^2 \Big\} = M\beta_{jj}\beta_{jf}. \tag{3.43}$$

The last part in equation (3.41) is

$$\begin{aligned}\mathrm{E}\ &\Big\{ \big| (\tau)^{-1}\mathbf{s}^T\mathbf{n}_j^H\mathbf{h}_{jj}q_j \big|^2 \Big\} \\ &= \mathrm{E}\Big\{ (\tau)^{-1}\delta_q^2\mathbf{s}^T\mathbf{n}_j^H\mathbf{R}_{jj}\mathbf{n}_j \Big\} \\ &= \frac{\delta_n^2\delta_q^2}{\tau} M\beta_{jj},\end{aligned} \tag{3.44}$$

where $\mathbf{R}_{jj}$ in equation (3.44) is the channel covariance matrix. δ_q^2 is the power of uplink data signal, δ_n^2 is noise power. According to equation (3.42), equation (3.43)

and equation (3.44), the power of desired user is

$$\begin{aligned}\mathbf{P}_j^{sig} &= \mathrm{E}\Big\{ \big| \boldsymbol{E}_{jf} q_j \big|^2 \Big\} + \mathrm{E}\Big\{ \big| \sum_{f\neq j}^{L} \mathbf{h}_{jf}^H \mathbf{h}_{jj} q_{ju} \big|^2 \Big\} + \mathrm{E}\Big\{ \big| (\tau)^{-1} \mathbf{s}_l^T \mathbf{n}_j^H \mathbf{h}_{jj} q_j \big|^2 \Big\} \\ &= M^2 \beta_{jj}^2 \delta_q^2 + M \delta_q^2 \beta_{jj} \sum_{f\neq j}^{L} \beta_{jf} + \frac{\delta_n^2 \delta_q^2}{\tau} M \beta_{jj}. \end{aligned} \tag{3.45}$$

Since the β_{jf} is geometric attenuation and shadowing of interference cells, it is much smaller than β_{jj}. Besides, the number of antennas at each BS in Massive MIMO systems is large, thus $M^2 \gg M$. Follow this way, desired signal power can be approximated by

$$\mathbf{P}_j^{sig} \approx M^2 \beta_{jj}^2 \delta_q^2. \tag{3.46}$$

The power of interference signals and noise can be obtained by following the familiar way, respectively. The power of interference is given by

$$\begin{aligned}\mathbf{P}_j^{inter} &= \mathrm{E}\Big\{ \Big\| \big(\mathbf{h}_{jj} + \sum_{f\neq j}^{L} \mathbf{h}_{jf} + \mathbf{n}_j \mathbf{s}/\tau_s\big)^H \big(\sum_{f\neq j}^{L} \mathbf{h}_{jf} \mathbf{q}_f \big) \Big\|_F^2 \Big\} \\ &\approx M^2 \delta_q^2 \sum_{f\neq j}^{L} \beta_{jf}^2. \end{aligned} \tag{3.47}$$

The noise power is

$$\begin{aligned}\mathbf{P}_j^{n} &= \mathrm{E}\Big\{ \Big\| \big(\mathbf{h}_{jj} + \sum_{f\neq j}^{L} \mathbf{h}_{jf} + \mathbf{n}_j s/\tau_s\big)^H \mathbf{n}_j \Big\|_F^2 \Big\} \\ &= \delta_n^2 \mathrm{E}\Big\{ \big(\mathbf{h}_{jj} + \sum_{f\neq j}^{L} \mathbf{h}_{jf} + \mathbf{n}_j s/\tau_s\big)^H \big(\mathbf{h}_{jj} + \sum_{f\neq j}^{L} \mathbf{h}_{jf} + \mathbf{n}_j s/\tau_s\big) \Big\} \\ &= \delta_n^2 \big(M\beta_{jj} + M \sum_{f\neq j}^{L} \beta_{jf} + \delta_n^2 \big). \end{aligned} \tag{3.48}$$

Therefore, the signal to interference plus noise ratio(SINR) at the jth BS is able to achieved based on equation (3.46), equation (3.47) and equation (3.48)

$$\mathrm{SINR}_j \approx \frac{M^2 \beta_{jj}^2 \delta_q^2}{M^2 \delta_q^2 \sum\limits_{f\neq j}^{L} \beta_{jf}^2 + \delta_n^2 \big(M\beta_{jj} + M \sum\limits_{f\neq j}^{L} \beta_{jf} + \delta_n^2 \big)}. \tag{3.49}$$

If there is no pilot contamination in the system, then equation (3.49) can be written as

$$\mathrm{SINR}_j \approx \frac{M^2 \beta_{jj}^2 \delta_q^2}{\delta_n^2 \big(M\beta_{jj} + M \sum\limits_{f\neq j}^{L} \beta_{jf} + \delta_n^2 \big)}. \tag{3.50}$$

When M goes to infinity, the ratio in equation (3.50) is also tends to infinity. Therefore, the SINR_j goes to infinity with the increase of M.

When pilot contamination exists, if the number of antennas at BS goes to infinity, then the effect of noise can be ignored. Under this situation, the SINR in equation (3.49) reaches a limit

$$\lim_{M\to\infty} \mathrm{SINR}_j = \frac{\beta_{jj}^2}{\sum\limits_{f\neq j}^{L} \beta_{jf}^2}. \tag{3.51}$$

The denominator in equation (3.51) is the interference from adjacent cells which is caused by pilot contamination.

It can be seen that this interference due to pilot contamination will not decrease, even though the number of antennas at each BS increases without bound. In other words, the SINR is bounded due to pilot contamination. Moreover, the capacity of uplink system which denoted as C is the function of SINR, which is given as

$$\begin{aligned} \lim_{M\to\infty} \mathrm{C}_j &= \log_2\left(1+\mathrm{SINR}_j\right) \\ &= \log_2\left(1+\frac{\beta_{jj}^2}{\sum\limits_{f\neq j}^{L} \beta_{jf}^2}\right). \end{aligned} \tag{3.52}$$

Based on equation (3.51) and equation (3.52), we can get the conclusion that pilot contamination is the main restriction on the increase of SINR and achievable capacity of a Massive MIMO system. Therefore, pilot contamination is a serious challenge for Massive MIMO systems.

3.4 Simulation Results

In this section, we mainly discuss the performance of three estimation methods mentioned in this chapter. Since $\mathbf{H}_j$ is joint Gaussian distribution, the MMSE estimation method is equal to MAP method. Thus MAP estimation method and LS estimation method are considered in this section. The model of the system includes hexagonally shaped cells, and all users are distributed at the circle with 600 meters to their own BSs. In this section, we assume each cell there is a user. Each BS has a linear array

Table 3.1: Basic Simulation Parameters of Chaper 3

Number of paths (P)	60
Number of cells (L)	2
Carrier frequency	2GHz
Gain factor exponent (γ)	2.8
Pilot length (τ)	16
Antenna space (D)	$\mu/2$

with 60 antennas. The cell radius is 800 meters and the signal-to-noise ratio (SNR) at the user is 18 dB. Moreover, there is no overlap between AOAs for interfering and desired channels. Other basic parameters are shown in table 3.1.

The normalized mean square error (NMSE) of CSI and capacity per cell are considered as standards to test the performance of channel estimation methods. The capacity is given by

$$\mathrm{C} \triangleq \frac{\log_2\big(1+\sum_{j=1}^{L}\mathrm{SINR}_j\big)}{L}. \tag{3.53}$$

The NMSE is calculated as follows

$$\mathrm{err}_{\mathbf{h}} \triangleq 10\log_{10}\left[\frac{\sum_{j=1}^{L}\left\|\hat{\mathbf{h}}_{jj}-\mathbf{h}_{jj}\right\|_F^2}{\sum_{j=1}^{L}\left\|\mathbf{h}_{jj}\right\|_F^2}\right], \tag{3.54}$$

The desired cell in Fig. 3.3 is $j = 1$. It can be seen that, under pilot contamination, MAP method can get lower NMSE than LS method. Moreover, with increase of the number of antennas at the BS, the NMSE of MAP method becomes lower and lower, but there is no decrease for NMSE of LS method. Therefore, MAP estimation method enables to obtain much better estimate than LS method suffering pilot contamination.

Fig.3.4 describes the capacity per cell. The parameters in this figure are the same as Fig.3.3. The capacity of MAP method grows much more quickly than the capacity of LS method as the number of antennas increases. Therefore MAP estimation method can obtain much more capacity than LS method under pilot contamination.

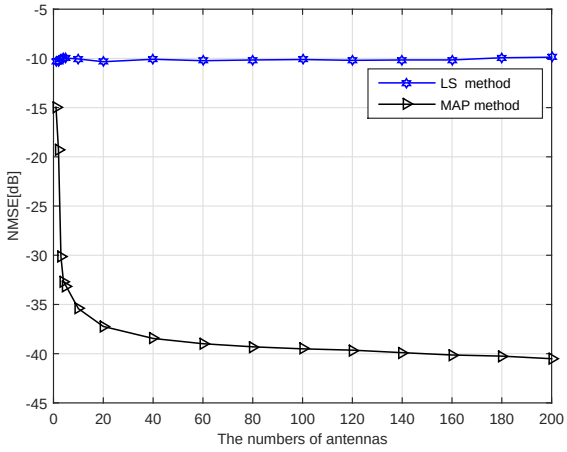

Figure 3.3: The NMSE of the CSI, 2-cell network.

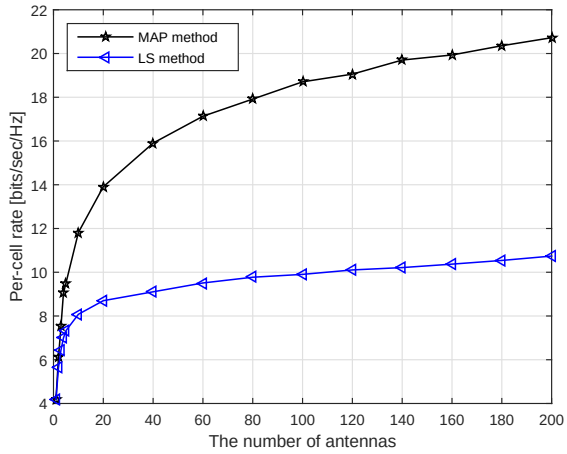

Figure 3.4: The capacity of per cell, 2-cell network.

3.5 Conclusion

In this chapter, the system model of Massive MIMO is firstly introduced, then several channel estimation methods, such as LS mehod, MMSE method and MAP estimation method are discussed. The MMSE estimation method and MAP estimation method are examples of covariance-added estimation methods, which are able to achieve much better estimation results than the LS estimation method. However, these covariance-added estimation methods need to estimate every channel covariance matrix separately, which is a difficult and complex work in Massive MIMO due to pilot contamination.

Chapter 4

Novel Estimation Methods for New Scenarios

4.1 Introduction

Based on Chapter 3 we can draw a conclusion that, under pilot contamination, covariance-aided estimation methods (such as the MMSE estimation method) is able to achieve much better estimation results than the LS estimation method. However, the MMSE estimation method requires the estimate of the covariance matrix for each channel. Current research which uses the covariance matrix to estimate CSI assumes that the covariance matrices can be estimated accurately prior to being used in channel estimation.

Some papers assume that covariance matrix of each cell can be estimated by getting the users in different cells to transmit pilot sequences one at a time [11]. The pilot sequences in the same cell are mutually orthogonal. In this way, every time only users in one cell transmit pilot sequences, so that there is no pilot contamination and covariance matrices of each cell can be estimated separately. After estimating all covariance matrices, then all users in the system begin to transmit signals simultaneously. However, estimating covariance matrices in this way has some limitations. For example, this estimation method can only fit the situation in which no new users come into cell and locations of users are fixed. If a new user enters the cell with unknown covariance

matrix or covariance matrices are initially known but change over time due to mobility of users, then the method which makes cells switch on and off separately to estimate the new covariance matrices is no longer desirable, because it requires too much time. During this estimation period, each time only users in one cell are transmitting pilots, which increase the overhead and reduces the performance of system. It would be better if all users can continue to transmit pilots and data whilst the BS tracks the changing covariance matrices. This chapter investigates how the tracking process may be implemented and performance results are obtained.

4.2 System Model for New Scenarios

As mentioned in Section 3.1, the system model in this chapter is also a network of L time-synchronized cells, and the channel is block-fading. Each block consists of uplink transmissions (including uplink training and uplink data transmission) and downlink transmission. We focus on uplink transmission in this chapter. Moreover, the covariance matrices of channels in this chapter are constant until the arrival of the new user or the mobility of users. In this chapter, in the same cell pilot sequences are mutually orthogonal, but in different cells they are the same group. Since we will be focusing on the performance of users that share the same pilot sequence, the number of users per cell in this chapter is $U = 1$. The number of cells in this chapter is $L = 4$ or $L = 7$. The other parameters are the same as Section 3.1. The system model of this chapter (4 cells) is shown in Fig. 4.1. From this figure, we can see that in each cell there is one user with single antenna, and each BS have a antenna array with M antennas. The user in the cell can move, or a new user comes into one cell (no user in the cell moves). The BS in the jth cell receives signals from all users

$$\mathbf{y}_j(n) = \sum_{l=1}^{L} \mathbf{h}_{jl}(n)\mathbf{s}_l^T(n) + \mathbf{n}_j(n), \tag{4.1}$$

where $n = 1, 2, \cdots, N$ is the number of blocks. The $\mathbf{h}_{jl}(n)$ is a $M \times 1$ vector which is a constant in the nth block but changes during different blocks (the so called block fading model [49]). $\mathbf{y}_j(n)$ and $\mathbf{n}_j(n)$ are$M \times \tau$ matrices. $\mathbf{y}_j(n)$ and $\mathbf{n}_j(n)$ are variables

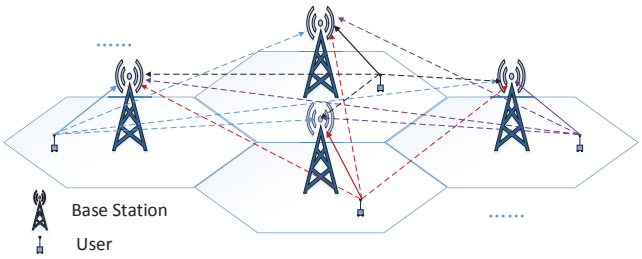

Figure 4.1: The system model for new scenarios.

in the nth block. The $\mathbf{s}(n)$ denotes the pilot sequences in the nth block, since in every block the pilot sequences are the same, we describe the pilot sequences in the nth block as $\mathbf{s}$ instead of $\mathbf{s}(n)$. We choose a multipath model, the channel be written as

$$\mathbf{h}_{jl}(n) = \frac{1}{P}\sum_{p=1}^{P}\zeta_{jlp}(n)\mathbf{a}(\varphi_{jlp}(n)), \tag{4.2}$$

where P is the number of i.i.d paths. The $\zeta(n)_{jlp}$ is the complex gain factor in the nth block, which is related to distance and SNR between the user and the jth BS. $\mathbf{a}(\varphi_{jlp}(n))$ is the signature vector of ULA based on $\varphi_{jlp}(n)$ which is the angle of arrival (AOA) of pth path of the channel between the lth user and the jth BS [51]. The $\mathbf{a}(\varphi_{jlp}(n))$ is

$$\mathbf{a}(\varphi_{jlp}(n)) \triangleq \begin{pmatrix} 1 \\ e^{-i2\pi\frac{D}{\mu}cos(\varphi_{jlp}(n))} \\ \vdots \\ e^{-i2\pi\frac{(M-1)D}{\mu}cos(\varphi_{jlp}(n))} \end{pmatrix}, \tag{4.3}$$

in which D is the antenna space and μ is the wavelength of the signal. In this chapter, we consider $D = \mu/2$. Moreover, $\varphi_{jlp}(n)$ is the random AOA, such that $\varphi_{jlp}(n) \in [-\pi, \pi]$, Since $\varphi_{jlp}(n) \in [-\pi, 0]$ and $\varphi_{jlp}(n) \in [0, \pi]$ can get the same result for $\cos(\varphi_{jlp}(n))$, we just consider the case that the $\varphi_{jlp}(n)$ is between 0 and π.

Moreover, $\zeta_{jlp}(n)$ in equation (4.2) is modeled by

$$\zeta_{jlp}(n) = \sqrt{\beta_{jl}(n)}e^{i\phi_{jlp}}, \tag{4.4}$$

where the ϕ_{jlp} is the i.i.d. random phase, and it is independent across different j, l and p. The $\sqrt{\beta_{jl}}$ is large-scale signal attenuation due to distance and shadowing and we denote $\beta_{jl} = \alpha/d_{jl}^{\gamma}$ as in Chapter 3.

4.3 Novel Estimation Methods for New Scenarios

In this section, we will firstly research a situation in which there is a new user who comes into a cell with unknown covariance matrix, and we then study another scenario: all covariance matrices change due to mobility of users. Covariance matrices in every block keep the same until users change their locations or a new user comes into the cell.

4.3.1 Novel Estimation Method for the First Scenario

We now describe the first scenario: a new user enters one cell has a fixed but unknown covariance matrix. Moreover, in this scenario, the scheduled number of served users for each BS at the same time is fixed. Therefore, the initial user in that cell drops out its service when the new user is scheduled. We suppose the new entrant uses the same pilot sequence as the initial one, but its location in the cell is dramatically different from the initial one. Besides, after the new user comes into the cell, its covariance matrix will keep the same over time. The users in other cells also have fixed covariance matrices and we assume that they are known from initial training [11]. Therefore, we only need to estimate the new covariance matrix related to the channels of the new user.

The key idea that we explain is that the sum of all the covariance matrices for cell j can be estimated using the received signal $\mathbf{y}_j$ directly.

The covariance matrices of received signal is shown as

$$\begin{aligned}
\mathbf{R}_j^y &= \mathrm{E}\big[\mathbf{y}_j(n)\mathbf{y}_j^H(n)\big] \\
&= \mathrm{E}\Big[\Big(\sum_{l=1}^{L}\mathbf{h}_{jl}(n)\mathbf{s}_l^T + \mathbf{n}_j(n)\Big)\Big(\sum_{l=1}^{L}\mathbf{h}_{jl}(n)\mathbf{s}_l^T + \mathbf{n}_j(n)\Big)^H\Big] \\
&= \mathrm{E}\Big[\sum_{l=1}^{L}\mathbf{h}_{jl}(n)\mathbf{s}_l^T\mathbf{s}_l{\mathbf{h}_{jl}}^H(n) + \sum_{l=1}^{L}\mathbf{h}_{jl}(n)\mathbf{s}_l^T\mathbf{n}_j^H(n) \\
&+ \sum_{l=1}^{L}\sum_{b\neq l}^{L}\mathbf{h}_{jl}(n)\mathbf{s}_l^T\mathbf{s}_b{\mathbf{h}_{jb}}^H(n) + \mathbf{n}_j(n)\mathbf{n}_j^H(n)\Big]. \qquad (4.5)
\end{aligned}$$

The first part and the last part in equation (4.5) are

$$\begin{aligned} \mathrm{E}\Big[\sum_{l=1}^{L}\mathbf{h}_{jl}(n)\mathbf{s}_l^T\mathbf{s}_l{\mathbf{h}_{jl}}^H(n)\Big] &= \tau\mathrm{E}\Big[\sum_{l=1}^{L}\mathbf{h}_{jl}(n){\mathbf{h}_{jl}}^H(n)\Big] \\ &= \tau\sum_{l=1}^{L}\mathbf{R}_{jl}, \end{aligned} \tag{4.6}$$

$$\mathrm{E}\Big[\mathbf{n}_j(n)\mathbf{n}_j^H(n)\Big] = \sum_{l=1}^{L}\mathbf{R}_{jl}^{noise}, \tag{4.7}$$

where $\mathbf{R}_{jl}$ and $\mathbf{R}_{jl}^{noise}$ are the channel covariance matrix and the noise covariance matrix, respectively.

The $\mathbf{h}_{jl}(n)$ and $\mathbf{n}_j(n)$ are independent and $\mathbf{n}_j(n)$ is zero-mean random vector, so the second part in equation (4.5) is equal to zero. Following a similar methodology, the third part in equation (4.5) is also zero. Thus, the final expression of equation (4.5) is

$$\mathbf{R}_j^y = \tau\sum_{l=1}^{L}\mathbf{R}_{jl} + \mathbf{R}_j^{noise}, \tag{4.8}$$

in which the $\sum_{l=1}^{L}\mathbf{R}_{jl}$ is a sum of channel covariance matrices. The covariance matrix of the noise $\mathbf{R}_j^{noise}$ is $\omega^2\mathbf{I}_M$. Therefore, we can obtain the desired channel covariance matrix (assume the new entrant comes into vth cell) by equation (4.8), which is

$$\begin{aligned} \mathbf{R}_{jv} &= \frac{1}{\tau}\Big[\mathbf{R}_j^y - \mathbf{R}_j^{noise}\Big] - \sum_{l\neq v}^{L}\mathbf{R}_{jl} \\ &= \frac{1}{\tau}\Big[\mathbf{R}_j^y - \omega^2\mathbf{I}_M\Big] - \sum_{l\neq v}^{L}\mathbf{R}_{jl}. \end{aligned} \tag{4.9}$$

The $\mathbf{R}_{jl}$ in equation (4.9) are assumed to be estimated correctly before the new user's entry. Based on the discussion above, only $\mathbf{R}_{jv}$ changes and $\sum_{l\neq v}^{L}\mathbf{R}_{jl}$ remains the same. The only unknown part of equation (4.9) is $\mathbf{R}_j^y$ which can be estimated directly from the received signal by:

$$\hat{\mathbf{R}}_j^y(n) = \frac{1}{n}\big[\sum_{b=1}^{n}\mathbf{y}_j(b)\mathbf{y}_j(b)^H\big]. \tag{4.10}$$

Therefore, the estimation of $\mathbf{R}_{jv}$ can be written as

$$\hat{\mathbf{R}}_{jv}(n) = \frac{1}{\tau}\Big[\hat{\mathbf{R}}_j^y(n) - \omega^2\mathbf{I}_M\Big] - \sum_{l\neq v}^{L}\mathbf{R}_{jl}. \tag{4.11}$$

The MAP estimation equation based on estimated covariance matrices is (refer to 3.33)

$$\hat{\mathbf{h}}_{jv}(n) = \hat{\mathbf{R}}_{jv}(n)\Big[\sum_{l=1}^{L}\hat{\mathbf{R}}_{jl}(n) + \omega^2\mathbf{I}_M\Big]^{-1}\widetilde{\mathbf{S}}^H\mathbf{Y}_j(n). \tag{4.12}$$

Our assumption in this section is that the covariance matrix of the new user does not change over time. It is feasible to estimate the new covariance matrix by taking n in equation (4.10) large enough.

The details of the proposed algorithm are summarized as follows:

Step 1) Before new entrant comes into the vth cell, the jth BS achieves the correct estimation value of original $\hat{\mathbf{h}}_{jl}$ and $\hat{\mathbf{R}}_{jl}$ through training;

Step 2) The new user enters the vth cell, then the jth BS collects $\mathbf{y}_j(n)$ and calculates $\hat{\mathbf{R}}_j^y(n)$ based on the equation (4.10);

Step 3) Compute the nth block estimated $\hat{\mathbf{R}}_{jv}(n)$ based on the equation (4.11);

Step 4) Estimate the $\mathbf{h}_{jl}(n)$ by equation (4.12) based on $\hat{\mathbf{R}}_j(n)$ in Step 3).

4.3.2 Novel Estimation Method for the Second Scenario

In this section, we study a second new scenario: all users in the system move slowly so that their covariance matrices change slowly with time. Even though covariance matrices under this condition change a little each time, if we do not update them, the new covariance matrices will be considerably different from the original ones after several periods of time. At that time, new covariance matrices need to be estimated separately again. Instead, we try to track the estimated covariance matrices as they change over time.

We briefly analyze a special case: one user in one cell moves slowly, so the channel covariance matrix of that cell changes (assume the user in the jth cell moves). Users in other cells have fixed covariance matrices. The unknown covariance matrix $\mathbf{R}_{jj}$ in this case can also be estimated by equation (4.11). But it is better to track the changing covariance matrix $\mathbf{R}_{jj}$ using the exponential moving average with forgetting

factor $\lambda_j \in [0,1]$

$$\hat{\mathbf{R}}_{jj}(n+1) \;=\; \lambda_j \boldsymbol{\Delta}_{\hat{\mathbf{R}}_{jj}(n)} + \big(1-\lambda_j\big)\hat{\mathbf{R}}_{jj}(n), \tag{4.13}$$

where

$$\boldsymbol{\Delta}_{\hat{\mathbf{R}}_{jj}(n)} = \left\{ \frac{1}{\tau}\left[\hat{\mathbf{R}}_j^y(n) - \omega^2 \mathbf{I}_M\right] - \sum_{l\neq j}^{L} \hat{\mathbf{R}}_{jl}(n) \right\}, \tag{4.14}$$

and

$$\hat{\mathbf{R}}_j^y(n) \;=\; (1-\lambda_j^y)\hat{\mathbf{R}}_j^y(n-1) + \lambda_j^y \mathbf{y}(b)\mathbf{y}(b)^H. \tag{4.15}$$

Note that we are also tracking the covariance matrix of $\mathbf{y}_j$, which is also changing due to the mobility of the user in jth cell. Therefore, we use a different forgetting factor $\lambda_j^y \in [0,1]$ in equation (4.15). The $\hat{\mathbf{R}}_{jj}(n)$ is the estimated value of $\mathbf{R}_{jj}(n)$ from previous block n and the $\boldsymbol{\Delta}_{\hat{\mathbf{R}}_{jj}(n)}$ is the updated part which is described in equation (4.14). Because only covariance matrix of jth cell changes, and $\sum_{l\neq v}^{L} \hat{\mathbf{R}}_{jl}(n)$ remain the same as n increases. Therefore, we can get an estimate of $\mathbf{R}_{jj}(n)$ by equation (4.13).

Now we consider the general situation of second scenario: covariance matrices of all channels are changing due to all users moving slowly. Under this condition, the estimation equation is the same as equation (4.13), but $\sum_{l\neq j}^{L} \hat{\mathbf{R}}_{jl}(n)$ also must be tracked. It appears that there are too many unknowns to track all the changing covariance matrices. However, we exploit the fact that the distance between each interfering user to the jth BS is typically much larger than the distance between user in the jth cell to the jth BS. If all users in the system move with the same speed, $\mathbf{R}_{jl}(n)$ changes more slowly than $\mathbf{R}_{jj}(n)$ changes. The further the distance between the interfering user and the jth BS is, the more slowly change in $\mathbf{R}_{jl}(n)$ compared with $\mathbf{R}_{jj}(n)$. If interfering users are originally distributed at the cell edges which are not very close to the jth BS, we can concentrate on estimating $\mathbf{R}_{jj}(n)$ and suppose interfering covariance matrices change much more slowly compare with the covariance matrices of the jth cell with movements of all users. Under this situation, the estimation equations are described as

$$\hat{\mathbf{R}}_{jl}(n+1) = \lambda_l \boldsymbol{\Delta}_{\hat{\mathbf{R}}_{jl}(n)} + \big(1-\lambda_l\big)\hat{\mathbf{R}}_{jl}(n), \tag{4.16}$$

where

$$\mathbf{\Delta}_{\hat{\mathbf{R}}_{jj}(n)} = \left\{ \frac{1}{\tau} \left[\hat{\mathbf{R}}_j^y(n) - \omega^2 \mathbf{I}_M \right] - \sum_{l \neq j}^{L} \hat{\mathbf{R}}_{jl}(n) \right\}, \tag{4.17}$$

and

$$\mathbf{\Delta}_{\hat{\mathbf{R}}_{jl}(n)} = \begin{cases} \frac{1}{n} \sum_{b=1}^{n} \hat{\mathbf{h}}_{jl}(b) \hat{\mathbf{h}}_{jl}(b)^H & , \text{if } n \leq W \\ \frac{1}{W} \sum_{b=n-W+1}^{n} \hat{\mathbf{h}}_{jl}(b) \hat{\mathbf{h}}_{jl}(b)^H, & \text{if } n > W, \end{cases} \tag{4.18}$$

in which $l \neq j$ in equations (4.16-4.18). The W is the length of window. Here we assume that $\hat{\mathbf{R}}_{jl}(n)$ can be updated in a decision-directed manner, in which channel estimates $\hat{\mathbf{h}}_{jl}(b)$ are assumed reliable enough to estimate the covariance matrices. We tried to estimate the covariance matrix of the desired user (with $l = j$) by equation (4.18) but it did not work: the estimates of $\mathbf{h}_{jj}(n)$ were apparently unreliable very quickly. When $n > W$, the value of $\mathbf{\Delta}_{\hat{\mathbf{R}}_{jl}(n)}$ in equation (4.18) is the sample average which just depends on the recent W blocks. Therefore, equation (4.18) tracks the new covariance matrices of received signal when users change their locations. The $\hat{\mathbf{R}}_j^y(n)$ in equation (4.17) can be achieved by equation (4.15). Therefore, we can estimate the channel covariance matrices by equation (4.16-4.18).

The proposed algorithm for the case which all users move is summarized below

Step 1) Before users moving, the jth BS achieves the original estimate value of $\hat{\mathbf{h}}_j$ and $\hat{\mathbf{R}}_{jl}$ through training;

Step 2) All users in system move slowly. The jth BS collects $\mathbf{y}_j(n)$ and calculates $\hat{\mathbf{R}}_j^y(n)$ based on equation (4.15) ;

Step 3) The jth BS achieves the value of $\hat{\mathbf{R}}_{jl}(n)$ based on equations (4.16-4.18);

Step 4) Estimate the $\mathbf{h}_{jl}(n)$ by the equation (4.12) based on $\hat{\mathbf{R}}_{jl}(n)$ in Step 3).

4.4 Simulation Results

In this section, we use hexagonally shaped cells in which all users are originally distributed at the circle with 800 meters to their own BSs. The cell radius is 1000 meters

Table 4.1: Basic Simulation Parameters of Chapter 4

Number of paths (P)	40
Number of cells (L)	4, 7
Carrier frequency	2GHz
Gain factor exponent (γ)	3
Pilot length (τ)	8
Antenna space (D)	$\mu/2$

and the cell-edge signal-to-noise ratio (SNR) of jth cell is 20 dB. Moreover, we assume that there is no overlap between AOAs for interfering and desired channels all the time. Other basic parameters are shown in table 4.1. We choose the bit error rate (BER) of 16-level quadrature amplitude modulation (16-QAM) for uplink data transmission and the normalized mean square error (NMSE) of CSI as metrics to test the performance of the proposed algorithms. The NMSE of CSI is denoted as

$$\mathrm{err}_{\mathbf{h}} \triangleq 10\log_{10}\left[\frac{\sum_{j=1}^{L}\left\|\hat{\mathbf{h}}_{jj}(n)-\mathbf{h}_{jj}(n)\right\|_F^2}{\sum_{j=1}^{L}\left\|\mathbf{h}_{jj}(n)\right\|_F^2}\right]. \tag{4.19}$$

Fig. 4.2 depicts the NMSEs of CSI when a new entrant comes into jth cell with otherwise fixed covariance matrices. The initial user in the jth cell drops out its service and interfering users in system have fixed covariance matrices. We suppose $j = 1$ in this figure. The distance between user and 1st BS is d_{1l}, such as $d_{11} = 800m$, $d_{12} = 2532m$, $d_{13} = 2241m$, $d_{14} = 2501m$, $M = 50$. After the user enters the 1st cell, the distance between the new user and its BS is $d_{11} = 750$ meters. We assume the new user comes into the cell at the first block. After the new users comes into the cell, all covariance matrices are fixed. It can be seen that the performance of our proposed algorithm is not very good when n is small. But with the increase of n, its performance becomes much better than that obtained by the LS estimation method or the MAP estimation method with un-updated (incorrect) covariance matrices which is shown in Fig. 4.2. That is because, the sample covariance matrices $\hat{\mathbf{R}}_y(n)$ are very different from true values of the $\mathbf{R}_y$ when n is small. But the difference quickly decreases as n increases.

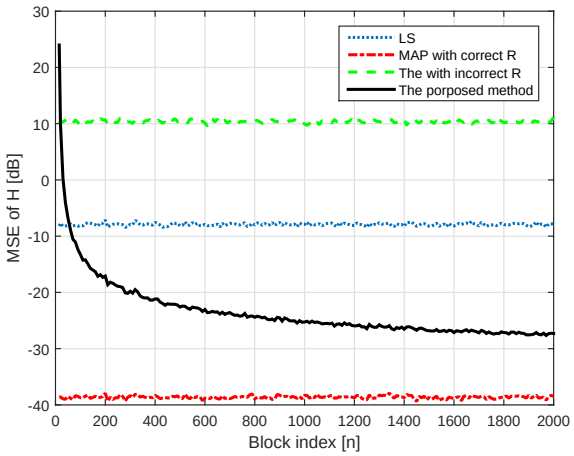

Figure 4.2: The NMSE of CSI for the case: new user enters a cell, $M = 50$, 4-cell network.

Therefore, the performance of our algorithm improves as n grows large.

We then study the second scenario: all users move slowly. In this situation, $M = 40$, $\lambda = 0.01$, $\lambda^y = 0.05$ and $W = 400$ (we did multiple experiments to tune these value). The details of the mobility of users are as follows: every user move slowly towards to their own BSs, we sample every 10 meters as a interval which is called one movement of users. Each movement lasts for 10 seconds. In summary, the user's movement is continuous but sampled. The channel covariance matrices are fixed during the same users' movements but change in different users' movements. In each movement of users, there are 62500 blocks transmitted (including training and data blocks) and 700 blocks are training blocks and used to track the new covariance matrices. Each training time, we use 700 blocks to track the changing covariance matrices, and then use these matrices to estimate CSI. These covariance matrices are averaged by 1000 drops of users.

Fig.4.3 depicts the NMSEs of $\mathbf{h}_{11}$. In this figure, the blocks refer to the training only. Users continue move slowly and we train 8 times. The distance between users and 1st BS is d_{1l}, such as $d_{11} = 800m$, $d_{12} = 2532m$, $d_{13} = 2457m$, $d_{14} = 2501m$. It can been seen from Fig.4.3 that when users change their locations, the NMSEs of proposed algorithm and MAP estimation with un-updated covariance matrices all rise. But the NMSEs of our proposed algorithm will come down quickly during the period in which

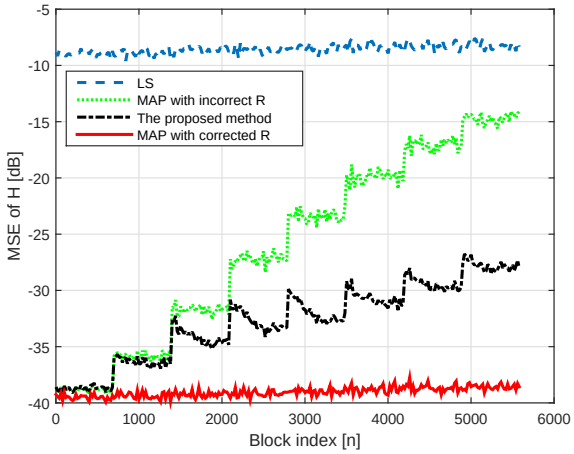

Figure 4.3: The NMSE of channel covariance matrix for the case: all users move slowly, $d_{11} = 800m$, $d_{12} = 2532m$, $d_{13} = 2457m$, $d_{14} = 2501m$, $M = 40$, 4-cell network,W=400.

there is no change, and then they keep relatively stable until the next movements of users. However, the NMSEs of MAP estimation with un-updated (incorrect) covariance matrices do not come down and continue rising with the movements of users. After users move 8 times, the NMSEs of proposed algorithm are clearly lower than those obtained by the LS method or the MAP method with un-updated (incorrect) covariance matrices.

Although the proposed method improves on the methods that do not attempt to track the changing covariance matrices, it is not clear if it tracks the changes sufficiently well to provide reliable communication. We examine this issue next.

Fig.4.4 describes the BER performance of 16-QAM. Parameters of this figure are the same as Fig. 4.3. There are two curves in this figure: BER performance with NMSE of $\mathbf{h}$ equal to -38 dB (when covariance matrices are estimated correctly); BER performance with NMSE of $\mathbf{h}$ equal to -30 dB. We argue that a BER of less than 10^{-6} is acceptable for data communication [52]. It can be seen that the curve which the NMSE of $\mathbf{h}$ equal to -30 dB is lower than 10^{-6} at SNR= 9 dB, which is an acceptable value in a practical digital system. According to Fig.4.4, for our proposed method, the NMSE= -30 dB after users move 7 times; however for the MAP estimation method with un-updated R, the NMSE= -30 dB when users only move 3 times.

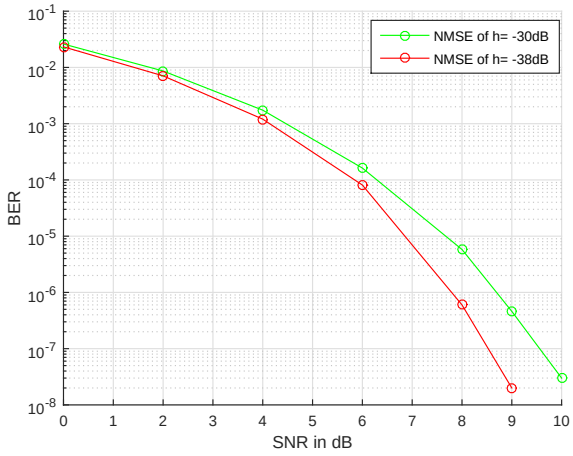

Figure 4.4: The BER performance of 16 QAM for the case: all users move slowly.

Therefore, it is reasonable to assume that before the covariance matrices would need to be re-estimated, the proposed method in the scenario illustrated in Fig. 4.3 provides satisfactory performance for about 437500 blocks (including training and data blocks), however, the MAP method with un-updated covariance matrices can provide good performance for only about 187500 blocks. After that the system has to estimate new covariance matrices separately again by the method in [11].

Fig. 4.5 shows performance of three estimation methods in a 7-cell network. In this figure, $d_{11} = 800m$, $d_{12} = 2532m$, $d_{13} = 2457m$, $d_{14} = 2457m$, $d_{15} = 2532m$, $d_{16} = 2241m$, $d_{17} = 2501m$. Other parameters are same as Fig. 4.3. From Fig. 4.5 we can see that, the NMSE of the proposed method is higher than that in Fig. 4.3. The reason is that the number of interfering cells in Fig. 4.5 is bigger than Fig. 4.3, which can reduce estimation accuracy of covariance matrices (refer equation (4.16-4.18)). Even though the NMSE of proposed method in 4.5 is higher than 4.3, it is still much better than the LS method and the MAP method with un-updated (incorrect) covariance matrices.

The Fig.4.6 shows the relationship between forgetting factor λ_j^y and the performance of proposed method. The forgetting factor of received signal λ_j^y in this figure is 0.05 or 0.01. Other parameters are the same as Fig.4.3. From Fig.4.6 we can get that the value of forgetting factor λ_j^y is able to affect the performance of our proposed method. The

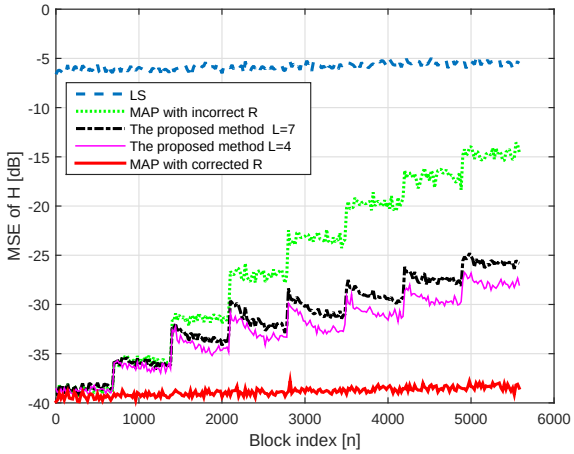

Figure 4.5: The NMSE of channel covariance matrix for the case: all users move slowly, 4 or 7-cell network W=400.

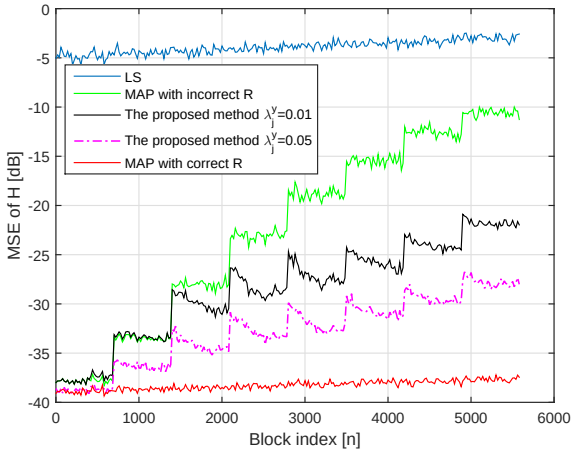

Figure 4.6: The NMSE of channel covariance matrix for the case: all users move slowly, 4-cell network, W=400, λ_j^y=0.05 or 0.01.

MSE of channel estimate with $\lambda_j^y = 0.05$ is lower than that with $\lambda_j^y = 0.01$. The reason is that, the forgetting factor balances between the updated part and the previous part. If λ_j^y is too small, it will not track the new covariance matrices very well during each user's movement and will increase the NMSE of channel estimate. Therefore, if users change quickly during every movement, λ_j^y should be relatively large to obtain new covariance matrices estimate.

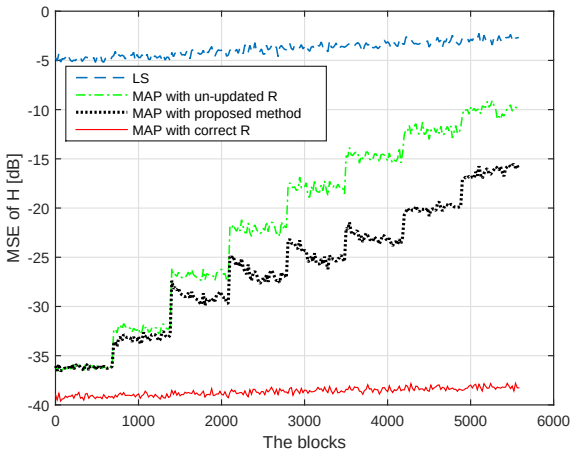

Figure 4.7: The NMSE of channel covariance matrix for the case: all users move slowly, $d_{11} = 800m$, $d_{12} = 1066m$, $d_{13} = 2241m$, $d_{14} = 1275m$, $M = 40$, 4-cell network, W=400.

In Fig.4.7, we want to test the influence of distances between interfering users and the jth BS on proposed method performance. In this figure, two interfering users' locations are close to the jth BS, such as $d_{11} = 800m$, $d_{12} = 1066m$, $d_{13} = 2241m$, $d_{14} = 1275m$, and other parameters are same as Fig.4.3. The NMSE in Fig.4.7 is bigger than that in Fig. 4.3. This is because the interfering users in Fig.4.7 are closer to jth BS compared with Fig.4.3, so the effect of interfering users' mobility on $\mathbf{R}_{1l}(n)$ is correspondingly more serious. Therefore, we can draw the conclusion that the further the interfering users are from the jth BS, the better the performance of the proposed algorithm.

4.5 Conclusion

This chapter studies covariance-aided estimation methods in new scenarios in which covariance matrices change due to the arrival of the new user or mobility of users. We analyze two different situations and propose algorithms to estimate the new covariance matrices and then obtain channel estimation results based on these estimated covariance matrices. Simulation results show that when the new covariance matrices are not directly available, our proposed algorithms can provide satisfactory performance than

those two algorithms (the LS method and MAP estimation method with un-updated (incorrect) covariance matrices) before the system re-estimates the new covariance matrices using the method in [11]. In other words, our proposed method reduces the frequency of pausing the system transmission to train the system one cell at a time.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

Pilot contamination has become the bottleneck for the performance of Massive MIMO systems. Covariance-aided channel estimation methods have been considered as a promising candidate to deal with this problem. Therefore, in this thesis, we focus on covariance-aided channel estimation methods under pilot contamination in Massive MIMO systems. Moreover, different from most covariance-aided methods which usually assume covariance matrices are fixed and known, we study new scenarios in which all or some of channel covariance matrices are not directly available. We focus on two specific scenarios: 1) a new user arrives in the cell with unknown covariance matrix; 2) covariance matrices are initially known but change over time due to mobility of users. For these cases, we develop novel algorithms to estimate new covariance matrices and use these estimated covariance matrices to obtain high quality channel estimates. Simulation results we obtain show that under pilot contamination, our proposed algorithms are superior to the Least Squares estimation method and the MAP estimation method with un-updated covariance matrices.

5.2 Future work

Although our proposed methods have the ability to obtain much better performance than Least Squares estimation and MAP estimation method with un-updated (incorrect) covariance matrices, it still can be improved. For example, the scenario which all users move regularly and slowly, when the interfering users are distributed at the edges close to the jth BS, there is a little benefit from our proposed algorithms. Therefore, some additional (or alternative) measures are needed to improve the performance for the situation which interfering users are distributed close to the jth BS, which is an interesting future topic for research.

References

[1] T. Marzetta, " The case for many (greater than 16) antennas as the base station," *Proc., Information Theory and its Applications (ITA), San Diego, CA*, 2007.

[2] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, " Massive MIMO for next generwation wireless systems," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 186–195, February 2014.

[3] T. Marzetta, " Noncooperative cellular wireless with unlimited numbers of base station antennas," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[4] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, " What will 5G be?" *Selected Areas in Communications, IEEE Journal on*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[5] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, " Massive MIMO performance evaluation based on measured propagation data," *Wireless Communications, IEEE Transactions on*, vol. 14, no. 7, pp. 3899–3911, July 2015.

[6] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, " Five disruptive technology directions for 5G," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 74–80, February 2014.

[7] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, " Pilot contamination problem in multi-cell TDD systems," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, June 2009, pp. 2184–2188.

[8] Z. Xiang, M. Tao, and X. Wang, " Massive MIMO multicasting in noncooperative cellular networks," *Selected Areas in Communications, IEEE Journal on*, vol. 32, no. 6, pp. 1180–1193, June 2014.

[9] Y. Zeng and T.-S. Ng, " A semi-blind channel estimation method for multiuser multiantenna OFDM systems," *Signal Processing, IEEE Transactions on*, vol. 52, no. 5, pp. 1419–1429, May 2004.

[10] Y. Li, N. Seshadri, and S. Ariyavisitakul, " Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *Selected Areas in Communications, IEEE Journal on*, vol. 17, no. 3, pp. 461–471, Mar 1999.

[11] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, " A coordinated approach to channel estimation in large-scale multiple-antenna systems," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 2, pp. 264–273, February 2013.

[12] A. Hu, T. Lv, and Y. Lu, " Subspace-based semi-blind channel estimation for large-scale multi-cell multiuser MIMO systems," in *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, June 2013, pp. 1–5.

[13] R. Muller, L. Cottatellucci, and M. Vehkapera, " Blind pilot decontamination," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 8, no. 5, pp. 773–786, Oct 2014.

[14] H. Q. Ngo and E. Larsson, " EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, March 2012, pp. 3249–3252.

[15] M. Alodeh, S. Chatzinotas, and B. Ottersten, " Spatial DCT-based channel estimation in multi-antenna multi-cell interference channels," *Signal Processing, IEEE Transactions on*, vol. 63, no. 6, pp. 1404–1418, March 2015.

[16] J. Hoydis, K. Hosseini, S. ten Brink, and M. Debbah, " Making smart use of excess antennas: Massive MIMO, small cells, and TDD," *Bell Labs Technical Journal*, vol. 18, no. 2, pp. 5–21, Sept 2013.

[17] I. E. Telatar, " Capacity of multi-antenna Gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[18] G. J. Foschini and M. J. Gans, " On limits of wireless communications in a fading environment when using multiple antennas," *Wireless personal communications*, vol. 6, no. 3, pp. 311–335, 1998.

[19] D. Tse and S. Hanly, " Linear multiuser receivers: effective interference, effective bandwidth and user capacity," *Information Theory, IEEE Transactions on*, vol. 45, no. 2, pp. 641–657, Mar 1999.

[20] O. Somekh, B. Zaidel, and S. Shamai, " Sum rate characterization of joint multiple cell-site processing," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4473–4497, Dec 2007.

[21] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, " Multi-cell MIMO Cooperative networks: A new look at interference," *Selected Areas in Communications, IEEE Journal on*, vol. 28, no. 9, pp. 1380–1408, December 2010.

[22] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, " Capacity limits of MIMO channels," *Selected Areas in Communications, IEEE Journal on*, vol. 21, no. 5, pp. 684–702, 2003.

[23] T. Yoo and A. Goldsmith, " On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 3, pp. 528–541, March 2006.

[24] J.-C. Shen, J. Zhang, and K. Letaief, " Downlink user capacity of Massive MIMO under pilot contamination," *Wireless Communications, IEEE Transactions on*, vol. 14, no. 6, pp. 3183–3193, June 2015.

[25] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, " Scaling up MIMO: Opportunities and challenges with very large arrays," *Signal Processing Magazine, IEEE*, vol. 30, no. 1, pp. 40–60, Jan 2013.

[26] W. Liu, "Efficient detection and scheduling for MIMO-OFDM systems," 2012.

[27] S. Lakshminarayana, M. Assaad, and M. Debbah, " Coordinated multicell beamforming for Massive MIMO: A random matrix approach," *Information Theory, IEEE Transactions on*, vol. 61, no. 6, pp. 3387–3412, June 2015.

[28] J. Chen and V. Lau, " Two-tier precoding for FDD multi-cell Massive MIMO time-varying interference networks," *Selected Areas in Communications, IEEE Journal on*, vol. 32, no. 6, pp. 1230–1238, June 2014.

[29] H. Huh, G. Caire, H. Papadopoulos, and S. Ramprashad, " Achieving "Massive MIMO" spectral efficiency with a not-so-large number of antennas," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 9, pp. 3226–3239, September 2012.

[30] S. A. Ramprashad, G. Caire, and H. Papadopoulos, " Cellular and network MIMO architectures: MU-MIMO spectral efficiency and costs of channel state information," in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, Nov 2009, pp. 1811–1818.

[31] H. Huh, A. Tulino, and G. Caire, " Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *Information Theory, IEEE Transactions on*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[32] T. Marzetta, " How much training is required for multiuser MIMO?" in *Signals, Systems and Computers, 2006. ACSSC '06. Fortieth Asilomar Conference on*, Oct 2006, pp. 359–363.

[33] C. Zhang and R. Qiu, " Massive MIMO as a big data system: Random matrix models and testbed," 2015.

[34] H. Q. Ngo, E. Larsson, and T. Marzetta, " Energy and spectral efficiency of very large multiuser MIMO systems," *Communications, IEEE Transactions on*, vol. 61, no. 4, pp. 1436–1449, April 2013.

[35] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, " Linear pre-coding performance in

measured very-large MIMO channels," in *Vehicular Technology Conference (VTC Fall), 2011 IEEE*, Sept 2011, pp. 1–5.

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

[37] J. Park and B. Clerckx, " Multi-user linear precoding for multi-polarized Massive MIMO system under imperfect CSIT," *Wireless Communications, IEEE Transactions on*, vol. 14, no. 5, pp. 2532–2547, May 2015.

[38] S. Park, C. byoung Chae, and S. Bahk, " Before/after precoding massive MIMO systems for cloud radio access networks," *Communications and Networks, Journal of*, vol. 15, no. 4, pp. 398–406, Aug 2013.

[39] H. Wang, W. Zhang, Y. Liu, Q. Xu, and P. Pan, " On design of non-orthogonal pilot signals for a multi-cell Massive MIMO system," *Wireless Communications Letters, IEEE*, vol. 4, no. 2, pp. 129–132, April 2015.

[40] L. You, X. Gao, X.-G. Xia, N. Ma, and Y. Peng, " Pilot reuse for Massive MIMO transmission over spatially correlated Rayleigh fading channels," *Wireless Communications, IEEE Transactions on*, vol. 14, no. 6, pp. 3352–3366, June 2015.

[41] X. Guozhen, L. An, J. Wei, X. Haige, and L. Wu, " Joint user scheduling and antenna selection in distributed massive MIMO systems with limited backhaul capacity," *Communications, China*, vol. 11, no. 5, pp. 17–30, May 2014.

[42] R. R. Mueller, M. Vehkaperae, and L. Cottatellucci, " Blind pilot decontamination," in *Smart Antennas (WSA), 2013 17th International ITG Workshop on*, March 2013, pp. 1–6.

[43] H. Yin, D. Gesbert, and L. Cottatellucci, " Dealing with interference in distributed large-scale MIMO systems: A statistical approach," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 8, no. 5, pp. 942–953, Oct 2014.

[44] X.-C. Gao, J.-K. Zhang, Z.-Y. Wang, and J. Jin, " Optimal precoder design maximizing the worst-case average received SNR for massive distributed MIMO Systems," *Communications Letters, IEEE*, vol. 19, no. 4, pp. 589–592, April 2015.

[45] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, " Pilot contamination and precoding in multi-cell TDD systems," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 8, pp. 2640–2651, August 2011.

[46] A. Liu and V. Lau, " Two-stage subspace constrained precoding in Massive MIMO cellular systems," *Wireless Communications, IEEE Transactions on*, vol. 14, no. 6, pp. 3271–3279, June 2015.

[47] T. Marzetta, " Massive MIMO: An introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–22, 2015.

[48] E. Bjornson, L. Sanguinetti, J. Hoydis, and M. Debbah, " Optimal design of energy-efficient multi-user MIMO systems: Is Massive MIMO the answer?" *Wireless Communications, IEEE Transactions on*, vol. 14, no. 6, pp. 3059–3075, June 2015.

[49] R. Knopp and P. Humblet, " On coding for block fading channels," *Information Theory, IEEE Transactions on*, vol. 46, no. 1, pp. 189–205, Jan 2000.

[50] S. K. Sengijpta, " Fundamentals of statistical signal processing: Estimation theory," *Technometrics*, vol. 37, no. 4, pp. 465–466, 1995.

[51] J.-A. Tsai, R. Buehrer, and B. Woerner, " The impact of AOA energy distribution on the spatial fading correlation of linear antenna array," in *Vehicular Technology Conference, 2002. VTC Spring 2002. IEEE 55th*, vol. 2, 2002, pp. 933–937 vol.2.

[52] E. A. Lee and D. G. Messerschmitt, *Digital communication.* Springer Science & Business Media, 2012.