UNIVERSITY OF CALIFORNIA, SAN DIEGO

TDD Massive MIMO Systems: Channel Estimation, Power Optimization, and Access Point Location Design

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Communication Theory and Systems)

by

Elina Nayebi

Committee in charge:

Professor Bhaskar D. Rao, Chair
Professor Robert R. Bitmead
Professor William S. Hodgkiss
Professor Laurence B. Milstein
Professor Paul H. Siegel

2018

The dissertation of Elina Nayebi is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2018

EPIGRAPH

Research is what I'm doing when

I don't know what I'm doing.

Wernher von Braun

Your assumptions are your windows on the world.

Scrub them off every once in a while, or the light won't come in.

Isaac Asimov

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

It is my pleasure to acknowledge the roles of several individuals who were instrumental for completion of my Ph.D. research. First, I would like to express my sincere gratitude to Professor Bhaskar D. Rao, who expertly guided me through my graduate education and who shared the excitement of five years of discovery. I vividly remember the birth of my interest in wireless communication from the first time I attended his class back in year 2013. He introduced me to the nature of research and taught me how to find and pursue the impactful research directions. He always encouraged me to keep close relations with industry and learn about the challenges in today's engineering world, which ultimately gave birth to my research topics. He continually and convincingly conveyed a spirit of adventure in regard to research and scholarship without which this dissertation would have not been possible. I am also thankful to my committee members, Professor Robert R. Bitmead, Professor William S. Hodgkiss, Professor Laurence B. Milstein, and Professor Paul H. Siegel for their invaluable recommendations on my research. I am also thankful for the lessons and skills I learned from many courses taught by these inspiring teachers.

My appreciation also extends to Bell Labs for letting me fulfill my dream of working there by offering internship opportunities for two consecutive summers at early stages of my graduate studies. The foundation of this dissertation is partially based on the experiences I have gained while working on Massive MIMO project at Bell Labs. I am really fortunate that I had the kind supervision of Dr. Alexei Ashikhmin. His exemplary guidance, constant encouragements, and careful monitoring throughout the internships were so great that even my most profound gratitude is not enough. I also salute Dr. Thomas L. Marzetta for his tenacity of purpose, outstanding leadership qualities, and for sharing his vision on the future of wireless communication, which opened my eyes to many new opportunities. I am also thankful to my

other mentor at Bell Labs, Dr. Hong Yang for his continued care and support.

I am also indebted to the Department of Electrical and Computer Engineering at UC San Diego for offering me the first-year fellowship along with the program admission, which gave me freedom to search and find the research field whom I have the most passion for. I want to give my thanks to my former and current colleagues at UCSD Digital Signal Processing Lab, Dr. Bang Nguyen, Dr. Ritwik Giri, Dr. Yonghee Han, Yacong Ding, David Ho, Soon-En Chiu, Govind Gopal, Igor Fedorov, Tharun Srikrishnan, Richard Bell, and Furkan Kovasoglu. I am thankful to them for attending my presentations, their helpful feedbacks, and most importantly, their moral support over the years.

I would also like to thank my family, my parents and my sister, for their unconditional love and support. They have always been there for me and motivated me throughout this journey, without which I would have not made it this far.

Chapter 2 contains material as it appears in [1], E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and Power Optimization in Cell-Free Massive MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4445-4459, July 2017, and is also based on the material as it appears in [2], E. Nayebi, A. Ashikhmin, T. L. Marzetta, and H. Yang, "Cell-Free Massive MIMO Systems," in *Proc. 49th Asilomar Conference on Signals, Systems and Computers*, Nov. 2015, pp. 695-699. The dissertation author was the primary investigator and author of these papers.

Chapter 3 contains material as it appears in [3], E. Nayebi, A. Ashikhmin, T. L. Marzetta, B. D. Rao, "Performance of Cell-Free Massive MIMO Systems with MMSE and LSFD Receivers," in *Proc. 50th Asilomar Conference on Signals, Systems and Computers*, Nov. 2016, pp. 203-207. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part is currently being prepared for submission for publication of the

material. The dissertation author is the primary investigator and author of this material and B. D. Rao supervised the research.

Chapter 5, in part, is a reprint of the material as it appears in [4], E. Nayebi, and B. D. Rao, “Semi-blind Channel Estimation for Multiuser Massive MIMO Systems," *IEEE Transactions on Signal Processing*, vol. 66, no.2, pp. 540-553, Jan. 2018, and is also based on the material as it appears in [5], E. Nayebi, and B. D. Rao, “Semi-blind Channel Estimation in Massive MIMO Systems with Different Priors on Data Symbols” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2018*, to appear. The dissertation author was the primary investigator and author of these papers.

VITA

2012	Bachelor of Science in Electrical Engineering, University of Tehran, Iran
2014	Master of Science in Electrical Engineering (Communication Theory and Systems), University of California, San Diego
2012-2017	Research Assistant, University of California, San Diego
2018	Doctor of Philosophy in Electrical Engineering (Communication Theory and Systems), University of California, San Diego

PUBLICATIONS

E. Nayebi, A. Ashikhmin, T. L. Marzetta, and H. Yang, "Cell-Free Massive MIMO Systems," in *Proc. 49th Asilomar Conference on Signals, Systems and Computers*, Nov. 2015, pp. 695-699.

E. Nayebi, A. Ashikhmin, T. L. Marzetta, B. D. Rao, "Performance of Cell-Free Massive MIMO Systems with MMSE and LSFD Receivers," in *Proc. 50th Asilomar Conference on Signals, Systems and Computers*, Nov. 2016, pp. 203-207.

E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and Power Optimization in Cell-Free Massive MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4445-4459, July 2017.

E. Nayebi and B. D. Rao, "Semi-blind Channel Estimation for Multiuser Massive MIMO Systems," *IEEE Transactions on Signal Processing*, vol. 66, no.2, pp. 540-553, Jan. 2018.

E. Nayebi and B. D. Rao, "Semi-blind Channel Estimation in Massive MIMO Systems with Different Priors on Data Symbols" in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2018*, to appear.

FIELDS OF STUDY

Major Field: Electrical Engineering

Studies in Communication Theory and Systems

Advisor: Bhaskar D. Rao

ABSTRACT OF THE DISSERTATION

TDD Massive MIMO Systems: Channel Estimation, Power Optimization, and Access Point Location Design

by

Elina Nayebi

Doctor of Philosophy in Electrical Engineering (Communication Theory and Systems)

University of California, San Diego, 2018

Professor Bhaskar D. Rao, Chair

With an ever-increasing demand for higher wireless throughput, there has been growing interest in massive multiple-input multiple-output (MIMO) as a key technology for future wireless networks. This dissertation addresses some of the key aspects of this technology that include: 1. precoding, power optimization, and access point (AP) location design in cell-free massive MIMO systems with distributed APs; 2. semi-blind channel estimation in massive MIMO systems.

Cell-free massive MIMO is a special deployment of massive MIMO systems with a large number of distributed low-cost low-power single antenna APs serving a much smaller

number of users. The cell-free system is not partitioned into cells and each user is served by all APs simultaneously. The downlink capacity lower bounds for conjugate beamforming and zero forcing precoders in cell-free systems are derived in this dissertation. To further increase the achievable throughput, max-min power optimization algorithms are formulated, and low complexity max-min power allocation algorithms are developed. We also introduce a technique that employs ℓ_1-norm sparsity penalty in the max-min power optimization for conjugate beamforming that helps us decrease the number of APs that serve a user in a practical system.

The uplink capacity lower bounds for minimum mean squared error (MMSE) and large scale fading decoding receivers in cell-free systems are provided. A deterministic approximation for signal-to-interference-plus-noise ratio of MMSE receiver is obtained with an unlimited number of APs and user devices.

Next, AP location design problem is investigated to maximize the sum-throughput and the minimum-throughput in uplink transmission of cell-free systems with an arbitrary user distribution. Utilizing compressed sensing techniques, the AP placement problems are formulated as convex optimization problems. An AP location design algorithm is also presented in an alternative small-cell system in which each user is served by only one AP.

Finally, semi-blind channel estimation for multiuser massive MIMO systems is investigated. Multiple semi-blind channel estimation techniques based on the expectation-maximization algorithm are developed by considering different priors on data symbols. Cramer Rao Bounds (CRBs) for semi blind channel estimation are derived for deterministic and stochastic (Gaussian) data symbol models to give us an analytical understanding of the semi blind scheme's performance. To get insight into the behavior of a massive MIMO system, the asymptotic behavior of the CRBs as the number of antennas at the base station grows is analyzed.

Chapter 1

Introduction

1.1 Background on Massive MIMO Systems

Multiple-input multiple-output (MIMO) technology is one of the key components to increase spectral efficiency of a wireless network [6–8], which can be categorized into three main groups: 1. point-to-point MIMO, 2. multiuser MIMO, 3. massive MIMO. Point-to-point MIMO is the simplest variant of a MIMO system, which refers to communication between an access point (AP) and a user that are equipped with multiple antennas. Point-to-point MIMO system improves reliability of communication via spatial diversity. Under favorable channel conditions, having multiple transmit and receive antennas provides us with the spatial degrees of freedom, which can be exploited to transmit multiple data streams over the MIMO channel and increase the channel capacity [8]. Define by $G \in \mathbb{C}^{M\times K}$ the channel matrix between the transmitter and receiver equipped with M and K antennas respectively. The spectral efficiency (in bits per second per hertz) of a point-to-point MIMO for a deterministic and constant channel with additive white Gaussian noise and perfect channel state information (CSI) at receiver is expressed as [9]

$$C = \log_2 \det\left(I_K + \frac{\rho}{M} G^H G\right), \tag{1.1}$$

where ρ is the signal-to-noise ratio (SNR). Let $\lambda_1, \cdots, \lambda_{\min(M,K)}$ be singular values of G. Then the spectral efficiency (1.1) can be rewritten as

$$C = \sum_{i=1}^{\min(M,K)} \log_2\left(1 + \frac{\rho}{M}\lambda_i^2\right). \tag{1.2}$$

From (1.2), it is observed that the capacity grows linear in $\min(M, K)$ without increasing transmit power or bandwidth. The upper and lower bounds on (1.2), when the channel matrix is normalized as $\text{tr}\left(G^H G\right) \approx MK$, are given by [10]

$$\log_2(1 + \rho K) \leq C \leq \min(M, K) \log_2\left(1 + \frac{\rho \max(M, K)}{M}\right). \tag{1.3}$$

The distribution of singular values of G has an important role on achievable rate of the system. In the rich scattering propagation environments and a high SNR regime, the achievable rate of a point-to point MIMO is proportional to the rank of the channel matrix, i.e, $\min(M, K)$. In other words, the upper bound in (1.3) is achieved when the singular values of G are all equal [10], [11] . However, line-of-site conditions cause rank deficiency in channel matrix and degrade performance of the point-to point MIMO systems. The lower bound in (1.3) is obtained when only one singular value of channel matrix is non-zero [10], [11].

In multiuser MIMO technology, a single transmitter serves multiple receivers that utilize the same time and frequency resources in a communication system. These systems provide advantages over point-to point MIMO: they are less sensitive to the propagation properties of the environment, and a single antenna suffices at each user device making them relatively cheap devices. Therefore, multiuser MIMO has been largely deployed in communication standards, such as 802.11 (WiFi), 802.16 (WiMAX), and long-term evolution (LTE). However, in the downlink transmission of multiuser MIMO, both AP and users must know the propagation channel. Thus, considerable amount of known pilot sequences should be used to learn the channel coefficients. Advantages and challenges associated with the multiuser MIMO systems and their performance analysis can be found in [12], [13], [14], and references therein. The sum spectral efficiency of multiuser MIMO with K single-antenna user devices in uplink and downlink transmissions is given by [9]

$$C^{\text{ul}} = \log_2 \det\left(I_M + \rho_u G G^H\right), \tag{1.4}$$

$$C^{\text{dl}} = \max_{P,\, \text{tr}(P)\leq 1} \log_2 \det\left(I_M + \rho_d G P G^H\right), \tag{1.5}$$

where ρ_u and ρ_d are the uplink and downlink SNRs respectively. In uplink spectral efficiency (1.4), the CSI is only required at base station. However, in the downlink spectral efficiency (1.5), the CSI is required at both base station and user devices.

Massive MIMO is a scalable form of multiuser MIMO, where APs are equipped with large number of antenna elements [15], [16]. One of the key distinctions between multiuser and massive MIMO systems is that the number of AP antennas is much larger than the number of users in massive MIMO (see Figure 1.1). With the growing demand for throughput in wireless networks, massive MIMO is a promising technology that achieves much better performance compared with the multiuser MIMO. With excessively large number of AP antennas, massive MIMO serves many users in the same time-frequency resources. Adding more antennas to the AP results in higher system throughput, and narrower beams (see [10, Figure 2]). Moreover, one can utilize large number of low-cost low-power antenna elements as opposed to expensive amplifiers that are used in conventional multiuser MIMO systems.

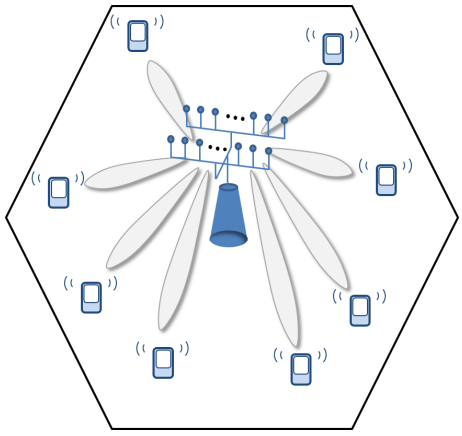

Figure 1.1: Massive MIMO.

In time-division duplex (TDD) protocol, the uplink and downlink physical channels are reciprocal [17]. Thus, propagation channel can be measured in uplink and used for both uplink data detection and downlink bemaforming. Most emphasis in massive MIMO systems is in TDD protocol rather than frequency-division duplex (FDD) due to the fact that the amount of pilot overhead to estimate the uplink channel coefficients is independent of the number of AP antennas, and the users are not required to feed the channel coefficients back to the AP. While

the focus of this dissertation is on massive MIMO systems with TDD protocol, it is important to note that in environments with enough special structure in the propagation, sparse properties of channel matrix in a particular domain can be exploited to use smaller number of downlink training pilots and uplink feedback to estimate the channel and hence one can utilize FDD operation in massive MIMO systems as well [18], [19].

TDD massive MIMO technology has many advantages over multiuser MIMO. It requires only AP to know the propagation channel. The number of training pilots depends only on the number of users and hence massive MIMO is scalable with respect to the number of AP antennas.

We denote the channel matrix between M antennas at the base station and K single antenna users in massive MIMO by $G \in \mathbb{C}^{M\times K}$, which can be modeled as the product of small scale fading factor $H \in \mathbb{C}^{M\times K}$ and a diagonal large scale fading factor $B \in \mathbb{R}^{K\times K}$ as follows

$$G = HB^{1/2}. \tag{1.6}$$

In massive MIMO, with independent small scale fading coefficients , columns of the channel matrix become asymptotically orthogonal, as the number of base station antennas increases [15], i.e.,

$$\frac{1}{M}G^H G \xrightarrow[M\to\infty]{a.s.} B. \tag{1.7}$$

This feature, which is a result of law of large numbers, is referred to as the channel hardening effect in massive MIMO. The achievable rates in (1.4) and (1.5) with infinite number of base

station antennas and favorable propagation conditions (1.7) given in [10] are as follows

$$C^{\text{ul}} \overset{M\gg K}{\approx} \sum_{i=1}^{K} \log_2\left(1 + M\rho_u\beta_i\right) \tag{1.8}$$

$$C^{\text{dl}} \overset{M\gg K}{\approx} \max_{\substack{p_1,\cdots,p_K\geq 0 \\ \sum_{k=1}^{K} p_k=1}} \sum_{i=1}^{K} \log_2\left(1 + M\rho_d p_i\beta_i\right). \tag{1.9}$$

In [11], it has been shown that conjugate beamforming, a simple linear precoding scheme, can achieve the achievable rate in (1.9) when the number of base station antennas grows to infinity. A similar result is also obtained in uplink: the simple linear matched-filtering (MF) processing at the base station achieves the multiuser uplink capacity defined in (1.8). This result shows that simple linear precoding and decoding schemes such as conjugate beamforing (CB) and zero-forcing (ZF) precoders can be used in massive MIMO. In [9] and [15], it is shown that the effects of uncorrelated noise, small scale fading, and non-coherent interference are eliminated when the number of base station antennas increases under the favorable propagation conditions (1.7). Therefore, simple power control algorithms can be deployed that no longer depend on the frequency.

One of the challenges in massive MIMO systems, is pilot contamination problem [15], [20], which emerges in multi-cell scenarios when the number of users is larger than the length of pilot sequences. Due to the limitation on channel coherence interval, the length of training pilots is limited. Therefore, training pilots of users in neighboring cells will be non-orthogonal. This leads to channel estimates that are contaminated by pilots transmitted by users in other cells and results in coherent interference that unlike non-coherent interference grows with the number of AP antennas. Pilot Contamination effect is a severe limiting factor for achievable data rate that degrades system performance considerably. Several channel estimation, precoding, and postcoding techniques have been introduced to mitigate the pilot contamination effect. For example, in [20], a multi-cell minimum mean squared error (MMSE) based precoding is

proposed. In [21] and [22], large scale fading decoding (also known as pilot contamination postcoding) for cellular massive MIMO is proposed, in which the base station of neighboring cells cooperate by sharing between themselves the large scale fading coefficients. In [23], the authors propose an eigenvalue decomposition-based approach to estimate channel coefficients that mitigates the pilot contamination problem to some extent.

1.2 Dissertation Contributions and Organization

Performance of massive MIMO systems with co-located antenna arrays are well studied in the literature, e.g., see [24], [11] and references therein. Another deployment of massive MIMO systems is a network comprising a large number of distributed single-antenna APs, where each user served simultaneously by all of the APs (see Figure 2.1). We call such a system *Cell-Free Massive MIMO* and investigate its performance in chapter 2. Since APs are spread out over a designated area in the cell-free systems, each user is close to a few APs, and thus can benefit from diversity against shadow fading. Hence, cell-free massive MIMO offers more coverage probability and power efficiency compared with the co-located systems. We derive capacity lower bounds for cell-free systems utilizing CB and ZF precoders and formulate max-min power control algorithms to provide equal throughput to all users. A number of low complexity power allocation algorithms for conjugate bramforming and zero-forcing precoders are also proposed that have significantly smaller computational complexity and yet achieve near-optimal performances. We quantitatively compare the performance of cell-free massive MIMO to that of a small-cell system in which each user is served by only on AP. We present a technique to reduce the number of APs serving each user for CB by utilizing ℓ_1-norm sparsity penalty term in the max-min optimization problem.

Chapter 3, provides uplink performance study of cell-free systems with MMSE and

large scale fading decoding (LSFD) receivers. The LSFD receiver maximizes the achievable throughput using only large scale fading coefficients between APs and users. Capacity lower bounds for MMSE and LSFD receivers are derived and a deterministic approximation for signal-to-interference-plus-noise ratio (SINR) of MMSE receiver is obtained when the number of APs and users grow infinitely large, which is an accurate approximation even for a small number of APs and users.

Performance of cell-free systems are largely influenced by AP locations, which needs to be optimized according to the user distribution in the system. In chapter 4, using compressed sensing techniques, two AP location design algorithms are obtained to maximize the sum-throughput and the minimum-throughput in uplink transmission of cell-free systems. We also introduce another AP location design algorithm in a small-cell system in which each user is served by only one AP.

In massive MIMO systems, the downlink linear precoding and uplink decoding operations require channel state information at the base station, and hence the actual propagation channels are required to be measured [16]. In chapter 5, we investigate semi-blind channel estimation for multiuser TDD massive MIMO systems with co-located antenna arrays. We derive a number of channel estimation algorithms using both uplink training pilots and data symbols. We derive a tractable expectation-maximization (EM) algorithm using a Gaussian prior for the unknown data symbols. An alternate EM algorithm is also derived by employing suitable priors on the channel coefficients, which outperform the EM algorithm with no channel priors in the low SNR regime. We further derive another semi-blind channel estimation algorithm based on the EM algorithm by using Gaussian mixture model (GMM) for the unknown data symbols that outperforms the EM algorithm with Gaussian prior as the SNR or as the number of antennas at the base station increases. Cramer Rao Bounds (CRBs) for semi blind channel estimation are derived for deterministic and stochastic (Gaussian) data symbol

models to give us an analytical understanding of the semi blind scheme performance. To get insight into the behavior of a massive MIMO system, the asymptotic behavior of the CRBs as the number of antennas at the base station grows is analyzed. In summary, semi-blind channel estimation methods become closer to the genie-aided maximum likelihood estimator based on known data symbols as the number of base station antennas increases, making semi-blind estimation more attractive for massive MIMO systems.

Chapter 2

Precoding and Power Optimization in Cell-Free Massive MIMO Systems

Cell-free massive multiple-input multiple-output (MIMO) comprises a large number of distributed low-cost low-power single antenna access points (APs) connected to a network controller. The number of AP antennas is significantly larger than the number of users. The system is not partitioned into cells and each user is served by all APs simultaneously. The simplest linear precoding schemes are conjugate beamforming and zero-forcing. Max-min power control provides equal throughput to all users and is considered in this chapter. Surprisingly, under max-min power control, most APs are found to transmit at less than full power. A power allocation algorithm is proposed for conjugate beamforming to reduce the number of effective APs that serve a particular user in cell-free massive MIMO. The zero-forcing precoder significantly outperforms conjugate beamforming. For zero-forcing, a near-optimal power control algorithm is developed that is considerably simpler than exact max-min power control. An alternative to cell-free systems is small-cell operation in which each user is served by only one AP for which power optimization algorithms are also developed. Cell-free massive MIMO is shown to provide five- to ten-fold improvement in 95%-likely per-user throughput over small-cell operation.

2.1 Introduction

A comprehensive wireless system should provide uniformly good service throughout a designated area. To that end, massive multiple-input multiple-output (MIMO) has attracted considerable attention as a candidate for the fifth generation physical layer technology [15], [10], [25].

Massive MIMO is a scalable form of multiuser MIMO. In [13], [14], [26] MIMO systems, with the assumptions that both ends of the link know the propagation channel and dirty paper coding is used, were studied. It was shown that such systems have very large

capacity that grows along with the number of base station antennas. The above assumptions, however, would forever limit the practical size of the wireless system. Instead massive MIMO systems proposed in [15] assume that only the base stations know the propagation channels and simple linear precoding is used instead of dirty paper coding. These assumptions make massive MIMO systems fully scalable with respect to the number of base station antennas.

Multi-cell massive MIMO systems, in which each cell is served by an array of co-located antennas, are well understood, e.g., see [24], [11] and references therein. A diametrically opposite massive MIMO deployment that would serve the same designated area, which could be an isolated village, a college campus, or an entire metropolitan area, is a network comprising a large number of randomly-located single-antenna access points (APs), with each user served simultaneously by all of the APs. We call such a system *Cell-Free Massive MIMO*.

Some of the limitations and advantages of the networks with distributed APs have been already studied under different names and settings, e.g., see [27–37]. In [31], the authors studied distributed algorithms for multi-cell beamforming and power allocation without data sharing among base stations. By contrast, in our work, all APs transmit data symbols to the users. In [32], performance of cooperative multi-cell zero-forcing (ZF) beamforming with a user selection scheme has been studied. The system performance is investigated in terms of sum-rate. In [33], downlink performance of multi-cell system with ZF beamforming in frequency division duplexing (FDD) system is studied in terms of sum-rate. In [34], the average sum-rate performance of distributed antennas for massive MIMO systems in uplink transmission is investigated, while all users transmit with the same power. In contrast, in this work, we consider time-division duplex operation (TDD) and analyze downlink performance of cell-free systems with emphasis on per-user throughput, rather than sum-throughput, by using max-min power allocation algorithms. In [35], capacity lower bounds of a multi-cell massive MIMO system has been derived for uplink and downlink transmission. In [36], asymptotic rate performance

of downlink multiuser systems with distributed antennas and perfect channel state information (CSI) was studied. By contrast, we take into account the effects of imperfect CSI with finite number of APs and users. In [37], the authors study uplink performance of large-scale distributed antenna settings with matched filtering (MF) receiver. They exploit low-rankness of users' channel covariance matrices to improve channel estimation and interference rejection under the assumption that all users share the same pilot sequence. However, further research is required in order to thoroughly understand these systems from the massive MIMO point of view with the goal of providing uniformly good service for all, or almost all, users.

In [38], [39], performance of conjugate beamforming (CB) in cell-free systems has been investigated with emphasis on pilot assignment algorithms to mitigate pilot contamination effect. The max-min power allocation in CB is used to increase the system performance. This power allocation algorithm involves a non-convex optimization problem with high computational complexity. In this work for CB precoder we propose low complexity power allocations algorithms (see section 2.3.2, 2.3.3, and 2.3.4) that have only moderate loss in terms of the system performance, but have significantly smaller complexity than the algorithm in [38], [39]. We further consider ZF precoder in cell-free systems with max-min power allocation, and introduce a simple near-optimal power control algorithm. We assume low mobility users. Since users move slowly the number of available orthogonal pilots is significantly larger than the number of users (see section 2.2.2), and therefore the pilot contamination is negligible in our systems.

Similar to [38], [39] we assume that all APs are connected to a network controller (NC) via an unspecified backhaul network. The controller conducts linear precoding and optimizes the transmit powers to improve the system performance. We propose several power allocation algorithms, with different levels of complexity. We further derive capacity lower bounds for cell-free systems utilizing CB and ZF precoders. These bounds take into account

the channel estimation error, the effective channel uncertainty at users, and other important parameters, that many papers ignore. The performance of cell-free systems in [38] is compared to that of small-cell systems using MF receiver. As a counterpart we also consider a *small-cell* system in which a user is served only by a single AP which is typically the closest one. For small-cell system we also consider MF receiver and ZF precoder, propose power optimization algorithms, and derive capacity lower bounds.

For performance criteria we use the *minimum rate* among all users and 5%-*outage rate*, which is the smallest rate among 95% of the best users. In future wireless systems all, or almost all, users will have to be served with large rates. So we believe that these criteria are more meaningful compared with the often used sum throughput. We formulate a number of max-min optimization problems to optimize the above criteria. The max-min optimization does not necessarily give the optimal 5%-outage rate. We still use the max-min optimization as an engineering tool for optimization of this criterion, and it leads to good system performance.

The main results of this work are the following. We show that cell-free systems give a very significant gain ($5 - 10$ fold) over small-cell systems, i.e., a system where each user is served by a single AP. The ZF precoder, significantly outperforms CB. We present a counter-intuitive result that the optimal max-min power control for CB requires that most APs would transmit with powers that are visibly smaller than the transmit power limit. Motivated by this result, we propose low complexity power allocation algorithms for CB precoder. In a real life cell-free system, each user will be served not by all APs, but by a subset of them located around the user, which can be viewed as an intermediate case between cell-free and small-cell systems. To this end, we further propose a max-min power allocation algorithm for CB precoding based on ℓ_1-norm penalty that decreases the number of APs that serve a user. For ZF precoder we propose a suboptimal power allocation algorithm scheme. We show that this algorithm has near-optimal performance while its complexity is very low. For small-cell

systems we also consider MF receiver and ZF precoder and power allocation algorithms. We conclude that cell-free systems, with ZF precoder in particular, outperform small-cell systems (about 10 times) in terms of 5%-outage rate. To the best of our knowledge, the above technical problems with our system settings have not been addressed in the literature yet.

The chapter is organized as follows. Section 2.2 describes the system model and channel estimation. In sections 2.3 and 2.4, we investigate CB and ZF precoders for cell-free massive MIMO system with max-min power control. In section 2.5, we consider small-cell systems. Finally, simulation results are presented in section 2.6.

Throughout the chapter superscripts ^{T},* , and ^{H} denote transpose, complex conjugate and hermitian operations respectively. Uppercase and bold symbols are utilized to denote matrices and vectors respectively, and $\mathbb{E}()$ is the expectation operator. Operator diag$\{A\}$ denotes the column vector of the main diagonal elements of matrix A.

2.2 System Model and Channel Estimation

2.2.1 Cell-Free System Model

Unlike cellular wireless networks, in a cell-free system we do not partition the network into cells and do not assign users to particular base stations. Instead we assume that a geographical area is covered by M randomly distributed single antenna APs. We assume that in this area there are K single antenna users, and that $M \gg K$. An example of a cell-free system is shown in Figure 2.1. In contrast to a standard cellular network, in a cell-free system each user is served not by one base station, but by all APs simultaneously. All APs are connected to a NC (not shown in Figure 2.1).

We use a flat fading channel model for each OFDM subcarrier. The OFDM subcarrier

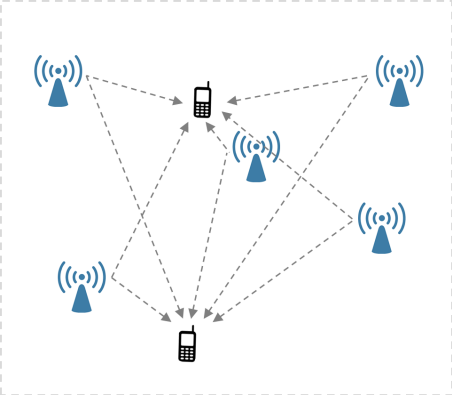

Figure 2.1: Cell-free system. All APs serve all users.

index is omitted for simplicity. Size of the overall area is considered small enough that the largest difference between propagation time from any two APs to a user is smaller than the duration of the OFDM cyclic prefix. The *channel coefficient* between AP m and user k is given by

$$g_{mk} = \sqrt{\beta_{mk}} h_{mk}, \tag{2.1}$$

where β_{mk} is the *large scale fading coefficient* which accounts for path loss and shadowing effects. This coefficient changes slowly and hence can be accurately estimated and tracked. Throughout the text we assume the NC knows the coefficients $\beta_{mk}, \forall m, k$. The second factor $h_{mk} \sim \mathcal{CN}(0, 1)$ is the *small scale fading coefficient*. We assume that these coefficients are i.i.d. random variables that stay constant during a coherent interval and are independent in different coherent intervals. For a wide-band OFDM system β_{mk} is independent of frequency, while h_{mk} has frequency dependence and a Nyquist sampling interval in frequency that is equal to the reciprocal of the channel delay-spread. We denote by $G \in \mathbb{C}^{M \times K}$, $[G]_{mk} = g_{mk}$ the *channel matrix* between all APs and users. We further assume channel reciprocity, i.e., that the uplink and downlink channel coefficients are the same.

We focus on the scenario of users with mobility less than 10km/h. In other words we

assume that most of our users are pedestrians, which is typically the case in real life scenarios.

2.2.2 Channel Estimation

The main idea of cell-free systems is that each user is served by all APs. In order to reduce interference between signals intended for different users the APs should form their transmitted signals by taking into account the channel coefficients. Thus, estimation of these coefficients is an important part of the communication protocol.

In this work we assume that the TDD protocol described in [40, Figure 3], is used. At the first step of this protocol all users simultaneously and synchronously transmit pilot sequences $\boldsymbol{\psi}_1, \cdots, \boldsymbol{\psi}_K \in \mathbb{C}^\tau$, which propagate to all M APs. At the second step all APs get estimates $\hat{g}_{mk}$ of g_{mk} and use these estimates to beamform data to all users.

We assume that pilot sequences $\boldsymbol{\psi}_i, i = 1, \cdots, K$, assigned to the corresponding users are orthonormal, i.e., $\boldsymbol{\psi}_i^H \boldsymbol{\psi}_j = \delta_{ij}$. As we mentioned before we assume the mobility of users less than 10km/h. For such speeds and carrier frequency of 1.9 GHz, the coherence interval is large, which enables using a large number of orthogonal pilots for channel estimation. Hence it is reasonable to assume that these pilots are assigned to users in such way that users with the same pilot are located far away from each other and the pilot contamination (coherent interference resulting from two or more users sharing the same pilot sequence) is negligible.

Remark 2.1. *In the case of cell-free systems with users of high mobility, the number of orthogonal pilots is significantly smaller and the pilot contamination caused by reuse of the same pilots may result in additional interference, see [39]. The main goal of this chapter, however, is to understand the phenomenology associated with precoding techniques (CB and ZF) and power allocation algorithms in cell-free systems. We believe that the obtained power allocation algorithms will be applicable in cell-free system where some users have high mobility.*

The received signal sequence in the training step at the m-th AP is

$$\boldsymbol{y}_m = \sqrt{\rho_r \tau} \sum_{i=1}^{K} g_{mi} \boldsymbol{\psi}_i + \boldsymbol{w}_m, \tag{2.2}$$

where ρ_r is the *uplink power* and $\boldsymbol{w}_m \sim \mathcal{CN}(0, I_\tau)$ is *additive noise* and τ is the *length of pilot sequences*. AP m computes the MMSE estimate of g_{mk} as

$$\hat{g}_{mk} = \frac{\sqrt{\rho_r \tau}\beta_{mk}}{1 + \rho_r \tau \beta_{mk}} \boldsymbol{\psi}_k^H \boldsymbol{y}_m. \tag{2.3}$$

Let $\tilde{g}_{mk} = g_{mk} - \hat{g}_{mk}$ be the channel estimation error. It is well known that $\hat{g}_{mk}$ and $\tilde{g}_{mk}$ are uncorrelated and (see [41, Chapter 11])

$$\hat{g}_{mk} \sim \mathcal{CN}\left(0, \alpha_{mk}\right), \quad \tilde{g}_{mk} \sim \mathcal{CN}\left(0, \beta_{mk} - \alpha_{mk}\right), \tag{2.4}$$

where $\alpha_{mk} = \frac{\rho_r \tau \beta_{mk}^2}{1+\rho_r \tau \beta_{mk}}$.

In the following sections we analyze two main linear precoding schemes in downlink transmission: conjugate beamforming precoding and zero-forcing precoding. As mentioned in section 2.1, throughout this chapter we consider the max-min optimization problems.

2.3 Conjugate Beamforming

In this section, we consider CB precoding combined with transmit power optimization.

Conjugate Beamforming with Power Optimization

- AP m estimates $\beta_{mk},\ k = 1, \cdots, K$ and sends them to the NC.
- NC computes power coefficients $\eta_{mk},\ \forall m,\ \forall k$ (defined later in this section) as a function of large scale fading coefficients (β_{mk}) and sends them to corresponding APs.
- Users synchronously transmit pilot sequences $\boldsymbol{\psi}_i, i = 1, \cdots, K$.
- AP m gets estimates $\hat{g}_{mk},\ k = 1, \cdots, K$.

- With conjugate beamforming precoding, the m-th AP transmits the signal

$$x_m = \sqrt{\rho_f} \sum_{i=1}^{K} \sqrt{\eta_{mi}} \hat{g}_{mi}^* s_i, \tag{2.5}$$

where s_i is *data signal* intended to user i, with $\mathbb{E}\left(|s_i|^2\right) = 1$, and ρ_f is the *transmit power limit* of each AP. The quantity η_{mi} is the *power coefficient* used by AP m for transmission to user i. By optimizing coefficients η_{mi} we hope to significantly increase the system performance.

The signal received by the k-th user is

$$y_k = \sum_{m=1}^{M} g_{mk} x_m + w_k, \tag{2.6}$$

where $w_k \sim \mathcal{CN}(0, 1)$ is additive noise. We assume that user k is only aware of the statistics of the estimated channel coefficients $\mathbb{E}\left(|\hat{g}_{mk}|^2\right) = \alpha_{mk},\ \forall m$, which is a result of channel hardening in massive MIMO systems [15]. A general capacity lower bound for massive MIMO systems has been derived in [35] and a more specific bound for cell-free systems is given in [39]. With our notations, the downlink achievable rate of user k for CB is $R_k = \log_2(1 + \text{SINR}_k)$, where

$$\text{SINR}_k = \frac{\rho_f \left(\sum_{m=1}^{M} \sqrt{\eta_{mk}} \alpha_{mk} \right)^2}{1 + \rho_f \sum_{i=1}^{K} \sum_{m=1}^{M} \eta_{mi} \beta_{mk} \alpha_{mi}}. \tag{2.7}$$

For the sake of completeness, we present a proof of this bound in appendix 2.9.A.

Note that the achievable rate using the SINR expression in (2.7) is obtained under the assumption that users are only aware of statistics of channel coefficients. In [39, Figure 2] it is shown that in massive MIMO systems the achievable rate (R_k) obtained by this assumption is close to the achievable rate in the case where the users know the instantaneous channel gain. Also note that the achievable rate using (2.7) is a function of only the large scale fading coefficients and not of the small scale fading coefficients. Therefore, for conducting transmit power

optimization it is enough for APs to transmit (by backhaul network) only coefficients β_{mk} to the NC. Using these coefficients, the NC finds optimal, or near optimal, power coefficients η_{mk} and conveys them to the corresponding APs. Note that coefficients β_{mk} do not depend on OFDM subcarrier index and change slowly (about 40 times slower than small scale fading coefficients [42]). Thus, the overall needed backhaul traffic is quite small. In a wide-band system the SINR has no frequency dependence, and power coefficients are independent of frequency as well.

In the following subsections we present optimal and suboptimal power optimization algorithms.

2.3.1 Optimal Power Allocation

We would like to find power coefficients $\eta_{mk},\ \forall m,\ \forall k$, that maximize the minimum $\text{SINR}_k,\ \forall k$, under the constraint that the transmit power of each AP is limited by ρ_f . Using (2.5), we obtain that the expected transmitted power of the m-th AP is equal to

$$\mathbb{E}\left(|x_m|^2\right) = \rho_f \sum_{i=1}^{K} \eta_{mi}\alpha_{mi}.$$

Denoting $\boldsymbol{\eta} = (\eta_{mk} : m = 1, \cdots, M,\ k = 1, \cdots, K)$, we formulate the following max-min power allocation problem:

$$\max_{\boldsymbol{\eta}} \min_{k}\ \text{SINR}_k(\boldsymbol{\eta}) = \frac{\rho_f \left(\sum_{m=1}^{M} \sqrt{\eta_{mk}}\alpha_{mk}\right)^2}{1 + \rho_f \sum_{i=1}^{K} \sum_{m=1}^{M} \eta_{mi}\beta_{mk}\alpha_{mi}} \tag{2.8a}$$

$$\text{s.t.} \sum_{i=1}^{K} \eta_{mi}\alpha_{mi} \leq 1, \quad m = 1, \cdots, M. \tag{2.8b}$$

The problem in (2.8) is quasiconcave (see [39, Proposition 1]), which allows us to use the bisection method (see [43, chapter 4.2.5]). In order to employ the bisection method, we

first formulate the following equivalent problem

$$\max_{\boldsymbol{\eta},t} \; t$$

$$\text{s.t. } \text{SINR}_k(\boldsymbol{\eta}) \geq t, \quad k = 1, \cdots, K, \text{ and} \tag{2.9a}$$

$$\sum_{i=1}^{K} \eta_{mi}\alpha_{mi} \leq 1, \quad m = 1, \cdots, M. \tag{2.9b}$$

Since $\text{SINR}_k(\boldsymbol{\eta})$ is quasiconcave, $\text{SINR}_k(\boldsymbol{\eta}) \geq t,\ k = 1, \cdots, K$, is a convex set. Thus, for a fixed t, the domain of constraints in (2.9) is convex and it is easy to determine whether a given t is feasible or not. Hence we can apply the bisection method as follows. First, we choose an interval (t_l, t_u) that contains the optimal value t^{opt}. Next, we check the feasibility of the midpoint $t = \frac{(t_l+t_u)}{2}$. If t is feasible, the search interval is updated to (t, t_u), otherwise it is updated to (t_l, t). We continue until the search interval is small enough.

The complexity of the bisection method is quite high. In next three subsections we consider algorithms for suboptimal power allocations with significantly smaller complexities.

2.3.2 Full Power Transmission

A simpler power control strategy, is to permit every AP to transmit with full power, i.e., $\sum_{i=1}^{K} \eta_{mi}\alpha_{mi} = 1$. In this case the denominator of SINR in (2.7) becomes a constant (not a function of η_{mk}) and the max-min power allocation problem can be formulated as follows

$$\max_{\boldsymbol{\eta}} \min_{k} \text{SINR}_k(\boldsymbol{\eta}) = \frac{\rho_f \left(\sum_{m=1}^{M} \sqrt{\eta_{mk}}\alpha_{mk} \right)^2}{1 + \rho_f \sum_{m=1}^{M} \beta_{mk}} \tag{2.10a}$$

$$\text{s.t. } \sum_{i=1}^{K} \eta_{mi}\alpha_{mi} = 1, \quad m = 1, \cdots, M. \tag{2.10b}$$

Proposition 2.1. *The objective function* $\min_k \text{SINR}_k(\boldsymbol{\eta})$ *in* (2.10a) *is a concave function of* $\boldsymbol{\eta}$ *and the problem (2.10) is convex.*

Proof. See Appendix 2.9.B. □

Numerical algorithms for solving convex optimization problems are well known, e.g., see [43], and have significantly smaller complexity than the bisection method.

Rather surprisingly, our numerical results presented in Figure 3.1 in section 2.6 show that the full power transmission results in more than 100% degradation in terms of 5%-outage rate compared to the optimal max-min power allocation.[1] This indicates that the optimal max-min power control requires that some of the APs transmit at less than full power.

To explain why the optimal $\min_k \text{SINR}_k$ in (2.8) is not achieved with full power, we point out that the problem is quadratic and hence, the optimal point is likely in the interior of the admissible set. This is in contrast to linear optimization, where the optimal value is always achieved on the boundary, i.e., full power transmission.

This observation prompted us to check how the optimal transmit power of AP m (at the optimal solution of (2.8)) formulated by

$$p_m^{\text{opt}} = \rho_f \sum_{i=1}^{K} \eta_{mi} \alpha_{mi},$$

depends on the maximal large scale fading coefficient $\beta_m^{\max} = \max_k \beta_{mk}$ associated with this AP. In Figure 2.3 in section 2.6, we present the scatter plot of p_m^{opt} versus $\beta_m^{\max}$. One can observe that if the number of APs (M) is large, the optimal powers of *most* of the APs are quite small and that p_m^{opt} and $\beta_m^{\max}$ are strongly correlated.

Motivated by this nontrivial observation, we propose below two heuristic algorithms that significantly improve the performance of full power transmission and have moderate complexity.

[1] Parameters of the numerical experiments are presented in full detail in section 2.6.

2.3.3 Heuristic Fixed-Power Transmission

With heuristic fixed-power transmission we assume that AP m transmits with fixed total power $p_m(\beta_m^{\max}) = e^{(-\lambda\beta_m^{\max})}$, where the exponent factor λ is chosen to fit the best exponential function to the power scatter plot in Figure 2.3 for a given M and K. Note that both the power scatter plot and λ are obtained offline "once and forever" and therefore they do not contribute to the complexity of the power optimization.

With this assumption, the max-min power allocation is given by

$$\max_{\boldsymbol{\eta}} \min_{k} \text{SINR}_k(\boldsymbol{\eta}) = \frac{\rho_f \left(\sum_{m=1}^{M} \sqrt{\eta_{mk}} \alpha_{mk} \right)^2}{1 + \rho_f \sum_{m=1}^{M} p_m \beta_{mk}} \tag{2.11a}$$

$$\text{s.t.} \sum_{i=1}^{K} \eta_{mi} \alpha_{mi} = p_m, \quad m = 1, \cdots, M. \tag{2.11b}$$

This algorithm has the same online complexity as the optimization problem in (2.10). However we get a significant, about 140%, improvement in terms of 5%-outage rate (see Figure 3.1).

2.3.4 Heuristic Uniform Power Coefficients

Now we would like to drastically reduce the complexity of finding power coefficients η_{mk} without loosing much of the performance.

We again assume that AP m transmits with fixed power $p_m(\beta_m^{\max}) = e^{(-\lambda\beta_m^{\max})}$ and that power coefficients are only functions of m. Thus, η_{mk} is the same for all users, i.e., $\eta_{mk} = \eta_m, \ \forall k$, and the power coefficients can be calculated directly from the per antenna power constraints (2.11b) as

$$\eta_m = \frac{p_m \left(\beta_m^{\max} \right)}{\sum_{i=1}^{K} \alpha_{mi}}. \tag{2.12}$$

From online computational complexity point of view, the optimal power allocation algorithm in section 2.3.1 is the heaviest amongst all methods. The full power transmission

and the heuristic fixed-power transmission in sections 2.3.2 and 2.3.3 have similar complexity, while for large M the performance of the heuristic fixed-power transmission is closer to that of the optimal power allocation algorithm (see section 2.6). At each step of the optimal power allocation algorithm, we need to solve the convex feasibility problem (2.9). Whereas in the heuristic fixed-power transmission, we only need to solve one convex problem (2.11). Moreover, the constraints in (2.11b) are active, which can reduce the complexity of the search in the convex optimization problem. The heuristic uniform power control (2.12) is the simplest method with almost zero online complexity. In addition, this power control can be performed at each AP independently rather than at the NC, and therefore *does not require a backhaul link*, which could be a crucially important feature for building practical communication systems. Moreover, according to the results of section 2.6, the performance of this scheme is quite close to the performance of the scheme with heuristic fixed-power transmission. Note that the heuristic power control methods require obtaining the power scatter plot and the fitting of the exponential function which are done offline only one time.

Remark 2.2. *Note that the exponential behavior of power scatter plot in Figure 2.3 holds only for large M. Therefore, the performance of the heuristic power control methods is closer to the optimal one if M is considerably larger than K, which is the case in massive MIMO systems.*

2.3.5 Access Point Selection Scheme

In practical cell-free systems, each user will be served by only a portion of APs. In numerical results for optimal max-min power allocation algorithm, we show that in fact this is the case and only a fraction of APs serve each user in cell-free systems (see Figure 2.8). In this section, we propose a power allocation algorithm for CB precoder to further reduce the

number of APs that serve each user in cell-free massive MIMO. We achieve this by imposing an ℓ_1-norm penalty of the square root of power coefficients in max-min power allocation and enforcing the power coefficients to be sparse. Let t^{opt} be the optimal $\min_k$ SINR$_k$ at the optimal solution of (2.9). For a constant parameter $0 < c < 1$, we formulate the max-min optimization problem by adding a sparsity penalty to the objective function as follows

$$\min_{\boldsymbol{\eta}} \quad \sum_{i=1}^{K} \sum_{m=1}^{M} \sqrt{\eta_{mi}} \tag{2.13a}$$

$$\text{s.t. } \text{SINR}_k(\boldsymbol{\eta}) \geq c \times t^{\text{opt}}, \quad k = 1, \cdots, K, \text{ and} \tag{2.13b}$$

$$\sum_{i=1}^{K} \eta_{mi} \alpha_{mi} \leq 1, \quad m = 1, \cdots, M.$$

We can turn problem (2.13) into a convex optimization problem by a simple change of variable $\varkappa_{mi} = \sqrt{\eta_{mi}}$. Note that in order to solve problem (2.13), one needs to first obtain the optimal $\min_k$ SINR$_k$ in (2.9) that is denoted by t^{opt}. The ℓ_1-norm penalty in (2.13), simply enforces the power coefficients to be sparse. In other words, η_{mk} will be close to zero for most values of m meaning that the corresponding APs will not be transmitting to the k-th user. Therefore, the effective number of APs serving each user will be smaller than M. At the optimal solution of (2.13), SINR of all users will be equal to $c \times t^{\text{opt}} < t^{\text{opt}}$, where c is a predetermined constant that controls the sparsity level of the power coefficients. When c is small, the power coefficients will be more sparse. However, it comes at the expense of lower achievable rates (see Figure 2.8).

2.4 Zero-Forcing

In this section, we investigate the downlink performance of a cell-free system with ZF precoder.

Zero-Forcing Precoder with Power Optimization

- AP m estimates $\beta_{mk},\ k = 1, \cdots, K$ and sends them to the NC.
- Users synchronously transmit pilot sequences $\boldsymbol{\psi}_i, i = 1, \cdots, K$.
- AP m gets estimates $\hat{g}_{mk},\ k = 1, \cdots, K$ and sends them to the NC.
- NC computes power coefficients as a function of large scale fading coefficients (β_{mk}).
- NC forms ZF precoding vectors using channel estimates ($\hat{g}_{mk}$) and power coefficients (η_{mk}), and then sends precoding vectors to the corresponding APs.

The idea behind ZF linear precoder is that signal transmitted to a user does not create interference to other users [8]. Usually ZF precoder is defined by the pseudo inverse matrix $A = \hat{G}^* \left(\hat{G}^T \hat{G}^*\right)^{-1}$, where $[\hat{G}]_{mk} = \hat{g}_{mk}$ is an estimate of channel matrix G. With this A we have $\hat{G}^T A = I_K$, i.e., the interference is completely canceled. In [44], it was pointed out that this ZF precoder is in general suboptimal for the power control problem subject to per antenna power constraints and finding an optimal precoder involves numerical algorithms. However, for simplicity, we will use the above ZF precoding matrix. Let $H \in \mathbb{R}^{M \times K}$ be a matrix with entries $[H]_{mk} = \sqrt{\eta_{mk}}$. We define precoding matrix for the cell-free power allocation problem as $A_{ZF} = A \odot H$, where $\odot$ is the Hadamard or entry-wise product. Elements of A_{ZF} are

$$[A_{ZF}]_{mk} = a_{mk}^{ZF} = \sqrt{\eta_{mk}} a_{mk}, \quad m = 1, \cdots, M,\ k = 1, \cdots, K. \tag{2.14}$$

In order for $\hat{G}^T A_{ZF}$ to be diagonal, to avoid interference, it is necessary to have $\eta_{1k} = \cdots = \eta_{Mk}$ for any k. Thus power coefficients should be only functions of k, i.e., $\eta_{mk} = \eta_k$, and therefore

$$A_{ZF} = \hat{G}^* (\hat{G}^T \hat{G}^*)^{-1} P, \tag{2.15}$$

where P is a diagonal matrix with $\sqrt{\eta_1}, \ldots, \sqrt{\eta_K}$ on its diagonal. The k-th user receives

$$\begin{aligned} y_k &= \sqrt{\rho_f}\boldsymbol{g}_k^T A_{ZF}\boldsymbol{s} + w_k = \sqrt{\rho_f}(\hat{\boldsymbol{g}}_k + \tilde{\boldsymbol{g}}_k)^T \hat{G}^*(\hat{G}^T\hat{G}^*)^{-1}P\boldsymbol{s} + w_k \\ &= \underbrace{\sqrt{\rho_f\eta_k}s_k}_{J_0} + \underbrace{\sqrt{\rho_f}\tilde{\boldsymbol{g}}_k^T\hat{G}^*(\hat{G}^T\hat{G}^*)^{-1}P\boldsymbol{s}}_{J_1} + w_k, \end{aligned} \tag{2.16}$$

where $\boldsymbol{g}_k = (g_{1k}, \ldots, g_{Mk})^T$, $\boldsymbol{s} = (s_1, \ldots, s_K)^T$, J_0 is the signal part, and J_1 results from the channel estimation error. The achievable rate expression with ZF precoder is presented below.

Theorem 2.1. *An achievable rate of user k with ZF precoder is* $\log_2(1 + \text{SINR}_{k,ZF})$*, with*

$$\text{SINR}_{k,ZF} = \frac{\rho_f\eta_k}{1 + \rho_f\sum_{i=1}^K \eta_i\gamma_{ki}}, \tag{2.17}$$

where γ_{ki} is the i-th element of the following vector

$$\boldsymbol{\gamma}_k = \textit{diag}\left\{\mathbb{E}\left((\hat{G}^T\hat{G}^*)^{-1}\hat{G}^T\mathbb{E}\left(\tilde{\boldsymbol{g}}_k^*\tilde{\boldsymbol{g}}_k^T\right)\hat{G}^*(\hat{G}^T\hat{G}^*)^{-1}\right)\right\}, \tag{2.18}$$

and $\mathbb{E}\left(\tilde{\boldsymbol{g}}_k^*\tilde{\boldsymbol{g}}_k^T\right)$ is a diagonal matrix with $(\beta_{mk} - \alpha_{mk})$ on its m-th diagonal element.

Proof. See Appendix 2.9.C. □

2.4.1 Optimal Power Allocation

Let $\boldsymbol{a}_{[1]}^T, \cdots, \boldsymbol{a}_{[M]}^T$ and $\hat{\boldsymbol{g}}_{[1]}^T, \cdots, \hat{\boldsymbol{g}}_{[M]}^T$ be the rows of matrices A_{ZF} and $\hat{G}$ respectively. We define vector

$$\boldsymbol{\delta}_m = \text{diag}\left\{\mathbb{E}\Big((\hat{G}^T\hat{G}^*)^{-1}\hat{\boldsymbol{g}}_{[m]}\hat{\boldsymbol{g}}_{[m]}^H(\hat{G}^T\hat{G}^*)^{-1}\Big)\right\}. \tag{2.19}$$

The transmitted power from AP m is given by

$$\begin{aligned} \rho_f\mathbb{E}\left(\boldsymbol{a}_{[m]}^T\boldsymbol{a}_{[m]}^*\right) &= \rho_f\mathbb{E}\Big(\left(\hat{\boldsymbol{g}}_{[m]}^H(\hat{G}^T\hat{G}^*)^{-1}P\right)\left(\hat{\boldsymbol{g}}_{[m]}^H(\hat{G}^T\hat{G}^*)^{-1}P\right)^H\Big) \\ &= \rho_f\,\text{trace}\left(P^2\mathbb{E}\Big((\hat{G}^T\hat{G}^*)^{-1}\hat{\boldsymbol{g}}_{[m]}\hat{\boldsymbol{g}}_{[m]}^H(\hat{G}^T\hat{G}^*)^{-1}\Big)\right) \\ &= \rho_f\sum_{i=1}^K \eta_i\delta_{mi}. \end{aligned} \tag{2.20}$$

We formulated the max-min power allocation problem with per antenna power constraints as follows

$$\max_{\boldsymbol{\eta}} \min_{k} \text{SINR}_{k,ZF}(\boldsymbol{\eta}) = \frac{\rho_f \eta_k}{1 + \rho_f \sum_{i=1}^{K} \eta_i \gamma_{ki}} \tag{2.21a}$$

$$\text{s.t.} \sum_{i=1}^{K} \eta_i \delta_{mi} \leq 1, \quad m = 1, ..., M. \tag{2.21b}$$

As we mentioned earlier, the power coefficients must be functions of k only in order to cancel the interference to other users. Despite this fact, the transmitted power from AP m to user k, $\rho_f |a_{mk}^{ZF}|^2 = \rho_f \eta_k |a_{mk}|^2$, is function of both m and k.

The numerator and denominator of $\text{SINR}_{k,ZF}(\boldsymbol{\eta})$ in (2.21a) are linear functions of $\boldsymbol{\eta}$. Thus $\text{SINR}_{k,ZF}(\boldsymbol{\eta})$ is a quasilinear function and we can again use the bisection method to solve (2.21). The feasibility problem at each step of bisection method can be expressed as

$$\text{find} \ \ \boldsymbol{\eta}$$

$$\text{s.t. } \text{SINR}_{k,ZF}(\boldsymbol{\eta}) \geq t, \quad k = 1, \cdots, K, \tag{2.22a}$$

$$\sum_{i=1}^{K} \eta_i \delta_{mi} \leq 1, \quad m = 1, ..., M. \tag{2.22b}$$

The following Lemma is very useful since it allows us to significantly reduce the complexity of problem (2.22) and further problems (2.21), (2.32), (2.37), and (2.38). Despite the claim in Lemma 2.1 being simple, the proof turned out to be nontrivial (see Appendix 2.9.D).

Lemma 2.1. *Suppose $\boldsymbol{\eta}$ and $\boldsymbol{\eta}' > 0$ are two vectors such that*

$$\text{SINR}_{k,ZF}(\boldsymbol{\eta}) = t \quad \textit{and} \quad \text{SINR}_{k,ZF}(\boldsymbol{\eta}') \geq t, \quad \textit{for } k = 1, \cdots, K.$$

Then we have $0 \leq \boldsymbol{\eta} \leq \boldsymbol{\eta}'$, where the vector inequality is applied element-wise.

Proof. See Appendix 2.9.D. □

Corollary 2.1. *If* $\text{SINR}_{k,ZF}(\boldsymbol{\eta}) = t,\ \forall k$*, doesn't have a feasible solution satisfying constraints in (2.22b), then* $\text{SINR}_{k,ZF}(\boldsymbol{\eta}) \geq t,\ \forall k$*, doesn't have a feasible solution either.*

Proof. See Appendix 2.9.D. □

According to Corollary 2.1, to check the feasibility of $\text{SINR}_{k,ZF}(\boldsymbol{\eta}) \geq t$, it suffices to only check the feasibility of $\text{SINR}_{k,ZF}(\boldsymbol{\eta}) = t$, which is equivalent to solving a system of K linear equations. This allows us to avoid solving the convex feasibility problem (2.22).

Note that $\boldsymbol{\gamma}_k$ in (2.18) and $\boldsymbol{\delta}_m$ in (2.19) are functions of only the large scale fading coefficients and the expectation is taken over small scale fading coefficients. The NC can estimate them in two ways:

1. Every time the large scale fading coefficients change, NC generates multiple dummy random variables $\hat{g}_{mk}^{\text{dummy}} \sim \mathcal{CN}(0, \alpha_{mk})$ that have the same statistics as the channel coefficients defined in (2.4). These random variables are then used to compute $\boldsymbol{\gamma}_k$ and $\boldsymbol{\delta}_m$.

2. An alternative way is to update $\boldsymbol{\delta}_m$ using exponential smoothing. Denote the current estimate by $\hat{\boldsymbol{\delta}}_m$. Every time the NC obtains new estimates of channel coefficients $\hat{G}$, it updates $\hat{\boldsymbol{\delta}}_m = \alpha\hat{\boldsymbol{\delta}}_m + (1-\alpha)\boldsymbol{\delta}_{m,\text{curr}}$, where $\boldsymbol{\delta}_{m,\text{curr}} = \text{diag}\left\{(\hat{G}^T\hat{G}^*)^{-1}\hat{\boldsymbol{g}}_{[m]}\hat{\boldsymbol{g}}_{[m]}^H(\hat{G}^T\hat{G}^*)^{-1}\right\}$ and $0 < \alpha < 1$ is a constant. Parameter $\boldsymbol{\gamma}_k$ can be estimated using a similar procedure.

Note that if we use method 1 to compute $\boldsymbol{\gamma}_k$ and $\boldsymbol{\delta}_m$, the NC needs to perform the optimal power control algorithm in (2.21) only when the large scale fading coefficients change.

2.4.2 Low Complexity Power Allocation Algorithm

The computational complexity of power allocation can be further reduced by obtaining an accurate approximation of the optimal solution of (2.21) instead of finding the true solution. When the channel estimation error is zero, i.e., $\tilde{g}_{mk} = 0$, the max-min power control for ZF

precoder has a closed form solution which we use to obtain simple heuristic power coefficients ($\boldsymbol{\eta}^{\text{apx}}$) as an approximate solution to problem (2.21). If we assume that $\tilde{g}_{mk}$ is zero, then the optimization problem takes the form

$$\max_{\boldsymbol{\eta}^{\text{pc}}} \min_{k} \text{SINR}_{k,ZF}^{\text{pc}}(\boldsymbol{\eta}^{\text{pc}}) = \rho_f \eta_k^{\text{pc}} \tag{2.23a}$$

$$\text{s.t.} \sum_{i=1}^{K} \eta_i^{\text{pc}} \delta_{mi}^{\text{pc}} \leq 1, \quad m = 1, ..., M, \tag{2.23b}$$

where the superscript pc denotes *perfect CSI* and δ_{mi}^{pc} is the i-th element of the following vector

$$\boldsymbol{\delta}_m^{\text{pc}} = \text{diag}\left\{\mathbb{E}\Big((G^T G^*)^{-1} \boldsymbol{g}_{[m]} \boldsymbol{g}_{[m]}^H (G^T G^*)^{-1}\Big)\right\}, \tag{2.24}$$

which is equivalent to the vector $\boldsymbol{\delta}_m$ in (2.19) with $\tilde{g}_{mk} = 0$.

Lemma 2.2. *Suppose $\boldsymbol{\eta}^{\text{pc}}$ is the optimal solution of the optimization problem (2.23), then there exists $m' \in \{1, \cdots, M\}$ such that AP m' transmits with full power, i.e., $\sum_{i=1}^{K} \eta_i^{\text{pc}} \delta_{m'i}^{\text{pc}} = 1$. Furthermore, the optimal power coefficients are given by $\eta_i^{\text{pc}} = 1/\sum_{i=1}^{K} \delta_{m'i}^{\text{pc}}$, $\forall\, i = 1, \cdots, K$.*

Proof. See Appendix 2.9.E. □

This Lemma prompts us to use the heuristic solution

$$\eta_1^{\text{apx}} = \cdots = \eta_K^{\text{apx}} = 1/\left(\max_{m} \sum_{i=1}^{K} \delta_{mi}\right), \tag{2.25}$$

as an approximation of the optimal solution of (2.21). Note that δ_{mi} in (2.25) is defined in (2.19) and contains the effect of channel estimation error.

Results presented in Figures 3.1 and 3.2 in section 2.6, show that the rates obtained by this approximation, i.e., $\log_2\left(1 + \text{SINR}_{k,ZF}\left(\boldsymbol{\eta}^{\text{apx}}\right)\right)$, are virtually optimal, while the computational complexity of (2.25) is drastically smaller than the one of (2.21).

An intuitive explanation of the virtual optimality of (2.25) is the following. Lemma 2.2 is based on the assumption that channel estimation error is zero, i.e., $\tilde{g}_{mk} = 0$. In real life

scenario with enough uplink pilot transmit power, we are close to this regime. According to (2.4), $\text{Var}(\tilde{g}_{mk})/\text{Var}(\hat{g}_{mk}) = 1/\rho_r\tau\beta_{mk}$. If AP m is close to user k, then the above ratio is typically small, which can be interpreted as that $\tilde{g}_{mk}$ is negligible. The APs that are far from user k may have larger estimation error, however, their contribution in transmission to user k is not significant, and hence the estimation error does not degrade the performance.

The following proposition rigorously shows that with enough uplink power, approximate power control results in near-optimal SINR values.

Proposition 2.2. *Let $\boldsymbol{\eta}^{\text{opt}}$ and $\boldsymbol{\eta}^{\text{apx}}$ respectively denote the optimal solution of the max-min power control problem (2.21) and the power coefficients defined by (2.25). For every $\epsilon > 0$ there exists a threshold $\rho_r(\epsilon)$ such that whenever the uplink pilot transmit power $\rho_r > \rho_r(\epsilon)$,*

$$\left\|\text{SINR}_{k,ZF}\left(\boldsymbol{\eta}^{\text{opt}}\right) - \text{SINR}_{k,ZF}\left(\boldsymbol{\eta}^{\text{apx}}\right)\right\| < \epsilon, \quad k = 1, \cdots, K. \tag{2.26}$$

Proof. See Appendix 2.9.F. □

Note that since all norms, e.g., 1-norm, Frobenius norm, and infinity-norm, on a finite-dimensional Banach space are equivalent, any norm can be used in Proposition 2.2.

We now evaluate the complexity of computing power coefficients $\boldsymbol{\eta}^{\text{apx}}$ in (2.25), which is equivalent to computing $\boldsymbol{\delta}_m, \ \forall m$. The dominant factor in computing $\boldsymbol{\delta}_m$ in both methods discussed in section 2.4.1, is computing $\boldsymbol{\delta}_{m,\text{curr}} = \text{diag}\left\{(\hat{G}^T\hat{G}^*)^{-1}\hat{\boldsymbol{g}}_{[m]}\hat{\boldsymbol{g}}_{[m]}^H(\hat{G}^T\hat{G}^*)^{-1}\right\}$ (either using true channel estimates or dummy random variables). Calculation of $\boldsymbol{\delta}_{m,\text{curr}}$ consists of multiplying matrices of size $K \times M$ and $M \times K$, a matrix inversion of size $K \times K$, multiplication of matrix of size $K \times K$ by vector of size $K \times 1$ for M times, and MK multiplications. Thus, under the assumption of $M \gg K$, the dominant factor is of order $\mathcal{O}(MK^2)$, which is the same as the computational complexity of ZF precoding matrix $A = \hat{G}^*(\hat{G}^T\hat{G}^*)^{-1}$. Therefore, the power control method defined in (2.25) is not the limiting factor in terms of complexity.

2.5 Small-Cell

Small-cell systems are often considered as a candidate for future generations of wireless systems. Challenges and potentials of small-cell systems are addressed, for example, in [45], [39]. Below we estimate the performance of small-cell systems and compare it with the performance of cell-free systems. We use the same description of small-cell systems as in [39].

We assume that M APs and K users are randomly located within the same area as in the case of cell-free systems. Each AP serves at most one user at a time and $M - K$ access points remain silent as shown in Figure 2.2. This can be viewed as M small cells, each equipped with a single antenna AP. Optimal assignment of K out of M access points to K users is a hard combinatorial problem. A simple greedy solution is to assign to user k the AP m_k with the largest slow fading coefficient ($\beta_{m_k k}$) among available APs, i.e.,

$$m_k = \underset{n\text{: AP } n \text{ is available}}{\operatorname{argmax}} \beta_{nk}. \tag{2.27}$$

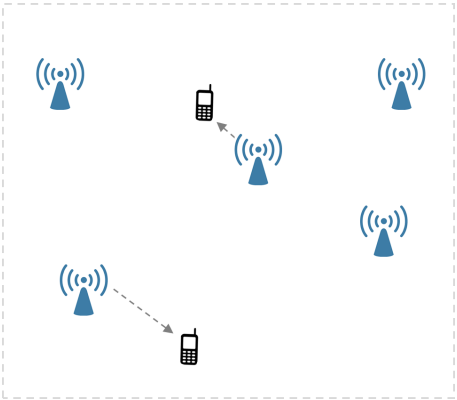

Figure 2.2: Small-cell system. Each user is served by one AP.

Since each user is served by only one AP, channel does not harden and we assume that user k knows $\hat{g}_{m_k,k}$ either via a downlink training sequence or error-free feedback from AP

m_k. AP m_k does not conduct any precoding, but simply transmits data symbol s_k and user k receives

$$y_k = \sqrt{\rho_f \eta_k}\, g_{m_k,k}\, s_k + \sum_{i=1, i\neq k}^{K} \sqrt{\rho_f \eta_i}\, g_{m_i,k}\, s_i + w_k, \tag{2.28}$$

where $w_k \sim \mathcal{CN}(0,1)$ is additive Gaussian noise and $\eta_k \in [0,1]$ is power coefficient of AP m_k. User k uses MF receiver, i.e., $\hat{s}_k = \frac{\hat{g}^*_{m_k,k}}{|\hat{g}_{m_k,k}|} y_k$, to recover data signal s_k. Achievable rate of the k-th user with MF is given by [39]

$$\mathbb{E}\left(\log_2\left(1 + \text{SINR}_k\right)\right) = \frac{1}{\ln 2} e^{\mu_k} \text{Ei}(\mu_k), \tag{2.29}$$

where

$$\text{SINR}_k = \frac{\rho_f\, \eta_k |\hat{g}_{m_k,k}|^2}{1 + \rho_f \sum_{i=1}^{K} \eta_i\, \beta_{m_i,k} - \rho_f \eta_k \alpha_{m_k,k}}, \tag{2.30}$$

and

$$\mu_k = \frac{1 + \rho_f \sum_{i=1}^{K} \eta_i\, \beta_{m_i,k} - \rho_f \eta_k \alpha_{m_k,k}}{\rho_f\, \eta_k \alpha_{m_k,k}}, \tag{2.31}$$

and $\text{Ei}(x) = \int_x^\infty \frac{e^{-t}}{t} \mathrm{d}t$ is the exponential integral. Similar to the cell-free system, we also apply the max-min power allocation with per antenna power constraints

$$\begin{aligned} &\max_{\boldsymbol{\eta}} \min_{k} \text{SINR}_k(\boldsymbol{\eta}) \\ &\qquad \text{s.t. } 0 \leq \eta_k \leq 1, \quad k = 1, \cdots, K. \end{aligned} \tag{2.32}$$

Similar to (2.21) this optimization problem is quasiconcave and therefore can be solved by using the bisection method. A lemma and corollary similar to Lemma 2.1 and Corollary 2.1 can be proved for the above optimization problem. Therefore for the feasibility check it is enough to solve a system of K linear equations.

Simulation results in section 2.6 show that ZF precoder in cell-free systems significantly outperforms CB precoder. Hence it is natural to ask whether small-cell systems with

ZF precoder would have better performance compared to MF receiver. Below we formulate the corresponding power allocation problem for ZF precoder in small-cell systems and answer this question negatively.

We assume that K selected APs denoted by $m_1, \cdots, m_K$, are connected via a backhaul network and jointly form the precoding matrix A to serve K users. Denote by $\hat{G} = [\hat{g}_{m_i,k}]_{1\leq i,k\leq K}$ the estimated channel matrix between all selected APs and all users in the small-cell system. With full rank assumption of $\hat{G}$, the precoding matrix is $A = \left(\hat{G}^T\right)^{-1} P$, where P is a diagonal matrix with $\sqrt{\eta_1}, \cdots, \sqrt{\eta_K}$ on its diagonal. The received signal at the k-th user is given by

$$y_k = \underbrace{\sqrt{\rho_f \eta_k} s_k}_{J_0} + \underbrace{\sqrt{\rho_f} \tilde{\boldsymbol{g}}_k^T (\hat{G}^T)^{-1} P \boldsymbol{s}}_{J_1} + w_k, \tag{2.33}$$

where $\tilde{\boldsymbol{g}}_k = [\tilde{g}_{m_1,k}, \cdots, \tilde{g}_{m_K,k}]^T$ is the estimation error of channel coefficients between all selected APs and user k and $w_k \sim \mathcal{CN}(0,1)$ is additive Gaussian noise.

Theorem 2.2. *An achievable rate of the k-th user in small-cell scheme with ZF precoding is* $\log(1 + \text{SINR}_{k,ZF})$*, where* $\text{SINR}_{k,ZF}$ *is given by*

$$\text{SINR}_{k,ZF} = \frac{\rho_f \eta_k}{1 + \rho_f \sum_{i=1}^{K} \eta_i \nu_{ki}}, \tag{2.34}$$

where ν_{ki} is the i-th element of the following vector

$$\boldsymbol{\nu}_k = \mathit{diag}\left\{ \mathbb{E}\left((\hat{G}^*)^{-1} \mathbb{E}\left(\tilde{\boldsymbol{g}}_k^* \tilde{\boldsymbol{g}}_k^T \right) (\hat{G}^T)^{-1} \right) \right\}, \tag{2.35}$$

and $\mathbb{E}\left(\tilde{\boldsymbol{g}}_k^* \tilde{\boldsymbol{g}}_k^T \right)$ is a diagonal matrix with $\beta_{m_i,k} - \frac{\rho_r \tau \beta_{m_i,k}^2}{1+\rho_r \tau \beta_{m_i,k}}$ on its i-th diagonal element.

Proof. See Appendix 2.9.G. □

Let $\boldsymbol{\alpha}_{[m_1]}^T, \cdots, \boldsymbol{\alpha}_{[m_K]}^T$ be the rows of matrix $\left(\hat{G}^T\right)^{-1}$ and define vector $\boldsymbol{\sigma}_{m_k}$ by

$$\boldsymbol{\sigma}_{m_k} = \text{diag}\left\{ \boldsymbol{\alpha}_{[m_k]}^* \boldsymbol{\alpha}_{[m_k]}^T \right\}. \tag{2.36}$$

Then the instantaneous transmitted power from AP m_k is $\rho_f \sum_{i=1}^{K} \eta_i \sigma_{m_k,i}$. The power allocation problem with ZF precoder and *instantaneous* transmit power constraints can be expressed as

$$\max_{\boldsymbol{\eta}} \min_{k} \text{SINR}_{k,ZF}(\boldsymbol{\eta}) = \frac{\rho_f \eta_k}{1 + \rho_f \sum_{i=1}^{K} \eta_i \nu_{ki}}$$
$$\text{s.t.} \sum_{i=1}^{K} \eta_i \sigma_{m_k,i} \leq 1, \quad k = 1, ..., K. \tag{2.37}$$

Let $\boldsymbol{\zeta}_{m_k} = \mathbb{E}(\boldsymbol{\sigma}_{m_k})$, where the expectation is taken over small scale fading coefficients. The max-min power allocation problem with *average* per antenna power constraints is given by

$$\max_{\boldsymbol{\eta}} \min_{k} \text{SINR}_{k,ZF}(\boldsymbol{\eta}) = \frac{\rho_f \eta_k}{1 + \rho_f \sum_{i=1}^{K} \eta_i \nu_{ki}}$$
$$\text{s.t.} \sum_{i=1}^{K} \eta_i \zeta_{m_k,i} \leq 1, \quad k = 1, ..., K. \tag{2.38}$$

Both problems in (2.37) and (2.38) are quasilinear and similar to (2.21) bisection method can be used to solve them. Corollary 2.1 reduces the complexity by solving a system of K linear equations to check the feasibility of the problem instead of solving an optimization problem.

2.6 Numerical Results

We consider a square dense urban area of 2×2 km^2 which is wrapped around to avoid boundary effects. We also consider M randomly placed APs and K randomly placed users. For large scale fading coefficients we use the COST Hata model

$$10 \log_{10}(\beta_{mk}) = -136 - 35 \log_{10}(d_{mk}) + X_{mk}, \tag{2.39}$$

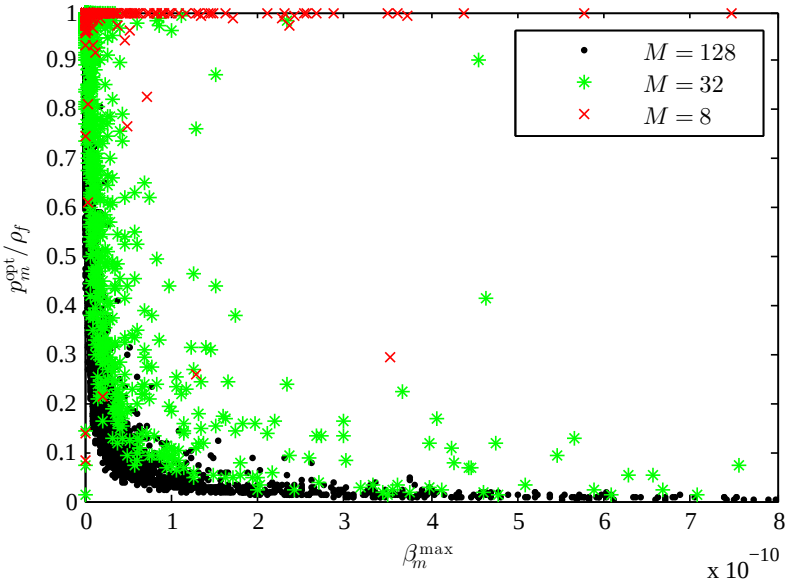

Figure 2.3: Scatter plot of p_m^{opt}/ρ_f versus $\beta_m^{\max}$ of all APs for CB under optimal power allocation with different number of APs and $K = 4$.

where d_{mk} is the distance between AP m and user k in kilometers and $X_{mk} \sim \mathcal{N}(0, \sigma^2_{\text{shad}})$ with $\sigma_{\text{shad}} = 8$ dB. The noise variance at the receiver is assumed to be $\sigma_w^2 = 290 \times \kappa \times B \times NF$, where κ, B, and NF are Boltzmann constant, bandwidth (20 MHz), and noise figure (9 dB) respectively. The maximum transmitted power of each AP antenna and user, ρ_f and ρ_r, is 200 mW, unless stated otherwise. Length of all pilot sequences is $\tau = K$.

Experiment 1: We would like to show that in the case of CB the max-min optimization (2.8) requires that most of APs must transmit with low powers as we indicated in section 2.3.2.

Let $p_m^{\text{opt}} = \rho_f \sum_{i=1}^{K} \eta_{mi}\alpha_{mi}$ denote the optimal transmit power of AP m defined by (2.8). By $\beta_m^{\max}$ we denote the largest slow fading coefficient between AP m and users, i.e, $\beta_m^{\max} = \max_k \beta_{mk}$. Figure 2.3, shows the scatter plot of p_m^{opt}/ρ_f versus $\beta_m^{\max}$ for $m = 1, \cdots, M$, over 100 iterations for three values of M, and $K = 4$. It can be observed that for $M = 8$, almost all APs transmit with full power which shows that when M is small, the full power transmission scheme given in (2.10), whose complexity is relatively small, is near

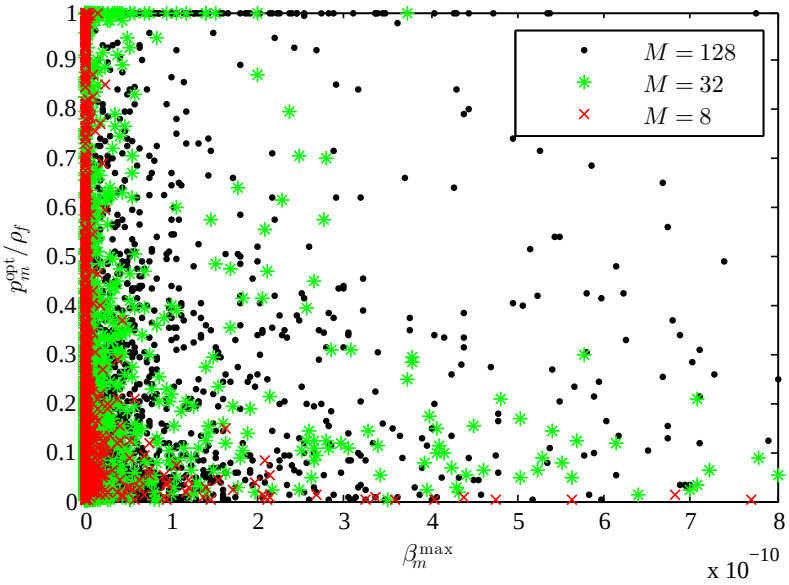

Figure 2.4: Scatter plot of p_m^{opt}/ρ_f versus $\beta_m^{\max}$ of all APs for ZF under optimal power allocation with different number of APs and $K = 4$.

optimal. However, for larger values of M, the APs that have large $\beta_m^{\max}$ should transmit with low power. Typically a large $\beta_m^{\max}$ indicates that the distance between AP m and at least one of the users is small. So we conclude that in systems with a large number of APs (M), the APs that are closer to users usually transmit with smaller power. Comparing cases $M = 32$ and $M = 128$ we see that this behavior becomes more and more pronounced as M grows. We conduct the same experiment for ZF precoding. The results are presented in Figure 2.4. Note that for ZF precoding, we have also observed that optimal power control strategy does not use full power for most of the APs.

Experiment 2: In this experiment, we compare achievable per-user rates of cell-free and small-cell systems with CB and ZF precoding and the power optimization algorithms considered in the previous sections. For small-cell system we consider two possible scenarios. In the first scenario, we assume that each AP antenna has the same maximum transmit power as in the cell-free system, i.e., 200mW. Thus, the hardware cost of APs for both small-cell and

cell-free systems is about the same. In this scenario, however, the total transmit power limit of cell-free system, in which all M APs are active, is larger than small-cell system, when only K APs are active. For this reason, we also consider the second scenario in which the maximum transmit power of each small-cell AP is equal to $M/K \cdot 200$mW. In this case the total transmit power limit for both systems is about the same. However, in the latter case more powerful amplifiers are needed for each AP antenna in the small-cell system. Also note that according to Figure 2.3 in the cell-free system, the total transmit power of CB with optimal and heuristic power allocation algorithms is much less than the total power limit when M is large.

The empirical CDFs of per-user rates for all scenarios with $M = 128$ and $K = 16$ are plotted in Figure 3.1. The horizontal line corresponds to 5%-outage rates. The curve "Small-Cell MF Full Power" corresponds to the small-cell system with equal power coefficients ($\eta_1 = \cdots = \eta_K = 1$), i.e., each AP transmits at full power. The CDF "Cell-Free ZF Low Complexity Power Allocation" (dashed red) corresponds to $R_k^{\text{subopt}} = \log_2(1 + \text{SINR}_{k,ZF}(\boldsymbol{\eta}^{\text{apx}}))$, where $\boldsymbol{\eta}^{\text{apx}}$ is the approximate power coefficients defined in (2.25) and $\text{SINR}_{k,ZF}$ is given in (2.17). The curve "Cell-Free ZF Optimal Power Allocation" (solid red) corresponds to $R_k^{\text{opt}} = \log_2(1 + \text{SINR}_{k,ZF}(\boldsymbol{\eta}))$, where $\boldsymbol{\eta}$ is the optimal solution of (2.21).

First, we note that the performance of cell-free ZF system with low complexity approximate power allocation defined in (2.25) is virtually indistinguishable from the system with optimal powers, which coincides with the results of Proposition 2.2. Second, we see that the cell-free system with ZF precoder significantly outperforms cell-free system with CB, and all small-cell systems. Third, the small-cell system with MF and full power transmission provides superior 5%-outage rate compared to the small-cell system with ZF precoder and max-min power allocation. Finally, we note that in the small-cell system, ZF precoding with instantaneous power constraints significantly outperforms ZF precoding with average power constraints.

Next, we observe that cell-free systems are visibly better than small-cell systems. The optimal power allocation in cell-free CB and cell-free ZF systems respectively give 5.36 and 10.02 fold improvements in terms of 5%-outage rate over the small-cell system with full power transmission ($\eta_1 = \cdots = \eta_K = 1$) and maximum antenna transmit power 200mW. In the scenario when the maximum antenna transmit power of small-cell system is equal $128/16 \cdot 200$mW, we get 5.26 and 9.83 fold improvements respectively.

Furthermore, it is important to note that 5%-outage rate of cell-free CB with optimal max-min power allocation is close to 5%-outage rates of the cell-free with heuristic power allocations.

It is also interesting to note that small-cell with max-min power allocation produces inferior performance in terms of 5%-outage rate compared to the small-cell system with full power transmission. This indicates that max-min optimization is not the near optimal tool to maximize 5%-outage rates in small-cell system.

Figure 3.2 presents similar results but for worst rate (minimum rate) among K users for given realization of large-scale coefficients β_{mk}. One can see that again cell-free systems significantly outperform small-cell systems. Note that with max-min power control, rates of all users are the same for each realization.

Experiment 3: We further investigate the performance of small-cell systems in which the total number of APs (M) is reduced but their individual transmit power is increased. The results are shown in Figure 3.3. In all curves $K = 32$. One can see that as we decrease the number of APs from 256 to 128 and further to 64 the performance of small-cell system in terms of 5%-outage rate significantly degrades despite the fact that the maximum transmit power is growing by a factor of 4 and 8 respectively. In the small-cell scheme, each user selects the largest of a multiplicity of (heavy-tailed log-normal) large scale fading coefficients. Therefore, the small-cell system benefits from super-diversity gain for large values of M. This fact also

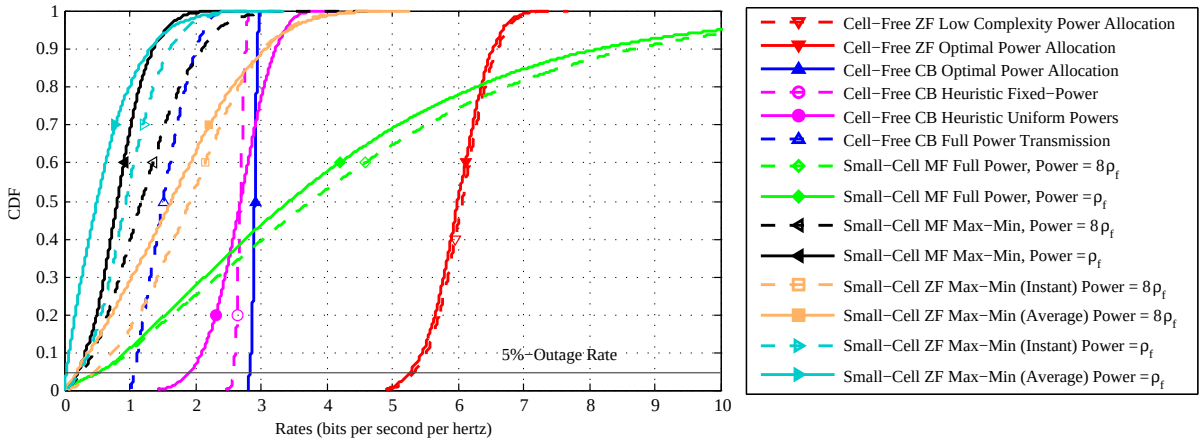

Figure 2.5: CDFs of the achievable per-user rates for cell-free and small-cell schemes under different power allocations with $M = 128$ and $K = 16$.

explains why in small-cell scheme some users may have rates several times greater than the typical rate of users in cell-free system, as evident by studying Figure 3.3.

Experiment 4: In this experiment, we study the effective number of APs that are contributing in transmission to a particular user. The *effective number of APs* serving users is defined as the minimum number of APs that provide 95% of the total transmitted power to a particular user. Figure 2.8 shows the CDF of the effective number of APs serving each user with CB and ZF precoders using the optimal max-min power allocation and the AP selection algorithm in Section 2.3.5 for $M = 128$ and $K = 16$. In problem (2.13), we set $c = \frac{(1+t^{\text{opt}})^{\alpha}-1}{t^{\text{opt}}}$, where t^{opt} is the optimal $\min_k$ SINR$_k$ in (2.9), and $0 \leq \alpha \leq 1$ is the ratio between achievable rate obtained by problems (2.13) and (2.9). For example, with $\alpha = 0.9$, the achievable rate obtained by problem (2.13) is 90% of the optimal achievable rate obtained by (2.9). In Figure 2.8, the effective number of APs obtained by problem (2.13) is plotted for different values of α. One can see that only a fraction of APs are contributing in transmission to a user. It can be observed that the effective number of APs transmitting to a particular user is much smaller with the power allocation algorithm (2.13). More specifically, with only 1% smaller achiev-

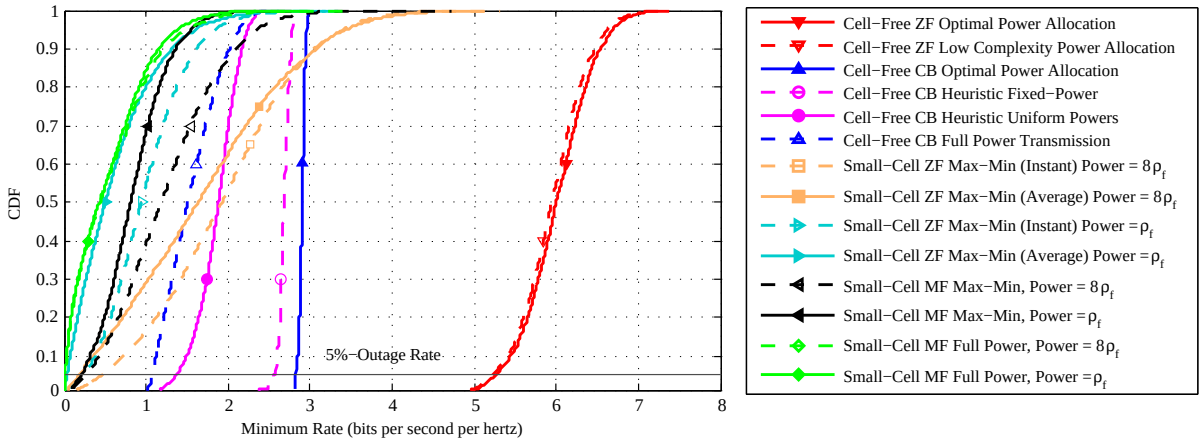

Figure 2.6: CDFs of the worst achievable rate for cell-free and small-cell schemes under different power allocations with $M = 128$ and $K = 16$.

able rate ($\alpha = 0.99$), the median of effective number of APs obtained by (2.13) is about 36% less than that of the the problem (2.9). Figure 2.9 shows the CDF of per-user achievable rates obtained by power allocation algorithm (2.13) for different values of α. Note that the number of APs needed for achieving a good performance depends on many factors, such as channel model, density of APs and users, correlation between channel coefficients, etc.

Experiment 5: In this experiment, we compare performance of cell-free and small-cell systems as the number of APs and users increase with constant ratio. Table 2.1 contains per-user 5%-outage and median rates of cell-free system with max-min optimal power allocation and small-cell system with full power transmission. The ratio between number of APs and users is 8 in all cases. We observe that the 5%-outage rate of the small-cell scheme with full power transmission remains almost unchanged as the number of APs and users increase (with constant ratio). On the other hand, such an increase in the network size directly translates to a superior performance for the cell-free scheme with CB and ZF precoders.

Finally, we note that CDFs of the cell-free system with the optimal power allocation in all figures are nearly vertical. This means that the performance of the system stays almost the

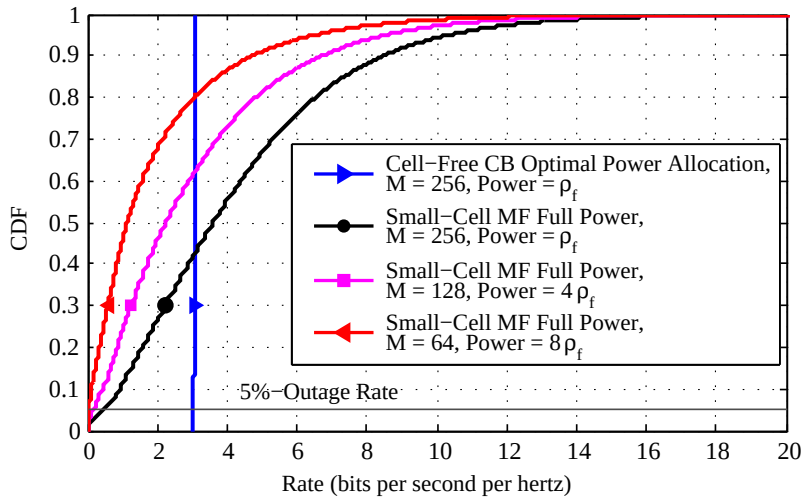

Figure 2.7: CDFs of the achievable per-user rates for cell-free and small-cell schemes with different M and transmit powers.

same for different channel realizations, which is a desired feature for wireless communication systems.

2.7 Conclusion

We studied downlink performance of cell-free massive MIMO systems in terms of the minimum rate among all users and 5%-outage rate. In cell-free massive MIMO, all distributed APs serve all users simultaneously providing uniformly good services to all users. We considered power optimization algorithms using max-min criterion for cell-free massive MIMO and small-cell systems with CB and ZF precoders. For cell-free systems with CB, we proposed low complexity heuristic power allocation algorithms. Numerical results indicate that these heuristic algorithms have only moderate loss in terms of 5%-outage rate, but have significantly lower computational complexity compared with the optimal power allocation. We proposed a technique to reduce the effective number of APs that serve a user for CB precoder by employing an ℓ_1-norm sparsity penalty in the max-min power optimization. For cell-free systems with

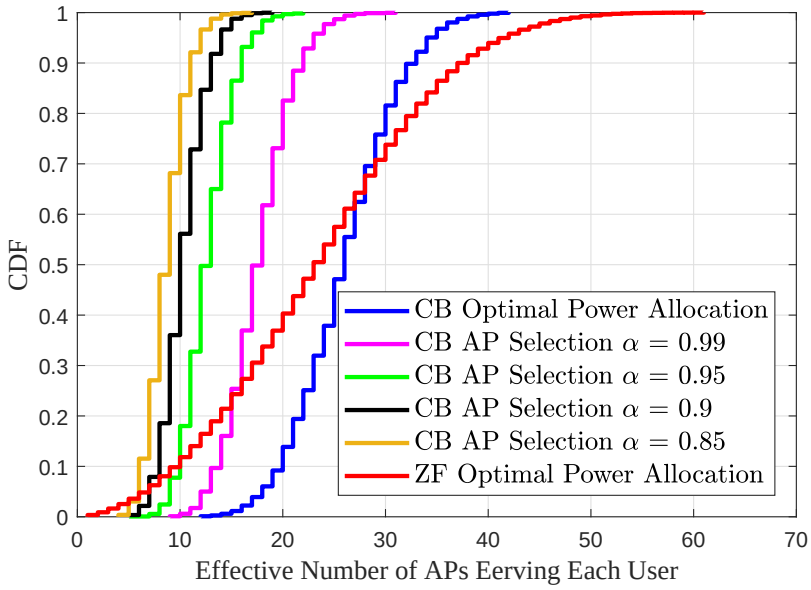

Figure 2.8: CDF of effective number of APs serving each user for cell-free CB and ZF precoders with $M = 128$ and $K = 16$.

ZF precoder, we proposed methods for finding optimal and suboptimal powers. The method for finding suboptimal powers is based on a perfect CSI model; it has very small complexity and its performance happened to be virtually optimal. We compared our results with a small-cell scheme in which each user is served by a single AP. Even though, in a small-cell scheme users are aware of CSI and power allocation algorithms are applied more frequently (about 40 times more than in a cell-free system), the simulation results show that we can achieve higher rates (in terms of 5%-outage and minimum rate) with the cell-free massive MIMO scheme.

2.8 Acknowledgment

This chapter contains material as it appears in [1], E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, “Precoding and Power Optimization in Cell-Free Massive MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4445-4459, July 2017, and is also based on the material as it appears in [2], E. Nayebi, A.

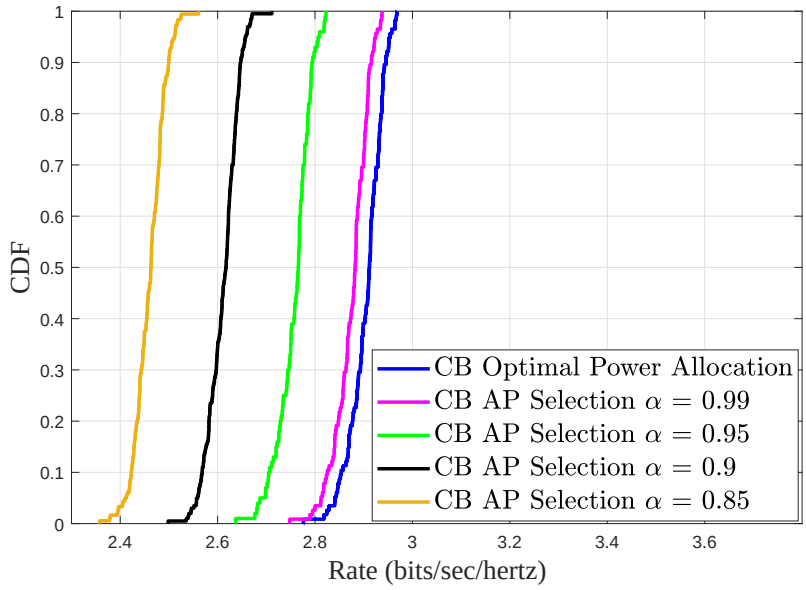

Figure 2.9: CDF of the achievable per-user rates for cell-free CB with AP selection algorithm with $M = 128$ and $K = 16$.

Ashikhmin, T. L. Marzetta, and H. Yang, "Cell-Free Massive MIMO Systems," in *Proc. 49th Asilomar Conference on Signals, Systems and Computers*, Nov. 2015, pp. 695-699. The dissertation author was the primary investigator and author of these papers.

2.9 Appendices

2.9.A Derivation of the Capacity Lower Bound

Proof. The received signal in (2.6) can be written as

$$y_k = \underbrace{\sum_{m=1}^{M} \sqrt{\rho_f \eta_{mk}} \mathbb{E}\left(|\hat{g}_{mk}|^2\right) s_k}_{T_0\text{: useful signal}} + \underbrace{\sum_{m=1}^{M} \sum_{i \neq k}^{K} \sqrt{\rho_f \eta_{mi}} \hat{g}_{mk} \hat{g}_{mi}^* s_i}_{T_1\text{: interference}}$$
$$+ \underbrace{\sum_{m=1}^{M} \sqrt{\rho_f \eta_{mk}} \left(|\hat{g}_{mk}|^2 - \mathbb{E}\left(|\hat{g}_{mk}|^2\right)\right) s_k}_{T_2\text{: lack of channel knowledge at user}} + \underbrace{\sum_{m=1}^{M} \sum_{i=1}^{K} \sqrt{\rho_f \eta_{mi}} \tilde{g}_{mk} \hat{g}_{mi}^* s_i}_{T_3\text{: channel estimation error}} + w_k. \quad (2.40)$$

Table 2.1: 5%-outage and median per-user rates of cell-free and small-cell schemes with different number of APs and users.

	Small-Cell MF Full Power			Cell-Free CB Optimal Power Allocation			Cell-Free ZF Optimal Power Allocation		
	$M=64$ $K=8$	$M=128$ $K=16$	$M=256$ $K=32$	$M=64$ $K=8$	$M=128$ $K=16$	$M=256$ $K=32$	$M=64$ $K=8$	$M=128$ $K=16$	$M=256$ $K=32$
5%-outage	0.5	0.51	0.51	2.38	2.82	3.05	3.16	5.28	7.17
Median	3.02	3.47	3.61	2.56	2.91	3.07	4.23	6.00	7.74

Since data signals intended to different users are uncorrelated and white additive noise is independent from data symbols and the channel coefficients, it is easy to check that terms T_0, T_1, T_2, T_3 and w_k are mutually uncorrelated. According to the results from [46], the worst case noise for mutual information is Gaussian additive noise with the variance equal to the variance of $T_1+T_2+T_3+w_k$. Hence the achievable rate is lower bounded by $\log_2(1+\text{SINR}_k)$, where

$$\text{SINR}_k = \frac{\mathbb{E}\left(|T_0|^2\right)}{1+\mathbb{E}\left(|T_1|^2+|T_2|^2+|T_3|^2\right)}. \tag{2.41}$$

Variances of T_0, T_1, T_2, and T_3 can be computed as

$$\begin{aligned}
&\mathbb{E}(|T_0|^2) = \rho_f \left(\sum_{m=1}^{M} \sqrt{\eta_{mk}}\alpha_{mk}\right)^2, \qquad \mathbb{E}(|T_1|^2) = \rho_f \sum_{i\neq k}^{K}\sum_{m=1}^{M} \eta_{mi}\alpha_{mk}\alpha_{mi},\\
&\mathbb{E}(|T_2|^2) = \rho_f \sum_{m=1}^{M} \eta_{mk}\alpha_{mk}^2, \qquad\qquad \mathbb{E}\left(|T_3|^2\right) = \rho_f \sum_{i=1}^{K}\sum_{m=1}^{M} \eta_{mi}\left(\beta_{mk}-\alpha_{mk}\right)\alpha_{mi}.
\end{aligned} \tag{2.42}$$

By substituting the variances in (2.41) we complete the proof. □

2.9.B Proof of Proposition 2.1

Proof. Define $c_{mk} = \frac{\alpha_{mk}}{\sqrt{1/\rho_f + \sum_{m=1}^{M} \beta_{mk}}}$. For $0 \leq \theta \leq 1$ we have

$$\begin{aligned}
&\text{SINR}_k(\theta\boldsymbol{\eta} + (1-\theta)\boldsymbol{\eta'}) = \left(\sum_{m=1}^{M} c_{mk}\sqrt{\theta\eta_{mk} + (1-\theta)\eta'_{mk}}\right)^2 \\
&= \sum_{m=1}^{M}\sum_{n=1}^{M} c_{mk}c_{nk}\left[\theta^2\eta_{mk}\eta_{nk} + (1-\theta)^2\eta'_{mk}\eta'_{nk} + \theta(1-\theta)(\eta_{mk}\eta'_{nk} + \eta_{nk}\eta'_{mk})\right]^{\frac{1}{2}} \\
&\overset{(a)}{\geq} \sum_{m=1}^{M}\sum_{n=1}^{M} c_{mk}c_{nk}\left[\left(\theta\sqrt{\eta_{mk}\eta_{nk}} + (1-\theta)\sqrt{\eta'_{mk}\eta'_{nk}}\right)^2\right]^{\frac{1}{2}} \\
&= \theta\text{SINR}_k(\boldsymbol{\eta}) + (1-\theta)\text{SINR}_k(\boldsymbol{\eta'}),
\end{aligned} \tag{2.43}$$

where (a) follows from the fact that Arithmetic mean is larger than the Geometric mean, i.e., $\eta_{mk}\eta'_{nk} + \eta_{nk}\eta'_{mk} \geq \sqrt{\eta_{mk}\eta'_{nk}\eta_{nk}\eta'_{mk}}$. The above inequality implies that $\text{SINR}_k(\boldsymbol{\eta})$ is a concave function of $\boldsymbol{\eta}$. Since the minimum of concave functions is a concave function, the objective function (2.10a) is concave and constraints (2.10b) are linear. Thus, the problem (2.10) is convex. □

2.9.C Proof of Theorem 2.1

Proof. Since data symbols, additive noise, and channel coefficients are mutually independent, it is easy to show that terms J_0, J_1, and w_k are mutually uncorrelated. Based on the worst-case uncorrelated additive noise [46], the achievable rate is lower bounded by $\log_2(1 + \text{SINR}_{k,ZF})$, where

$$\text{SINR}_{k,ZF} = \frac{\mathbb{E}\left(|J_0|^2\right)}{1 + \mathbb{E}\left(|J_1|^2\right)}. \tag{2.44}$$

Variances of J_0 and J_1 can be computed as $\mathbb{E}\left(|J_0|^2\right) = \rho_f \eta_k$ and

$$\begin{aligned}\mathbb{E}\left(|J_1|^2\right) &= \rho_f \mathbb{E}\left(\left|\tilde{\boldsymbol{g}}_k^T \hat{G}^* (\hat{G}^T \hat{G}^*)^{-1} P \boldsymbol{s}\right|^2\right) \\ &= \rho_f \,\text{trace}\Big\{P^2 \mathbb{E}\Big((\hat{G}^T \hat{G}^*)^{-1} \hat{G}^T \mathbb{E}\left(\tilde{\boldsymbol{g}}_k^* \tilde{\boldsymbol{g}}_k^T\right) \hat{G}^* (\hat{G}^T \hat{G}^*)^{-1}\Big)\Big\} \\ &= \rho_f \sum_{i=1}^{K} \eta_i \gamma_{ki}, \end{aligned} \tag{2.45}$$

where γ_{ki} is defined in Theorem 2.1. □

2.9.D Proof of Lemma 2.1

Before starting proof of Lemma 2.1, we remind the definition of M-matrices [47]: Matrix $A \in \mathbb{R}^{K\times K}$ is an M-matrix if it can be expressed in the form $A = sI - B$, where $B = (b_{ij})$ with $b_{ij} \geq 0$, for all $1 \leq i, j \leq K$, and s is greater than the maximum of the moduli of the eigenvalues of B, and I is an identity matrix.

Proof of Lemma 2.1 . Suppose $\boldsymbol{\eta}$ is the solution of $\text{SINR}_{k,ZF}(\boldsymbol{\eta}) = t,\ k = 1, \cdots, K$ and vector $\boldsymbol{\eta}' > 0$ satisfies $\text{SINR}_{k,ZF}(\boldsymbol{\eta}') \geq t,\ k = 1, \cdots, K$. Denote the maximum $\text{SINR}_{k,ZF}(\boldsymbol{\eta}')$ by $\text{SINR}_{j,ZF} = \max_i \text{SINR}_{i,ZF}(\boldsymbol{\eta}')$. It is clear that $\text{SINR}_{j,ZF}$ is an increasing function of η_j' and a decreasing function of η_i' for $i \neq j$. If we reduce η_j' then $\text{SINR}_{j,ZF}$ decreases and $\text{SINR}_{i,ZF}(\boldsymbol{\eta}')$ increases for all $i \neq j$. By continuing this procedure, we will achieve a vector $\boldsymbol{\eta}^{(2)} > 0$ such that $\text{SINR}_{k,ZF}(\boldsymbol{\eta}^{(2)}) = t_2,\ t_2 \geq t$ and $\boldsymbol{\eta}^{(2)} \leq \boldsymbol{\eta}'$. Let $\pi_1 = 1/t$ and $\pi_2 = 1/t_2$. Define matrix $B \in \mathbb{C}^{K\times K},\ [B]_{ij} = \gamma_{ij}$ for $i \neq j$ and $[B]_{ii} = 0$, and vector $\boldsymbol{d} = \left[\frac{1}{\rho_f}, \cdots, \frac{1}{\rho_f}\right]^T \in \mathbb{R}^K$. Then we have the following system of K linear equations

$$\left(B - \pi_1 I\right)\boldsymbol{\eta} + \boldsymbol{d} = \left(B - \pi_2 I\right)\boldsymbol{\eta}^{(2)} + \boldsymbol{d} = 0. \tag{2.46}$$

Since B is a non-negative matrix, from Perron-Frobenius theorem we know that there exists a left eigenvector $\boldsymbol{v}$ of B with corresponding eigenvalue r such that all components of $\boldsymbol{v}$

are non-negative and r is Perron-Frobenius eigenvalue with the largest absolute value among all other eigenvalues of B. Since all elements of vector $\boldsymbol{d}$ are positive, it follows from (2.46) that $B\boldsymbol{\eta}^{(2)} < \pi_2\boldsymbol{\eta}^{(2)}$. If we multiply both sides of $B\boldsymbol{\eta}^{(2)} < \pi_2\boldsymbol{\eta}^{(2)}$ by non-negative vector $\boldsymbol{v}$, we get

$$\boldsymbol{v}^T B\boldsymbol{\eta}^{(2)} = r\boldsymbol{v}^T\boldsymbol{\eta}^{(2)} < \pi_2\boldsymbol{v}^T\boldsymbol{\eta}^{(2)}. \tag{2.47}$$

Since $\boldsymbol{v}^T\boldsymbol{\eta}^{(2)}$ is positive, we can conclude that $r < \pi_2$. Since $\pi_1 \geq \pi_2$, from above equation we have $r < \pi_2 \leq \pi_1$, which yields that $(\pi_1 I - B)$ and $(\pi_2 I - B)$ are M-matrices. One of the properties of an M-matrix is that it is invertible and its inverse is a non-negative matrix. Therefore all entries of $\boldsymbol{\eta} = (\pi_1 I - B)^{-1}\boldsymbol{d} \geq 0$ are non negative. After adding $\pi_2\boldsymbol{\eta}$ to the both sides in (2.46) and rearranging we get

$$\boldsymbol{\eta}^{(2)} = \boldsymbol{\eta} + (\pi_2 I - B)^{-1}(\pi_1 - \pi_2)\boldsymbol{\eta}, \tag{2.48}$$

and since all elements of matrix $(\pi_2 I - B)^{-1}$ and vector $(\pi_1 - \pi_2)\boldsymbol{\eta}$ are non negative, we can conclude that $0 \leq \boldsymbol{\eta} \leq \boldsymbol{\eta}^{(2)} \leq \boldsymbol{\eta}'$ which completes the proof. □

Proof of Corollary 2.1. Suppose $\boldsymbol{\eta}^{(1)}$ is the solution of $\text{SINR}_{k,ZF}(\boldsymbol{\eta}^{(1)}) = t,\ k = 1, \cdots, K$, which doesn't satisfy feasibility constraints in (2.22b) and let $\boldsymbol{\eta}^{(2)}$ be any vector that satisfies $\text{SINR}_{k,ZF}(\boldsymbol{\eta}^{(2)}) \geq t,\ k = 1, \cdots, K$. From Lemma 2.1 we know $\boldsymbol{\eta}^{(1)} \geq 0$. Hence, the infeasibility of $\boldsymbol{\eta}^{(1)}$ means that there exists $m' \in \{1, \cdots, M\}$ such that $\sum_{i=1}^{K} \eta_i^{(1)}\delta_{m'i} \geq 1$. Since $\boldsymbol{\eta}^{(2)} \geq \boldsymbol{\eta}^{(1)}$ (see Lemma 2.1), then $\boldsymbol{\eta}^{(2)}$ is also infeasible, *i.e*,

$$1 < \sum_{i=1}^{K} \eta_i^{(1)}\delta_{m'i} \leq \sum_{i=1}^{K} \eta_i^{(2)}\delta_{m'i}. \tag{2.49}$$

□

2.9.E Proof of Lemma 2.2

Proof. Let $\boldsymbol{\eta}^{\text{pc}}$ be the optimal solution of the optimization problem (2.23). Denote the optimal power transmitted from AP m by

$$p_m = \sum_{i=1}^{K} \eta_i^{\text{pc}} \delta_{mi}^{\text{pc}}, \quad m = 1, \cdots, M. \tag{2.50}$$

It is clear that there exists $m' \in \{1, \cdots, M\}$ such that $p_{m'} = 1$, otherwise one can replace η_i with $\eta_i / \max_m p_m, \ \forall i$, which would improve the optimal value while still satisfying the problem constraints. Next we define

$$\eta_l^{\text{pc}} = \min_k \eta_k^{\text{pc}} \leq \max_k \eta_k^{\text{pc}} = \eta_u^{\text{pc}}. \tag{2.51}$$

Now assume $\eta_l^{\text{pc}} < \eta_u^{\text{pc}}$. In the following we show the existence of a new set of feasible power coefficients denoted by $\tilde{\boldsymbol{\eta}}^{\text{pc}}$ such that $\min_k \text{SINR}_{k,ZF}^{\text{pc}}(\tilde{\boldsymbol{\eta}}^{\text{pc}}) > \min_k \text{SINR}_{k,ZF}^{\text{pc}}(\boldsymbol{\eta}^{\text{pc}})$, which is a contradiction to the optimality assumption. First we assume η_l^{pc} is the unique minimizer. Define $c = \delta_{m'l}^{\text{pc}} / \delta_{m'u}^{\text{pc}}$ and $\epsilon = \min \left\{ \frac{\eta_u^{\text{pc}} - \eta_l^{\text{pc}}}{1+c}, \ \zeta \right\}$, where

$$\zeta = \min_m \left\{ \frac{1 - p_m}{\delta_{ml}^{\text{pc}} - c\, \delta_{mu}^{\text{pc}}} \,\middle|\, \delta_{ml}^{\text{pc}} - c\, \delta_{mu}^{\text{pc}} > 0, p_m \neq 1 \right\}. \tag{2.52}$$

The new power coefficients given by

$$\tilde{\eta}_k^{\text{pc}} = \begin{cases} \eta_l^{\text{pc}} + \epsilon, & k = l, \\ \eta_u^{\text{pc}} - c \times \epsilon, & k = u, \\ \eta_k^{\text{pc}}, & \text{otherwise}, \end{cases} \tag{2.53}$$

yield in new set of feasible transmitted power constraints, which can be written as

$$\begin{aligned} \sum_{i=1}^{K} \tilde{\eta}_i^{\text{pc}} \delta_{mi}^{\text{pc}} &= p_m + \epsilon(\delta_{ml}^{\text{pc}} - c\, \delta_{mu}^{\text{pc}}) \\ &\leq \begin{cases} p_m, & \delta_{ml}^{\text{pc}} - c\, \delta_{mu}^{\text{pc}} \leq 0, \\ 1, & \delta_{ml}^{\text{pc}} - c\, \delta_{mu}^{\text{pc}} > 0, \end{cases} \end{aligned} \tag{2.54}$$

From the definition of ϵ it follows that

$$\tilde{\eta}_l^{\text{pc}} = \eta_l^{\text{pc}} + \epsilon \leq \eta_l^{\text{pc}} + \frac{\eta_u^{\text{pc}} - \eta_l^{\text{pc}}}{1+c} = \eta_u^{\text{pc}} - c\,\frac{\eta_u^{\text{pc}} - \eta_l^{\text{pc}}}{1+c} \leq \eta_u^{\text{pc}} - c\epsilon = \tilde{\eta}_u^{\text{pc}}. \tag{2.55}$$

Therefore, $\min_k \text{SINR}_{k,ZF}^{\text{pc}}(\tilde{\boldsymbol{\eta}}^{\text{pc}}) = \rho_f \tilde{\eta}_l^{\text{pc}} > \rho_f \eta_l^{\text{pc}} = \min_k \text{SINR}_{k,ZF}^{\text{pc}}(\boldsymbol{\eta}^{\text{pc}})$, which is a contradiction. This completes the proof for $\eta_1^{\text{pc}} = \cdots = \eta_K^{\text{pc}}$, which immediately translates to

$$\eta_1^{\text{pc}} = \cdots = \eta_K^{\text{pc}} = 1/\left(\sum_{i=1} \delta_{m'i}^{\text{pc}}\right). \tag{2.56}$$

Note that if η_l^{pc} is not the unique minimizer, similar proof follows by applying the same procedure for multiple times until all such η_l^{pc}s are replaced with a larger value. □

2.9.F Proof of Proposition 2.2

Proof. Let $\Gamma = \left[\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_K\right]$ and $\Delta = \left[\boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_M\right]$ be matrices containing vectors defined in (2.18) and (2.19) respectively. By Lemma 2.1, $\boldsymbol{\eta}^{\text{opt}}(\Delta, \Gamma)$ is a continuous function of Δ and Γ, i.e., for every $e > 0$ there exists $e_1, e_2 > 0$ such that

$$\|\boldsymbol{\eta}^{\text{opt}}(\Delta, \Gamma) - \boldsymbol{\eta}^{\text{opt}}(\Delta_0, \Gamma_0)\| < \frac{e}{2}, \tag{2.57}$$

for all matrices Δ and Γ where $\|\Delta - \Delta_0\| < e_1$ and $\|\Gamma - \Gamma_0\| < e_2$. The power coefficients defined in (2.25) are also continuous functions of Δ, i.e., for every $e > 0$ there exists an $e_3 > 0$ such that

$$\|\boldsymbol{\eta}^{\text{apx}}(\Delta) - \boldsymbol{\eta}^{\text{apx}}(\Delta_0)\| < \frac{e}{2}, \tag{2.58}$$

for all Δ with $\|\Delta - \Delta_0\| < e_3$. Define matrix $\Delta^{\text{pc}} = \left[\boldsymbol{\delta}_1^{\text{pc}}, \cdots, \boldsymbol{\delta}_M^{\text{pc}}\right]$ containing vectors defined in (2.24). Let $\Gamma_0 = 0$ and $\Delta_0 = \Delta^{\text{pc}}$ in (2.57) and (2.58). For given $e_1, e_2, e_3 > 0$, we can find an uplink power ρ_r such that $\|\Delta - \Delta^{\text{pc}}\| < \min\{e_1, e_3\}$ and $\|\Gamma\| < e_2$. From (2.57)

and (2.58) we conclude that

$$\begin{aligned}\|\boldsymbol{\eta}^{\text{opt}}(\Delta,\Gamma)-\boldsymbol{\eta}^{\text{opt}}(\Delta^{\text{pc}},0)\|<\frac{e}{2},\\ \|\boldsymbol{\eta}^{\text{apx}}(\Delta)-\boldsymbol{\eta}^{\text{apx}}(\Delta^{\text{pc}})\|<\frac{e}{2}.\end{aligned} \tag{2.59}$$

From Lemma 2.2 we have $\boldsymbol{\eta}^{\text{opt}}(\Delta^{\text{pc}},0)=\boldsymbol{\eta}^{\text{apx}}(\Delta^{\text{pc}})$. So we have

$$\begin{aligned}\|\boldsymbol{\eta}^{\text{opt}}(\Delta,\Gamma)-\boldsymbol{\eta}^{\text{apx}}(\Delta)\|&=\|\boldsymbol{\eta}^{\text{opt}}(\Delta,\Gamma)-\boldsymbol{\eta}^{\text{apx}}(\Delta^{\text{pc}})+\boldsymbol{\eta}^{\text{apx}}(\Delta^{\text{pc}})-\boldsymbol{\eta}^{\text{apx}}(\Delta)\|\\ &\le\|\boldsymbol{\eta}^{\text{opt}}(\Delta,\Gamma)-\boldsymbol{\eta}^{\text{apx}}(\Delta^{\text{pc}})\|+\|\boldsymbol{\eta}^{\text{apx}}(\Delta)-\boldsymbol{\eta}^{\text{apx}}(\Delta^{\text{pc}})\|\\ &<\frac{e}{2}+\frac{e}{2}=e.\end{aligned} \tag{2.60}$$

Note that $\text{SINR}_{k,ZF}(\boldsymbol{\eta})$ is a continuous function of $\boldsymbol{\eta}$. Therefore we can conclude that for every $\epsilon>0$ there exists an uplink power $\rho_r(\epsilon)$ such that $\|\boldsymbol{\eta}^{\text{opt}}(\Delta,\Gamma)-\boldsymbol{\eta}^{\text{apx}}(\Delta)\|\le e$ implies

$$\left\|\text{SINR}_{k,ZF}(\boldsymbol{\eta}^{\text{opt}})-\text{SINR}_{k,ZF}(\boldsymbol{\eta}^{\text{apx}})\right\|<\epsilon,\quad k=1,\cdots,K. \tag{2.61}$$

□

2.9.G Proof of Theorem 2.2

Proof. It is clear that terms J_0, J_1, and w_k in (2.33) are mutually uncorrelated. Hence, from (2.33) we can obtain SINR of the k-th user as follows

$$\text{SINR}_{k,ZF}=\frac{\mathbb{E}\left(|J_1|^2\right)}{1+\mathbb{E}\left(|J_1|^2\right)} \tag{2.62}$$

where $\mathbb{E}\left(|J_1|^2\right)=\rho_f\eta_k$ and variance of J_1 is computed as

$$\begin{aligned}\mathbb{E}\left(|J_1|^2\right)&=\rho_f\,\mathbb{E}\left(\left|\tilde{\boldsymbol{g}}_k^T(\hat{G}^T)^{-1}P\boldsymbol{s}\right|^2\right)\\ &=\rho_f\,\text{trace}\left\{P^2E\left((\hat{G}^*)^{-1}E\left(\tilde{\boldsymbol{g}}_k^*\tilde{\boldsymbol{g}}_k^T\right)(\hat{G}^T)^{-1}\right)\right\}\\ &=\rho_f\sum_{i=1}^{K}\eta_i\nu_{ki},\end{aligned} \tag{2.63}$$

where ν_{ki} is defined in (2.35). □

Chapter 3

Performance of Cell-Free Massive MIMO Systems with MMSE and LSFD Receivers

Cell-free massive multiple-input multiple-output (MIMO) comprises a large number of distributed single-antenna access points (APs) serving a much smaller number of users. There is no partitioning into cells and each user is served by all APs.

In this chapter, the uplink performance of cell-free systems with minimum mean squared error (MMSE) and large scale fading decoding (LSFD) receivers is investigated. The main idea of LSFD receiver is to maximize achievable throughput using only large scale fading coefficients between APs and users. In this chapter, capacity lower bounds for MMSE and LSFD receivers are derived. An asymptotic approximation for signal-to-interference-plus-noise ratio (SINR) of MMSE receiver is derived as a function of large scale fading coefficients only. The obtained approximation is accurate even for a small number of AP antennas and users. MMSE and LSFD receivers demonstrate five-fold and two-fold gains respectively over matched filtering (MF) receiver in terms of 5%-outage rate.

3.1 Introduction

In recent years massive multiple-input multiple-output (MIMO) has attracted considerable attention as a candidate for the fifth generation physical layer technology [15], [10]. Cell-free massive MIMO is a particular deployment of massive MIMO systems with a network of randomly-located large number of single-antenna access points (APs), where the geographical area is not partitioned into cells and each user is served simultaneously by all of the APs [2], [39].

Some of the advantages and limitations of the networks with distributed APs can be found in [2,27,28,34,39]. In particular in [2], [39], the performance of downlink transmission and uplink transmission with matched filtering (MF) receiver in cell-free massive MIMO systems have been studied. In this chapter, we first consider uplink minimum mean squared error

(MMSE) receiver. We further propose a suboptimal MMSE receiver called *partial MMSE* (PMMSE) and demonstrate that it has virtually optimal performance. In [34], the authors study uplink performance of cellular massive MIMO systems with distributed antenna clusters in each cell. They consider MMSE and MF receivers with coordination among distributed antenna clusters in each cell. In contrast, we assume all distributed APs coordinate with each other to form the postcoding vectors and detect the signals transmitted by users. In [48], random matrix theory results are used to study performance of cellular massive MIMO systems. Motivated by [48], we applied random matrix theory to derive a tight approximation of the PMMSE in cell-free systems as a function of large scale fading coefficients with cooperation among distributed APs. Since the performance gap between MMSE and PMMSE receivers is negligible, our approximation is also very accurate for the optimal MMSE receiver, which is confirmed by the numerical results even for a small number of APs/users.

In [21] and [22], large scale fading decoding (LSFD), also known as pilot contamination postcoding, was proposed to reduce interference in cellular massive MIMO systems. In LSFD, base stations cooperate by using only the large scale fading coefficients. In this work, we propose generalization of the LSFD receiver for cell-free massive MIMO systems and derive its signal-to-interference-plus-noise ratio (SINR) expression as a function of large scale fading coefficients.

We further derive an expression for SINR of cell-free systems with MF receiver in the regime when the number of users is constant and the number of APs grows without limit. Our result shows that in this regime, the system performance is limited by the coherent interference resulting from two or more users sharing the same pilot sequence.

In numerical experiments, we evaluate the system performance under independent and correlated shadow fading models. Results show that MMSE and LSFD receivers provide significant gain over MF receiver. MMSE receiver outperforms LSFD receiver while the latter

has smaller complexity.

The chapter is organized as follows. Section 3.2 describes the system model and channel estimation. In section 3.3, we investigate MMSE, partial MMSE, and LSFD receivers in uplink transmission. Finally, numerical results are presented in section 3.4.

Throughout the chapter diag $(a_i)_{1\leq i\leq n}$ denotes diagonal matrix with $a_1, \cdots, a_n$ on its diagonal. If $\mathcal{S}_1 = \{\alpha_1, \cdots, \alpha_n\} \in \mathbb{N}^{n\times 1}$ and $\mathcal{S}_2 = \{\sigma_1, \cdots, \sigma_m\} \in \mathbb{N}^{m\times 1}$, where α_i and σ_is are in the increasing order, then operator $[v_i]_{i\in\mathcal{S}_1}$ denotes the column vector $[v_{\alpha_1}, \cdots, v_{\alpha_n}]^T$; and operator $[[v_{ij}]]_{i\in\mathcal{S}_1, j\in\mathcal{S}_2}$ denotes the $n \times m$ matrix $\begin{bmatrix} v_{\alpha_1\sigma_1}, & \cdots, & v_{\alpha_1\sigma_m} \\ \vdots & \ddots & \vdots \\ v_{\alpha_n\sigma_1}, & \cdots, & v_{\alpha_n\sigma_m} \end{bmatrix}$.

3.2 System Model and Channel Estimation

We consider a geographical area with M randomly distributed single-antenna APs and K single-antenna users, assuming that $K \ll M$. All APs are connected to a network controller (NC) via an unspecified backhaul network. All APs and users are perfectly synchronized in time. The channel coefficient between AP m and user k is given by

$$g_{mk} = \sqrt{\beta_{mk}} h_{mk}, \tag{3.1}$$

where β_{mk} is the large scale fading coefficient which accounts for path loss and shadow fading and $h_{mk} \sim \mathcal{CN}(0, 1)$ is the small scale fading coefficient. The large scale fading coefficients change slowly over time and are assumed to be known at the NC. The small scale fading coefficients are i.i.d. random variables that stay constant over a channel coherence interval.

We assume time-division duplex (TDD) protocol, i.e., all users synchronously send randomly assigned orthonormal pilot sequences ($\boldsymbol{\psi}_1, \cdots, \boldsymbol{\psi}_\tau \in \mathbb{C}^{\tau\times 1}$, where $\boldsymbol{\psi}_i^H \boldsymbol{\psi}_j = \delta(i - j)$) to allow APs to estimate channel coefficients, which they further send to the NC.

We consider short channel coherence interval (due to high user mobility) and therefore τ is small and $K > \tau$. Hence each pilot is reused by several users, which results in the pilot

contamination, [21], [22]. In [39], a greedy pilot assignment scheme in cell-free systems has been introduced, which is shown to improve the performance of cell-free system compared with the random pilot assignment scheme. However, for simplicity we consider the random pilot assignment in the cell-free systems.

All users are partitioned into τ sets $\mathcal{S}_1, \cdots, \mathcal{S}_\tau$ in a way that users in $\mathcal{S}_j$ use pilot $\boldsymbol{\psi}_j$. Let b_i be the index of the pilot sequence transmitted by the i-th user. The received signal in the first step of the TDD protocol at the m-th AP is given by

$$\boldsymbol{y}_m = \sqrt{\rho\tau} \sum_{i=1}^{K} g_{mi} \boldsymbol{\psi}_{b_i} + \boldsymbol{w}_m, \tag{3.2}$$

where ρ is the uplink transmit power of each user and $\boldsymbol{w}_m \in \mathbb{C}^{\tau \times 1} \sim \mathcal{CN}(0,1)$ is additive Gaussian noise. The m-th AP computes the MMSE estimate of g_{mk} as

$$\hat{g}_{mk} = \frac{\sqrt{\rho\tau}\beta_{mk}}{1 + \rho\tau \sum_{i \in \mathcal{S}_{b_k}} \beta_{mi}} \boldsymbol{\psi}_{b_k}^H \boldsymbol{y}_m. \tag{3.3}$$

It can be verified that $\hat{g}_{mk}$ and the channel estimation error $\tilde{g}_{mk} = g_{mk} - \hat{g}_{mk}$ are uncorrelated Gaussian random variables with distributions

$$\hat{g}_{mk} \sim \mathcal{CN}\left(0, \alpha_{mk}\right), \quad \tilde{g}_{mk} \sim \mathcal{CN}\left(0, \beta_{mk} - \alpha_{mk}\right), \tag{3.4}$$

where $\alpha_{mk} = \frac{\rho\tau\beta_{mk}^2}{1+\rho\tau \sum_{i \in \mathcal{S}_{b_k}} \beta_{mi}}$. Note that $\hat{g}_{mi} = \frac{\beta_{mi}}{\beta_{mk}} \hat{g}_{mk}$ for every $i, k \in \mathcal{S}_{b_k}$. Therefore, it is enough for AP m to choose one user $u_j \in \mathcal{S}_j$ and send only the channel estimates $\hat{g}_{mu_j},\ j = 1, \cdots, \tau$ to the NC.

Let η_i denote the power coefficient used by the i-th user to transmit uplink data. For notation convenience we define

$$\begin{aligned} A_i &\triangleq \text{diag}\left(\alpha_{mi}\right)_{1 \le m \le M}, \quad & B_i &\triangleq \text{diag}\left(\beta_{mi}\right)_{1 \le m \le M}, \\ C_i &\triangleq B_i - A_i, & D &\triangleq \rho \sum_{i=1}^{K} \eta_i C_i + I. \end{aligned} \tag{3.5}$$

3.3 Uplink Data Transmission

At the second step of the TDD protocol, users send data symbols and the m-th AP receives

$$y_m = \sqrt{\rho} \sum_{i=1}^{K} \sqrt{\eta_i} g_{mi} s_i + w_m, \tag{3.6}$$

where $w_m \sim \mathcal{CN}(0,1)$ is additive noise and s_i is the data signal transmitted by the i-th user. The NC uses estimates $\hat{g}_{mk}$ to form postcoding vectors $\boldsymbol{v}_k$ and obtains estimates of data signals $\hat{s}_k = \boldsymbol{v}_k^H \left[y_1, \cdots, y_M\right]^T,\ k = 1, \cdots, K$. Using the worst-case uncorrelated additive noise, the uplink achievable rate of the k-th user is $R = \mathbb{E}\left(\log_2\left(1 + \text{SINR}_k\right)\right)$, with

$$\text{SINR}_k\left(\boldsymbol{v}_k\right) = \frac{\rho \eta_k \boldsymbol{v}_k^H \hat{\boldsymbol{g}}_k \hat{\boldsymbol{g}}_k^H \boldsymbol{v}_k}{\boldsymbol{v}_k^H \left(\rho \sum_{i \neq k}^{K} \eta_i \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H + D\right) \boldsymbol{v}_k}, \tag{3.7}$$

where $\hat{\boldsymbol{g}}_i = \left[\hat{g}_{1i}, \cdots, \hat{g}_{Mi}\right]^T$. Note that the achievable SINR of the k-th user in (3.7) is obtained by taking into account the channel estimation error and pilot contamination effect.

3.3.1 MMSE Receiver

First, we consider MMSE receiver, which maximizes SINR of each user. The MMSE vector to decode the data symbol of the k-th user is given by

$$\boldsymbol{v}_k^{\text{MMSE}} = \sqrt{\rho \eta_k} \left(\rho \sum_{i=1}^{K} \eta_i \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H + D\right)^{-1} \hat{\boldsymbol{g}}_k. \tag{3.8}$$

Note that the MMSE vector in (3.8) contains channel estimates of all users in the network. Thus, it is optimal in the sense that it maximizes SINR of each user. Whereas in cellular systems, the MMSE vector at cell ℓ only contains the channel vectors of cell ℓ and the second-order statistics of channel coefficients between base station at cell ℓ and all users in the network

[34, 48]. The achievable SINR of the k-th user with MMSE receiver is given by

$$\text{SINR}_k^{\text{MMSE}} = \text{SINR}_k\left(\boldsymbol{v}_k^{\text{MMSE}}\right) = \frac{\hat{\boldsymbol{g}}_k{}^H \left(\rho \sum_{i=1}^K \eta_i \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H + D\right)^{-1} \hat{\boldsymbol{g}}_k}{\frac{1}{\rho\eta_k} - \hat{\boldsymbol{g}}_k{}^H \left(\rho \sum_{i=1}^K \eta_i \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H + D\right)^{-1} \hat{\boldsymbol{g}}_k}. \tag{3.9}$$

The Monte Carlo simulation of $R_k^{\text{MMSE}} = \log_2\left(1 + \text{SINR}_k^{\text{MMSE}}\right)$ requires long averaging over small scale fading coefficients h_{mk}. Hence, it is desirable to have an approximation of R_k^{MMSE} as a function of large scale fading coefficients only. The correlation between the channel estimates (i.e., $\hat{g}_{mi} = \frac{\beta_{mi}}{\beta_{mk}} \hat{g}_{mk}$ for $i, k \in \mathcal{S}_{b_k}$) does not allow us to use random matrix theory tools ([49, Theorem 1,2], [48]) to achieve this goal. Below we propose *partial MMSE* receiver whose performance is very close to the performance of the MMSE receiver and allows us to overcome this problem.

3.3.2 Partial MMSE Receiver

Let $\mathcal{I}_k = \mathcal{S}_{b_k} \cup \left\{u_1^{(k)}, \cdots, u_\tau^{(k)}\right\}$, where $u_j^{(k)} \in \mathcal{S}_j$ is the index of a user from $\mathcal{S}_j$, whose selection rule is discussed later. The partial MMSE vector for user k is then defined by

$$\begin{aligned}\boldsymbol{v}_k^{\text{PMMSE}} &= \sqrt{\rho\eta_k}\left(\rho \sum_{i\in\mathcal{I}_k} \eta_i \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H + \rho \sum_{i\notin\mathcal{I}_k} \mathbb{E}\left(\eta_i \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H\right) + D\right)^{-1} \hat{\boldsymbol{g}}_k \\ &= \sqrt{\rho\eta_k}\left(\rho \sum_{i\in\mathcal{I}_k} \eta_i \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H + Q\right)^{-1} \hat{\boldsymbol{g}}_k, \end{aligned} \tag{3.10}$$

where

$$Q = \rho \sum_{i\notin\mathcal{I}_k} \eta_i B_i + \rho \sum_{i\in\mathcal{I}_k} \eta_i C_i + I.$$

Note that $\mathcal{I}_k$ contains one user from each non-coherent interference group $\mathcal{S}_j,\ j \neq k$, and all users that cause coherence interference to user k. Recall that in massive MIMO systems, the coherent interference is the dominant impairment which limits the system performance when the number of antennas increases without bound [15]. Therefore, in the partial MMSE receiver we include channel vectors of all users that use the same pilot sequence as user k.

The users $u_1^{(k)}, \cdots, u_\tau^{(k)}$ should be chosen such that vectors $\hat{\boldsymbol{g}}_i, i \in \mathcal{I}_k$ in (3.10) have the major contribution in (3.8) and hence (3.10) becomes close to (3.8). Numerical results show that a random selection of users $u_1^{(k)}, \cdots, u_\tau^{(k)}$ from the corresponding sets $\mathcal{S}_1, \cdots, \mathcal{S}_\tau$ leads to poor performance (see Figure 3.1). A method for smart choice of these users can be formulated as following

$$u_j^{(k)} = \arg\max_{i \in \mathcal{S}_j} \boldsymbol{\beta}_k^T \boldsymbol{\beta}_i, \quad j = 1, \cdots, \tau, \tag{3.11}$$

where $\boldsymbol{\beta}_i = [\beta_{1i}, \cdots, \beta_{Mi}]^T$. In other words, we choose user $u_j^{(k)} \in \mathcal{S}_j$ that is in the close vicinity of the k-th user. The $\text{SINR}_k^{\text{PMMSE}}$ can be obtained by substituting $\boldsymbol{v}_k^{\text{PMMSE}}$ in (3.7).

In the following theorem we apply random matrix theory to obtain an asymptotic approximation of $R_k^{\text{PMMSE}} = \log_2\left(1 + \text{SINR}_k^{\text{PMMSE}}\right)$ when M and K grow infinitely large while the ratio M/K is finite. This asymptotic result is used as an approximation for finite values of M and K similar to [49] and [48], in which the approximations are derived for multiple-input single-output broadcast channel and cellular systems respectively.

Theorem 3.1. *Assume matrices A_i, C_i, $i = 1, \cdots, K$ have uniformly bounded spectral norms. For the partial MMSE receiver defined in (3.10), when M and K grow infinitely large such that $0 < \liminf_M \frac{M}{K} \leq \limsup_M \frac{M}{K} < \infty$, we have*

$$\text{R}_k^{\text{PMMSE}} - \log_2\left(1 + \widehat{\text{SINR}}_k^{\text{PMMSE}}\right) \xrightarrow[M,K\to\infty]{\text{a.s.}} 0, \tag{3.12}$$

where

$$\widehat{\text{SINR}}_k^{PMMSE} = \frac{\rho \eta_k \lambda_k^2}{\theta\left(D\right) + \rho \sum\limits_{i \in \mathcal{S}_{b_k} \setminus \{k\}} \eta_i \lambda_i^2 + \rho \sum\limits_{i \notin \mathcal{I}_k} \eta_i \theta\left(A_i\right) + \rho \sum\limits_{i \in \mathcal{I}_k \setminus \mathcal{S}_{b_k}} \frac{\eta_i \theta(A_i)}{\left(1 + \rho \frac{\eta_i}{M} \text{tr}\left(A_i T_i''\right)\right)^2}}, \tag{3.13}$$

and all parameters in $\widehat{\text{SINR}}_k^{\text{PMMSE}}$ are summarized in Tables 3.1 and 3.2.

Proof. See Appendix 3.7.A □

Table 3.1: Parameter definitions in Theorem 3.1 (Part I).

$\delta_i^{(t)}$	$\frac{\rho\eta_i}{M}\mathrm{tr}\, A_i\left(\frac{\rho}{M}\sum_{j\in\mathcal{I}_k\setminus\mathcal{S}_{b_k}}\frac{\eta_j A_j}{1+\delta_j^{(t-1)}}+\frac{1}{M}Q\right)^{-1}$
δ_i	$\lim_{t\to\infty}\delta_i^{(t)}$, with $\delta_i^{(0)}=M$
T	$\left(\frac{\rho}{M}\sum_{j\in\mathcal{I}_k\setminus\mathcal{S}_{b_k}}\frac{\eta_j A_j}{1+\delta_j}+\frac{1}{M}Q\right)^{-1}$
$[J]_{jl}$	$\frac{\frac{\rho^2}{M}\mathrm{tr}\left(\eta_j\eta_l A_j T A_l T\right)}{M\left(1+\delta_l\right)^2},\quad j,l\in\mathcal{I}_k\setminus\mathcal{S}_{b_k}$
$\boldsymbol{\delta}'(H)$	$\left[\delta_j'(H)\right]_{j\in\mathcal{I}_k\setminus\mathcal{S}_{b_k}}=(I-J)^{-1}\left[\frac{\rho\eta_j}{M}\mathrm{tr}\left(A_j T H T\right)\right]_{j\in\mathcal{I}_k\setminus\mathcal{S}_{b_k}}$
$T'(H)$	$THT+T\frac{\rho}{M}\sum_{j\in\mathcal{I}_k\setminus\mathcal{S}_{b_k}}\frac{\eta_j A_j\delta_j'(H)}{(1+\delta_j)^2}T$
$\delta''^{(t)}_{ni}$	$\frac{\rho\eta_i}{M}\mathrm{tr}\, A_i\left(\frac{\rho}{M}\sum_{j\in\mathcal{I}_k\setminus\{n\}}\frac{\eta_j A_j}{1+\delta''^{(t-1)}_{nj}}+\frac{1}{M}Q\right)^{-1}$
δ''_{ni}	$\lim_{t\to\infty}\delta''^{(t)}_{ni}$, with $\delta''^{(0)}_{ni}=M$
T''_n	$\left(\frac{\rho}{M}\sum_{j\in\mathcal{I}_k\setminus\{n\}}\frac{\eta_j A_j}{1+\delta''_{nj}}+\frac{1}{M}Q\right)^{-1}$

Table 3.2: Parameter definitions in Theorem 3.1 (Part II).

$\boldsymbol{\gamma}_i$	$\frac{\sqrt{\rho}}{M}\left[\sqrt{\eta_j}\text{tr}\left(A_i^{1/2}A_j^{1/2}T\right)\right]_{j\in\mathcal{S}_{b_i}}$
Γ	$I+\frac{\rho}{M}\left[\left[\sqrt{\eta_i\eta_j}\text{tr}\left(A_i^{1/2}A_j^{1/2}T\right)\right]\right]_{i\in\mathcal{S}_{b_k},j\in\mathcal{S}_{b_k}}$
$\boldsymbol{\nu}_i(H)$	$\frac{\sqrt{\rho}}{M^2}\left[\sqrt{\eta_j}\text{tr}\left(A_j^{1/2}A_i^{1/2}T'(H)\right)\right]_{j\in\mathcal{S}_{b_k}}$
$N(H)$	$\frac{\rho}{M^2}\left[\left[\sqrt{\eta_i\eta_j}\text{tr}\left(A_j^{1/2}A_i^{1/2}T'(H)\right)\right]\right]_{i\in\mathcal{S}_{b_k},j\in\mathcal{S}_{b_k}}$
λ_i	$\frac{1}{M}\text{tr}\left(A_i^{1/2}A_k^{1/2}T\right)-\boldsymbol{\gamma}_k^T\Gamma^{-1}\boldsymbol{\gamma}_i$
$\theta(H)$	$\frac{1}{M^2}\text{tr}\left(A_kT'(H)\right)-2\text{Re}\left(\boldsymbol{\nu}_k(H)^T\Gamma^{-1}\boldsymbol{\gamma}_k\right)$ $+\boldsymbol{\gamma}_k^T\Gamma^{-1}N(H)\Gamma^{-1}\boldsymbol{\gamma}_k$

Note that the approximation $\widehat{\text{SINR}}_k^{\text{PMMSE}}$ in (3.13) is a function of large scale fading coefficients *only*, and although it has a long formulation, it can be easily calculated numerically for large values of M and K.

3.3.3 Large Scale Fading Decoding

Next, we propose the LSFD receiver for cell-free systems. The main idea of LSFD receiver is that postcoding vector and power coefficients depend only on the large scale fading coefficients. Since these coefficients are independent of frequency and change (about 40 times) slower than small scale fading coefficients, LSFD allows one to reduce the computational complexity in NC, which is very desirable in real life systems.

The m-th AP sends y_m to the NC. Utilizing MF, the NC computes $\tilde{s}_{mi} = \hat{g}_{mi}^* y_m$ and estimates data symbol s_k by using linear combination of all received signals as following

$$\hat{s}_k = \sum_{m=1}^{M} \sum_{i=1}^{K} v_{mki}^* \tilde{s}_{mi}. \tag{3.14}$$

The NC computes postcoding coefficients v_{kmi} and power coefficients η_k as a function of large scale fading coefficients only.

Lemma 3.1. *The estimate of data symbol $\hat{s}_k$ in (3.14) can be simplified as*

$$\hat{s}_k = \boldsymbol{v}_k^H \tilde{\boldsymbol{s}}_k, \tag{3.15}$$

where $\boldsymbol{v}_k = [v_{1k}, \cdots, v_{Mk}]^T$ and $\tilde{\boldsymbol{s}} = [\tilde{s}_{1k}, \cdots, \tilde{s}_{Mk}]^T$.

Proof follows directly from the two facts. First, s_k and $\tilde{s}_{mi}, i \notin \mathcal{S}_{b_k}$ are uncorrelated and hence assignment $v_{mki} = 0, i \notin \mathcal{S}_{b_k}$ in (3.14) does not result in any performance loss, and the second fact is that $\tilde{s}_{mi} = \frac{\beta_{mi}}{\beta_{mk}} \tilde{s}_{mk},\ i \in \mathcal{S}_{b_k}$.

Theorem 3.2. *Achievable SINR of the k-th user with LSFD receiver is given by*

$$\text{SINR}_k\left(\boldsymbol{v}_k\right) = \frac{\rho\eta_k \boldsymbol{v}_k^H \boldsymbol{\mu}_k \boldsymbol{\mu}_k^H \boldsymbol{v}_k}{\rho \sum_{i\in\mathcal{S}_{b_k}\setminus\{k\}} \eta_i \boldsymbol{v}_k^H \boldsymbol{\mu}_i \boldsymbol{\mu}_i^H \boldsymbol{v}_k + \boldsymbol{v}_k^H \Lambda \boldsymbol{v}_k}, \tag{3.16}$$

where $\Lambda = \text{diag}\left(\rho \sum_{i=1}^{K} \eta_i \alpha_{mk} \beta_{mi} + \alpha_{mk}\right)_{1\leq m\leq M}$ *and* $\mu_i = \left[\frac{\rho\tau\beta_{mk}\beta_{mi}}{1+\rho\tau\sum_{j\in\mathcal{S}_{b_i}}\beta_{mj}}\right]_{1\leq m\leq M}$.

Proof. See Appendix 3.7.B □

Based on Rayleigh quotient theorem [50], the optimal $\boldsymbol{v}_k^{\text{LSFD}}$ which maximizes SINR of each user in (3.16) is given by

$$\boldsymbol{v}_k^{\text{LSFD}} = \left(\rho \sum_{i\in\mathcal{S}_{b_k}\setminus\{k\}} \eta_i \boldsymbol{\mu}_i \boldsymbol{\mu}_i^H + \Lambda\right)^{-1} \boldsymbol{\mu}_k. \tag{3.17}$$

The optimal SINR of the k-th user is then given by

$$\text{SINR}_k^{\text{LSFD}} = \rho\eta_k \boldsymbol{\mu}_k^H \left(\rho \sum_{i\in\mathcal{S}_{b_k}\setminus\{k\}} \eta_i \boldsymbol{\mu}_i \boldsymbol{\mu}_i^H + \Lambda\right)^{-1} \boldsymbol{\mu}_k. \tag{3.18}$$

The power coefficients $\boldsymbol{\eta} = \left[\eta_1, \cdots, \eta_K\right]^T$ in (3.18) can be obtained through solving the max-min power allocation problem with per user transmit power constraints, which is formulated as

$$\max_{\boldsymbol{\eta}} \min_{k} R_k^{\text{LSFD}} = \log_2\left(1 + \text{SINR}_k^{\text{LSFD}}\right), \tag{3.19a}$$

$$\text{s.t. } 0 \leq \eta_i \leq 1, \quad i = 1, \cdots, K. \tag{3.19b}$$

In order to solve the max-min optimization problem (3.19), we reformulate it as the following power optimization problem

$$\max_{\boldsymbol{\eta}, t} \; t$$

$$\text{s.t. } \text{SINR}_k^{\text{LSFD}}\left(\boldsymbol{\eta}\right) \geq t, \quad k = 1, \cdots, K, \tag{3.20a}$$

$$0 \leq \eta_i \leq 1, \quad i = 1, \cdots, K. \tag{3.20b}$$

Let t^{opt} denote the optimal value of t in (3.20). Next, we define the vector function $\boldsymbol{I}_t(\boldsymbol{\eta})$, which helps to determine if $t \leq t^{\text{opt}}$ or not. The k-th element of $\boldsymbol{I}_t(\boldsymbol{\eta})$ is defined as follows

$$\boldsymbol{I}_t^{(k)}(\boldsymbol{\eta}) \triangleq \frac{t}{\rho \boldsymbol{\mu}_k^H \left(\rho \sum_{i \in \mathcal{S}_{b_k} \setminus \{k\}} \eta_i \boldsymbol{\mu}_i \boldsymbol{\mu}_i^H + \Lambda\right)^{-1} \boldsymbol{\mu}_k}, \quad k = 1, \cdots, K. \tag{3.21}$$

It is easy to observe that the constraint in (3.20a) can be written as $\boldsymbol{\eta} \geq \boldsymbol{I}_t(\boldsymbol{\eta})$. We recall the definition of standard interference function from [51], which refers to any function $\boldsymbol{I}_t(\boldsymbol{\eta})$ that satisfies the following properties for all $\eta \geq 0$:

1. $\boldsymbol{I}_t(\boldsymbol{\eta}) \geq 0$ is positive.
2. If $\boldsymbol{\eta} \geq \boldsymbol{\eta}'$, then $\boldsymbol{I}_t(\boldsymbol{\eta}) \geq \boldsymbol{I}_t(\boldsymbol{\eta}')$.
3. For all $\alpha > 0$, $\alpha \boldsymbol{I}_t(\boldsymbol{\eta}) \geq \boldsymbol{I}_t(\alpha \boldsymbol{\eta})$.

Lemma 3.2. *For any fixed value $t > 0$, the function $\boldsymbol{I}_t(\boldsymbol{\eta})$ defined in (3.21) is a standard interference function.*

The proof follows from the definition in (3.21) and hence is left to the reader.

Denote by $\hat{\boldsymbol{\eta}}$ the fixed point solution of

$$\boldsymbol{\eta} = \min\{1, \boldsymbol{I}_t(\boldsymbol{\eta})\}. \tag{3.22}$$

It has been shown in [51] that the fixed point solution $\hat{\boldsymbol{\eta}}$ for any standard function $\boldsymbol{I}_t(\boldsymbol{\eta})$ has the following property that

$$\hat{\boldsymbol{\eta}} \geq \boldsymbol{I}_t(\hat{\boldsymbol{\eta}}) \quad \text{if and only if} \quad \exists\, 0 \leq \boldsymbol{\eta} \leq 1 \text{ such that } \boldsymbol{\eta} \geq \boldsymbol{I}_t(\boldsymbol{\eta}). \tag{3.23}$$

We refer to this evaluation as the feasibility check on t. This fact provides us a tool to develop bisection method to solve max-min optimization (3.20), which starts with two values of t_l and t_u for which the constraints (3.20a), (3.20b) are respectively feasible (t_l) and not feasible (t_u). The bisection method then recursively narrows down the interval (t_l, t_u) as summarized in Algorithm 1 until the interval is small enough.

Algorithm 1 LSFD max-min optimization problem.

1) Choose an interval (t_l, t_u) that contains the optimal t^{opt}, i.e., $t_l < t^{\text{opt}} < t_u$.

2) For midpoint $t = \frac{t_l + t_u}{2}$, find the fixed point solution $\hat{\boldsymbol{\eta}}$ using the following iteration

$$\boldsymbol{\eta}^{(j+1)} = \min\left\{1, \boldsymbol{I}_t\left(\boldsymbol{\eta}^{(j)}\right)\right\}. \tag{3.24}$$

3) If t is feasible meaning that the fixed point solution $\hat{\boldsymbol{\eta}}$ satisfies $\hat{\boldsymbol{\eta}} \geq \boldsymbol{I}_t\left(\hat{\boldsymbol{\eta}}\right)$, then set $t_l \leftarrow t$, otherwise set $t_u \leftarrow t$.

4) Continue steps 2 and 3 until the search interval is small enough, i.e., $t_u - t_l < \epsilon$.

3.3.4 Asymptotic Analysis of MF Receiver

In this subsection, we provide the SINR expression for MF receiver when the number of APs grows without limit.

Theorem 3.3. *Achievable uplink SINR of the k-th user for MF receiver, i.e., $\boldsymbol{v}_k^{\text{MF}} = [1, \cdots, 1]^T$, with unlimited number of APs ($M \to \infty$ and K = constant) and independent large scale fading coefficients is given by*

$$\text{SINR}_k\left(\boldsymbol{v}_k^{\text{MF}}\right) \xrightarrow[M\to\infty]{\text{a.s.}} \frac{\eta_k \bar{\mu}_k^2}{\sum\limits_{i \in \mathcal{S}_{b_k} \setminus \{k\}} \eta_i \bar{\mu}_i^2}, \tag{3.25}$$

where $\bar{\mu}_i$ is defined as

$$\bar{\mu}_i \triangleq \lim_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} \mu_{mi}$$

with $\mu_{mi} = \frac{\rho\tau\beta_{mk}\beta_{mi}}{1+\rho\tau\sum_{j\in\mathcal{S}_{b_i}}\beta_{mj}}$.

Proof. See Appendix 3.7.C. □

Note that the denominator in (3.25) corresponds to power of the pilot contamination related interference. Similar to the cellular massive MIMO systems, SINR of the k-th user using MF receiver is limited by the effect of pilot contamination. However, we observe that

by increasing M, the SINR of cell-free system becomes a deterministic constant whereas the SINR in cellular systems remains dependent on the large scale fading coefficients.

3.4 Numerical Results

We consider a square dense urban area of 2×2 km^2 with M randomly located APs and K randomly located users. The area is wrapped around to avoid boundary effects. For large scale fading coefficients, we consider a three-slope path loss model [52] as follows

$$\beta_{mk} = \begin{cases} c_0 & d_k \leq 0.01 \text{ km} \\ \dfrac{c_1}{d_{mk}^2} & 0.01\text{km} < d_k \leq 0.05 \text{ km}\,, \\ \dfrac{c_2 z_{mk}}{d_{mk}^{3.5}} & d_k > 0.05 \text{ km} \end{cases} \tag{3.26}$$

where d_{mk} is the distance in kilometers between user k and the AP m, and z_{mk} is the log-normal shadow fading, i.e., $10 \log_{10} z_{mk} \sim \mathcal{N}(0, \sigma_{\text{shad}}^2)$ with $\sigma_{\text{shad}} = 8$ dB. For $d_k > 0.05$ km we use COST-231 Hata propagation model

$$\begin{aligned} 10 \log_{10} c_2 = &- 46.3 - 33.9 \log_{10} f + 13.82 \log_{10} h_B \\ &+ (1.1 \log_{10} f - 0.7) h_R - (1.56 \log_{10} f - 0.8), \end{aligned} \tag{3.27}$$

where $f = 1900$ MHz is the carrier frequency, $h_B = 15$ m is the AP antenna hight, and $h_R = 1.65$ m is the user antenna hight. Parameters c_1 and c_2 in (3.26) are chosen in a way that path loss remains continuous at boundary points.

To model the correlation between large scale fading coefficients caused by closely located users and/or APs, we use the same correlation model as [39, Equations (53) and (54)] for shadow fading with parameters $\delta = 0.5$ and $d_{\text{decorr}} = 0.1$ km. The noise variance is $\sigma_v^2 = 290 \times \kappa \times B \times NF$, where κ, B, and NF are Boltzmann constant, bandwidth (20 MHz)

and noise figure (9 dB) respectively. We assume users transmit with equal power $\eta_i = 1,\ i = 1, \cdots, K$ and $\rho = 200$ mW.

Figure 3.1 depicts CDFs of R^{MMSE}, R^{PMMSE} with the heuristic user selection approach given in (3.11), R^{PMMSE} with random user selection, $\hat{R}^{\text{PMMSE}} = \log_2\left(1 + \widehat{\text{SINR}}^{\text{PMMSE}}\right)$, and R^{LSFD} with independent large scale fading coefficients. The CDF of per-user throughput achieved by MF receiver [39] is also included in the figure for comparison. The horizontal line corresponds to 5%-outage rate which represents the smallest rate among 95% of the best users. One can observe that the asymptotic approximation of MMSE receiver is very tight. MMSE and LSFD receivers respectively provide 5.1-fold and 2.6-fold gain over the MF receiver in terms of 5%-outage rate. Performance of the LSFD receiver lies between the simple MF receiver and MMSE receivers. Compared to the MMSE receiver, LSFD reduces the overall complexity of the system.

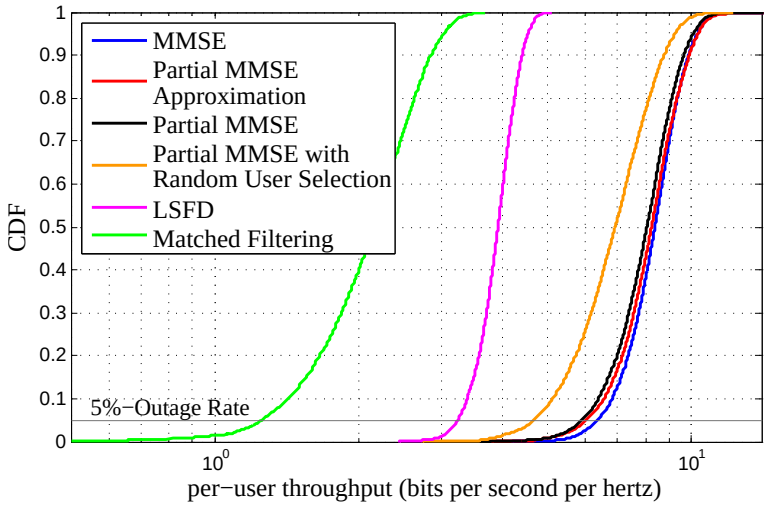

Figure 3.1: CDFs of the achievable per-user rates for LSFD and MMSE receivers with $M = 1000$, $K = 50$, and $\tau = 10$.

Figure 3.2 shows 5%-outage and mean values of R^{MMSE}, R^{PMMSE}, $\hat{R}^{\text{PMMSE}}$ versus the number of APs under independent and correlated shadow fading. We point out that in all con-

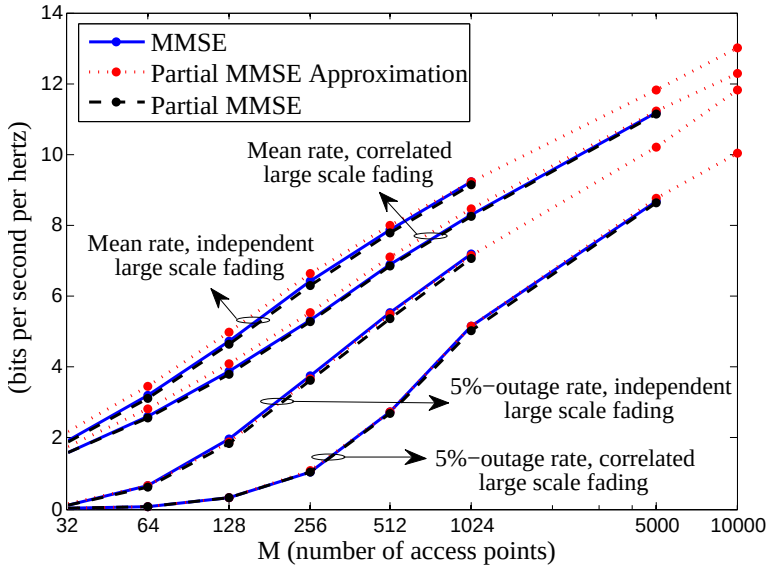

Figure 3.2: 5%-Outage and mean rates versus M for correlated and independent large scale fading with $K = 16$ and $\tau = 4$.

sidered scenarios, the partial MMSE is virtually optimal and our approximation $\hat{R}^{\text{partial}}_{\text{MMSE}}$ is very accurate. It is to be noted that the shadow fading correlation significantly affects the system performance.

The CDFs of per-user rates for different number of APs and users are plotted in Figure 3.3. The ratio between number of APs and users is constant in all cases, i.e., $^{M}/_{K} = 8$ and $^{K}/_{\tau} = 4$. We observe that the 5%-outage rate of MMSE and partial MMSE receivers increase as the network size increases.

3.5 Conclusion

In this chapter we studied the uplink performance of cell-free systems with MMSE and LSFD receivers. To study the asymptotic behavior of the cell-free systems, a more tractable (and hence suboptimal) MMSE receiver is introduced. The achievable rates of MMSE, PMMSE, and asymptotic approximation are shown to be very close. The numerical

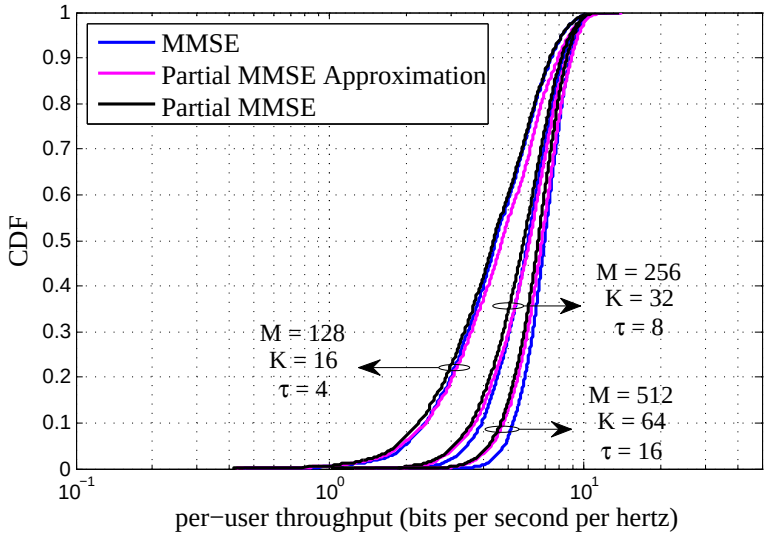

Figure 3.3: CDFs of the achievable per-user rates for MMSE receivers with different number of APs and users.

results confirm that the asymptotic approximation is very accurate even for small number of APs and users. A generalization of LSFD receiver for cell-free systems is introduced, which depends on the large scale fading coefficients only. While MMSE and LSFD demonstrate significant gains over MF, there is a considerable gap between MMSE and LSFD receivers.

3.6 Acknowledgment

This chapter contains material as it appears in [3], E. Nayebi, A. Ashikhmin, T. L. Marzetta, B. D. Rao, “Performance of Cell-Free Massive MIMO Systems with MMSE and LSFD Receivers," in *Proc. 50th Asilomar Conference on Signals, Systems and Computers*, Nov. 2016, pp. 203-207. The dissertation author was the primary investigator and author of this paper.

3.7 Appendices

3.7.A Proof of Theorem 3.1

For notational simplicity, we define the following matrices

$$U = Q + \rho \sum_{i \in \mathcal{I}_k \setminus \mathcal{S}_{b_k}} \eta_i \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H, \qquad R = I + \rho \left[\left[\sqrt{\eta_i \eta_j} \hat{\boldsymbol{g}}_i^H U^{-1} \hat{\boldsymbol{g}}_j \right] \right]_{i \in \mathcal{S}_{b_k}, j \in \mathcal{S}_{b_k}}$$

$$Z = \left[\left[\sqrt{\rho \eta_i} \hat{g}_{mi} \right] \right]_{1 \le m \le M, i \in \mathcal{S}_{b_k}}, \qquad W = \rho \sum_{i \in \mathcal{I}_k}^{K} \eta_i \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H + Q. \tag{3.28}$$

We will need the following preparatory lemma to prove Theorem 3.1.

Lemma 3.3. *Let $\hat{\boldsymbol{g}}_i \sim \mathcal{CN}(0, A_i)$ be the channel estimates defined in (3.3). Assume $H \in \mathbb{C}^{M \times M}$ is a Hermitian non-negative definite matrix with uniformly bounded spectral norm with respect to M. Then, for $i \in \mathcal{S}_{b_k}$, we have*

$$\hat{\boldsymbol{g}}_k U^{-1} \hat{\boldsymbol{g}}_i - \frac{1}{M} \mathrm{tr}\left(A_k^{1/2} A_i^{1/2} T \right) \xrightarrow[M,K \to \infty]{\text{a.s.}} 0, \quad \textit{and}$$

$$\hat{\boldsymbol{g}}_k^H U^{-1} H U^{-1} \hat{\boldsymbol{g}}_i - \frac{1}{M^2} \mathrm{tr}\left(A_i^{1/2} A_k^{1/2} T'(H) \right) \xrightarrow[M,K \to \infty]{\text{a.s.}} 0, \tag{3.29}$$

where T and $T'(H)$ are defined in Table 3.1.

Proof. Channel estimate $\hat{\boldsymbol{g}}_i$, for every $i \in \mathcal{S}_{b_k}$, can be written as $\hat{\boldsymbol{g}}_i = A_i^{1/2} \hat{\boldsymbol{h}}_k$, where $\hat{\boldsymbol{h}}_k \sim \mathcal{CN}(0, I)$ is complex normal random vector. For $i \in \mathcal{S}_{b_k}$, we have

$$\hat{\boldsymbol{g}}_k^H U^{-1} \hat{\boldsymbol{g}}_i = \frac{1}{M} \hat{\boldsymbol{h}_k}^H A_k^{1/2} \big(\frac{1}{M} U \big)^{-1} A_i^{1/2} \hat{\boldsymbol{h}}_k,$$

$$\hat{\boldsymbol{g}}_k^H U^{-1} H U^{-1} \hat{\boldsymbol{g}}_i = \frac{1}{M^2} \hat{\boldsymbol{h}}_k^H A_k^{1/2} \big(\frac{1}{M} U \big)^{-1} H \big(\frac{1}{M} U \big)^{-1} A_i^{1/2} \hat{\boldsymbol{h}}_k. \tag{3.30}$$

By [53, Theorem 3.4], we have

$$\frac{1}{M} \hat{\boldsymbol{h}_k}^H A_k^{1/2} \big(\frac{1}{M} U \big)^{-1} A_i^{1/2} \hat{\boldsymbol{h}}_k - \frac{1}{M} \mathrm{tr}\left(A_i^{1/2} A_k^{1/2} \big(\frac{1}{M} U \big)^{-1} \right) \xrightarrow[M,K \to \infty]{\text{a.s.}} 0,$$

$$\frac{1}{M^2} \hat{\boldsymbol{h}}_k^H A_k^{1/2} \big(\frac{1}{M} U \big)^{-1} H \big(\frac{1}{M} U \big)^{-1} A_i^{1/2} \hat{\boldsymbol{h}}_k - \frac{1}{M^2} \mathrm{tr}\left(A_i^{1/2} A_k^{1/2} \big(\frac{1}{M} U \big)^{-1} H \big(\frac{1}{M} U \big)^{-1} \right) \xrightarrow[M,K \to \infty]{\text{a.s.}} 0. \tag{3.31}$$

From [54, Theorem 1], we obtain

$$\frac{1}{M}\text{tr}\left(A_i^{1/2}A_k^{1/2}\big(\frac{1}{M}U\big)^{-1}\right)-\frac{1}{M}\text{tr}\left(A_i^{1/2}A_k^{1/2}T\right)\xrightarrow[M,K\to\infty]{\text{a.s.}}0. \tag{3.32}$$

By using [54, Theorem 2], we get

$$\frac{1}{M^2}\text{tr}\left(A_i^{1/2}A_k^{1/2}\big(\frac{1}{M}U\big)^{-1}H\big(\frac{1}{M}U\big)^{-1}\right)-\frac{1}{M}\text{tr}\left(A_i^{1/2}A_k^{1/2}T'\left(H\right)\right)\xrightarrow[M,K\to\infty]{\text{a.s.}}0, \tag{3.33}$$

which completes the proof. □

First, we divide the numerator and the denominator of $\text{SINR}_k^{\text{PMMSE}}$ by $\frac{1}{\rho\eta_k}$. Then, the useful signal power is given by

$$\frac{1}{\rho\eta_k}\left|\boldsymbol{v}_k^H\hat{\boldsymbol{g}}_k\right|^2=\underbrace{\rho\eta_k\left|\hat{\boldsymbol{g}}_k^HW^{-1}\hat{\boldsymbol{g}}_k\right|^2}_{T_0}. \tag{3.34}$$

The denominator of $\text{SINR}_k^{\text{PMMSE}}$ can be rewritten comprising of four components

$$\begin{aligned}\frac{1}{\rho\eta_k}\boldsymbol{v}_k^H\left(\rho\sum_{i\neq k}^{K}\eta_i\hat{\boldsymbol{g}}_i\hat{\boldsymbol{g}}_i^H+D\right)\boldsymbol{v}_k=&\underbrace{\hat{\boldsymbol{g}}_k^HW^{-1}DW^{-1}\hat{\boldsymbol{g}}_k}_{T_1}+\underbrace{\hat{\boldsymbol{g}}_k^HW^{-1}\rho\sum_{i\in\mathcal{S}_{b_k}\setminus\{k\}}^{K}\eta_i\hat{\boldsymbol{g}}_i\hat{\boldsymbol{g}}_i^HW^{-1}\hat{\boldsymbol{g}}_k}_{T_2}\\&+\underbrace{\hat{\boldsymbol{g}}_k^HW^{-1}\rho\sum_{i\notin\mathcal{I}_k}^{K}\eta_i\hat{\boldsymbol{g}}_i\hat{\boldsymbol{g}}_i^HW^{-1}\hat{\boldsymbol{g}}_k}_{T_3}+\underbrace{\hat{\boldsymbol{g}}_k^HW^{-1}\rho\sum_{i\in\mathcal{I}_k\setminus\mathcal{S}_{b_k}}^{K}\eta_i\hat{\boldsymbol{g}}_i\hat{\boldsymbol{g}}_i^HW^{-1}\hat{\boldsymbol{g}}_k}_{T_4}.\end{aligned} \tag{3.35}$$

Consider terms $T_0=\rho\eta_k\left|\hat{\boldsymbol{g}}_k^HW^{-1}\hat{\boldsymbol{g}}_k\right|^2$ and $T_2=\sum_{i\in\mathcal{S}_{b_k}\setminus\{k\}}^{K}\rho\eta_i\left|\hat{\boldsymbol{g}}_i^HW^{-1}\hat{\boldsymbol{g}}_k\right|^2$. Using matrix inversion lemma [55, Eq. C.7], term $\hat{\boldsymbol{g}}_iW^{-1}\hat{\boldsymbol{g}}_k$ can be written as

$$\hat{\boldsymbol{g}}_i^HW^{-1}\hat{\boldsymbol{g}}_k=\sqrt{\rho\eta_k}\hat{\boldsymbol{g}}_i^HU^{-1}\hat{\boldsymbol{g}}_k-\sqrt{\rho\eta_k}\hat{\boldsymbol{g}}_i^HU^{-1}ZR^{-1}Z^HU^{-1}\hat{\boldsymbol{g}}_k. \tag{3.36}$$

By applying Lemma 3.3 and the continuous mapping theorem [56], for $i\in\mathcal{S}_{b_k}$, we obtain

$$\hat{\boldsymbol{g}}_i^HU^{-1}Z-\boldsymbol{\gamma}_i^T\xrightarrow[M,K\to\infty]{\text{a.s.}}0,\qquad\text{and}\qquad R^{-1}-\Gamma^{-1}\xrightarrow[M,K\to\infty]{\text{a.s.}}0, \tag{3.37}$$

Combining (3.36) and (3.37), and using the continuous mapping theorem [56] yields

$$T_0 - \rho\eta_k\lambda_k^2 \xrightarrow[M,K\to\infty]{\text{a.s.}} 0, \qquad \text{and} \qquad T_2 - \sum_{i\in\mathcal{S}_{b_k}\setminus\{k\}} \rho\eta_i\lambda_i^2 \xrightarrow[M,K\to\infty]{\text{a.s.}} 0. \tag{3.38}$$

By matrix inversion lemma [55, Eq. C.7], term T_1 is given by

$$\begin{aligned} T_1 =& \hat{\boldsymbol{g}}_k^H U^{-1} D U^{-1} \hat{\boldsymbol{g}}_k - 2\text{Re}\left(\hat{\boldsymbol{g}}_k U^{-1} D U^{-1} Z R^{-1} Z^H U^{-1} \hat{\boldsymbol{g}}_k\right) \\ &+ \hat{\boldsymbol{g}}_k^H U^{-1} Z R^{-1} Z^H U^{-1} D U^{-1} Z R^{-1} Z^H U^{-1} \hat{\boldsymbol{g}}_k. \end{aligned} \tag{3.39}$$

By Lemma 3.3, it follows that

$$\begin{aligned} \hat{\boldsymbol{g}}_k^H U^{-1} D U^{-1} \hat{\boldsymbol{g}}_k - \frac{1}{M^2}\text{tr}\left(A_k T'\left(D\right)\right) &\xrightarrow[M,K\to\infty]{\text{a.s.}} 0, \\ \hat{\boldsymbol{g}}_k^H U^{-1} D U^{-1} Z - \boldsymbol{\nu}_k\left(D\right)^T &\xrightarrow[M,K\to\infty]{\text{a.s.}} 0 \\ Z^H U^{-1} D U^{-1} Z - N\left(D\right) &\xrightarrow[M,K\to\infty]{\text{a.s.}} 0. \end{aligned} \tag{3.40}$$

From (3.37) and (3.40), it follows that

$$T_1 - \theta\left(D\right) \xrightarrow[M,K\to\infty]{\text{a.s.}} 0. \tag{3.41}$$

Now consider term $\hat{\boldsymbol{g}}_k^H W^{-1} \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H W^{-1} \hat{\boldsymbol{g}}_k$ in T_3. From [53, Theorem 3.4], for $i \notin \mathcal{I}_k$, we obtain

$$\hat{\boldsymbol{g}}_k^H W^{-1} \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H W^{-1} \hat{\boldsymbol{g}}_k - \hat{\boldsymbol{g}}_k^H W^{-1} A_i W^{-1} \hat{\boldsymbol{g}}_k \xrightarrow[M,K\to\infty]{\text{a.s.}} 0. \tag{3.42}$$

Similar to (3.41), it is clear that $\hat{\boldsymbol{g}}_k^H W^{-1} A_i W^{-1} \hat{\boldsymbol{g}}_k - \theta\left(A_i\right) \xrightarrow[M,K\to\infty]{\text{a.s.}} 0$. Thus, we have

$$T_3 - \sum_{i\notin\mathcal{I}_k} \rho\eta_i \theta\left(A_i\right) \xrightarrow[M,K\to\infty]{\text{a.s.}} 0. \tag{3.43}$$

Define $W_{(i)} = W - \rho\eta_i \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H$. Note that where $W_{(i)}$ and $\hat{\boldsymbol{g}}_i$ are independent. By applying the matrix inversion lemma [53, Lemma 6.2], term $\rho\eta_i \hat{\boldsymbol{g}}_k^H W^{-1} \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H W^{-1} \hat{\boldsymbol{g}}_k$ in T_4 is written as

$$\rho\eta_i \hat{\boldsymbol{g}}_k^H W^{-1} \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H W^{-1} \hat{\boldsymbol{g}}_k = \rho\eta_i \frac{\hat{\boldsymbol{g}}_k^H W_{(i)}^{-1} \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H W_{(i)}^{-1} \hat{\boldsymbol{g}}_k}{\left(1 + \rho\eta_i \hat{\boldsymbol{g}}_i^H W_{(i)}^{-1} \hat{\boldsymbol{g}}_i\right)^2}. \tag{3.44}$$

Similar to the derivations in Lemma 3.3, we can show that

$$\hat{\boldsymbol{g}}_i^H W_{(i)}^{-1} \hat{\boldsymbol{g}}_i - \frac{1}{M} \mathrm{tr}\left(A_i T_i''\right) \xrightarrow[M,K\to\infty]{\text{a.s.}} 0, \tag{3.45}$$

where T'' is defined in Table 3.1. By [53, Theorem 3.4], for $i \in \mathcal{I}_k \setminus \mathcal{S}_{b_k}$, we have

$$\hat{\boldsymbol{g}}_k^H W_{(i)}^{-1} \hat{\boldsymbol{g}}_i \hat{\boldsymbol{g}}_i^H W_{(i)}^{-1} \hat{\boldsymbol{g}}_k - \hat{\boldsymbol{g}}_k^H W_{(i)}^{-1} A_i W_{(i)}^{-1} \hat{\boldsymbol{g}}_k \xrightarrow[M,K\to\infty]{\text{a.s.}} 0. \tag{3.46}$$

By (3.41) and [57, Rank-1 perturbation lemma], it follows that

$$\hat{\boldsymbol{g}}_k^H W_{(i)}^{-1} A_i W_{(i)}^{-1} \hat{\boldsymbol{g}}_k - \theta\left(A_i\right) \xrightarrow[M,K\to\infty]{\text{a.s.}} 0. \tag{3.47}$$

Thus, by the continuous mapping theorem [56], it yields

$$T_4 - \sum_{i\in\mathcal{I}_k\setminus\mathcal{S}_{b_k}} \rho\eta_i \frac{\theta\left(A_i\right)}{\left(1+\rho\frac{\eta_i}{M}\mathrm{tr}\left(A_i T_i''\right)\right)^2} \xrightarrow[M,K\to\infty]{\text{a.s.}} 0. \tag{3.48}$$

We complete the proof by combining (3.38), (3.41), (3.43), and (3.48).

3.7.B Proof of Theorem 3.2

The estimate of data symbol in (3.15) can be written as

$$\begin{aligned}
\hat{s}_{kl} = \sum_{m=1}^{M} v_{mk}^* \tilde{s}_{mk} = & \underbrace{\sum_{m=1}^{M} \sqrt{\rho\eta_k} v_{mk}^* E\left(\hat{g}_{mk}^* g_{mk}\right) s_k}_{T_0\text{: useful signal}} + \underbrace{\sum_{m=1}^{M} \sum_{i\in\mathcal{S}_{b_k}\setminus\{k\}} \sqrt{\rho\eta_i} v_{mk}^* E\left(\hat{g}_{mk}^* g_{mi}\right) s_i}_{T_1\text{: pilot contamination}} \\
& + \underbrace{\sum_{m=1}^{M} \sum_{i\in\mathcal{S}_{b_k}} \sqrt{\rho\eta_i} v_{mk^*} \left(\hat{g}_{mk}^* g_{mi} - E\left(\hat{g}_{mk}^* g_{mi}\right)\right) s_i}_{T_2\text{: lack of channel knowledge}} \\
& + \underbrace{\sum_{m=1}^{M} \sum_{i\notin\mathcal{S}_{b_k}} \sqrt{\rho\eta_i} v_{mk}^* \hat{g}_{mk}^* g_{mi} s_i}_{T_3\text{: interference}} + \underbrace{\sum_{m=1}^{M} v_{mk}^* \hat{g}_{mk}^* w_m}_{T_4\text{: noise}}.
\end{aligned} \tag{3.49}$$

Since data symbols of different users and additive Gaussian noise are mutually independent, terms T_0, T_1, T_2, T_3, and T_4 are mutually uncorrelated and have zero mean. According to [46],

the worst case noise in terms of mutual information is Gaussian noise with variance equal to the variance of $T_1+T_2+T_3+T_4$. Hence achievable rate is lower bounded by $\log\left(1+\text{SINR}_k\right)$, where

$$\text{SINR}_k = \frac{\mathbb{E}\left(|T_0|^2\right)}{\mathbb{E}\left(|T_1|^2+|T_2|^2+|T_3|^2+|T_4|^2\right)}. \tag{3.50}$$

Variances of terms T_0, T_1, T_2, T_3, and T_4 are given by

$$\begin{aligned}
&\mathbb{E}\left(|T_0|^2\right) = \rho\eta_k\left(\boldsymbol{v}_k^H\boldsymbol{\mu}_k\right)^2, && \mathbb{E}\left(|T_1|^2\right) = \sum_{i\in\mathcal{S}_{b_k}\setminus\{k\}}\rho\eta_i\left(\boldsymbol{v}_k^H\boldsymbol{\mu}_i\right)^2,\\
&\mathbb{E}\left(|T_2|^2\right) = \sum_{i\in\mathcal{S}_{b_k}}\rho\eta_i\sum_{m=1}^{M}|v_{mk}|^2\alpha_{mk}\beta_{mi}, && \mathbb{E}\left(|T_3|^2\right) = \sum_{i\notin\mathcal{S}_{b_k}}\rho\eta_i\sum_{m=1}^{M}|v_{mk}|^2\alpha_{mk}\beta_{mi},\\
&\mathbb{E}\left(|T_4|^2\right) = \sum_{m=1}^{M}|v_{mk}|^2\alpha_{mk}.
\end{aligned} \tag{3.51}$$

where $\boldsymbol{\mu}_i$ is defined in Theorem 3.2. Substituting the variance in (3.50) completes the proof.

3.7.C Proof of Theorem 3.3

The achievable uplink SINR of the k-th user with MF receiver is given by

$$\text{SINR}_k\left(\boldsymbol{v}_k=[1,\cdots,1]^T\right) = \frac{\rho\eta_k\left(\sum_{m=1}^{M}\mu_{mk}\right)^2}{\rho\sum_{i\in\mathcal{S}_{b_k}\setminus\{k\}}\eta_i\left(\sum_{m=1}^{M}\mu_{mi}\right)^2+\sum_{m=1}^{M}\alpha_{mk}\left(1+\rho\sum_{i=1}^{K}\eta_i\beta_{mi}\right)}, \tag{3.52}$$

From the central limit theorem and the continuous mapping theorem [56], we have

$$\left(\frac{1}{M}\sum_{m=1}^{M}\mu_{ik}\right)^2 - \bar{\mu}_k^2 \xrightarrow[M\to\infty]{\text{a.s.}} 0. \tag{3.53}$$

For constant K, we have

$$\frac{1}{M^2}\sum_{m=1}^{M}\alpha_{mk}\left(1+\rho\sum_{i=1}^{K}\eta_i\beta_{mi}\right) \xrightarrow[M\to\infty]{\text{a.s.}} 0. \tag{3.54}$$

By substituting (3.53) and (3.54) in (3.52), we complete the proof.

Chapter 4

Access Point Location Design

4.1 Introduction

Wireless networks with distributed access points (APs) have attracted a lot of attention due to their capability in providing enhanced network coverage, capacity, and power efficiency [34], [58], [59]. The advantages associated with distributed systems highly depend on the AP locations. In this chapter we investigate AP location design problem in cell-free massive multiple-input multiple-output (MIMO) and small-cell systems that were introduced in Chapter 2. We propose new algorithms to obtain optimal location of APs that are based on distribution of users in the network.

Cell-free massive MIMO is a particular deployment of massive MIMO with distributed APs, in which there are no cell boundaries and each user is served with all distributed APs (see Chapter 2). This system combines the notion of massive MIMO systems with distributed MIMO providing uniformly good service for all users in the network [1], [39].

An alternative wireless network with distributed APs is small-cell system in which the APs do not cooperate and each user is served by only one AP. Compared with the cell-free massive MIMO, small-cell systems are simpler as they require much less backhaul and coordination among APs at the cost of reduced per-user throughput as observed in Chapter 2 and [39]. However, these two systems are the extreme scenarios in distributed MIMO systems. In real-life systems with distributed APs, each user will be served by a few APs.

In order to maximize the average ergodic capacity, the authors in [60] propose a squared distance criterion for designing antenna locations in generalized distributed antenna systems. They minimize the expectation of the squared distance between a randomly located user and the nearest antenna port, which is equivalent to codebook design problem in vector quantization for which the Lloyd algorithm is a popular approach. In [61], the authors investigate the optimal deployment of APs and base stations (BSs) in a two-tiered wireless sensor

network to minimize the average sensor and AP powers. For a uniformly distributed one-dimensional network, they determine the optimal deployment of APs and BSs. They propose one- and two-tiered Lloyd algorithms to numerically optimize node deployment for general scenarios. In [62], the authors study uplink performance of a distributed massive MIMO system in a single cell scenario. They obtain optimal radius of the circular BS antenna array that maximizes the average rate and then numerically show that the circularly distributed massive MIMO can significantly outperform massive MIMO systems with co-located antenna arrays.

In this chapter, in contrast to other works, first we consider cell-free systems with zero-forcing (ZF) detector. For performance criteria, we use sum-throughput and 95%-likely per-user throughput, which is the smallest rate among 95% of the best users. To that end we develop two AP placement algorithms to maximize the sum-throughput and min-throughput of the users in the system. We call them max-sum and max-min algorithms respectively. The max-min optimization does not necessarily provide the optimal 95%-likely throughput. However, we use it as an engineering tool for optimization of this criterion. We formulate the AP location optimization problems in a way that resemble the general sparsity problems, and we leverage this structure to solve the AP placement problems efficiently. In this chapter, we consider uplink transmission. However, similar algorithms can be deployed in the downlink as well. Numerical results show that the proposed max-sum algorithm provides the highest sum-throughput compared with the Lloyd algorithm and random AP locations. However, the simple Lloyd algorithm results in the highest 95%-likely per-user throughput compared with the proposed max-min algorithm and random AP locations.

Next, we consider AP location design problem in small-cell systems. Since the APs do not cooperate in a small-cell system, we consider matched-filtering (MF) detection and propose an iterative algorithm to find optimal location of APs by taking into account the effect of interference in the system. The proposed algorithm is based on the k-means algorithm in

vector quantization that involves assignment and update steps. In [60], authors used the Lloyd algorithm to find the location of APs. In contrast to the Lloyd algorithm, which minimizes the average squared distance between a user and the nearest AP, our proposed method minimizes the average distance between a user and the serving AP as well as the average interference that the user experiences in the system. Numerical results show that we obtain about 1.4-fold improvement in 95%-likely per-user throughput over Lloyd algorithm. However, simplicity of the Lloyd algorithm and its good performance still makes it an attractive approach in selecting the location of distributed APs.

The chapter is organized as follows. Section 4.2 describes the system model and uplink transmission in cell-free massive MIMO. In section 4.3, the problem of AP location design in cell-free massive MIMO is discussed and two AP placement algorithms are proposed. Section 4.4 provides system model and achievable rate analysis in small-cell systems. In Section 4.5, the AP location optimization problem in small-cell systems is investigated. Numerical results are presented in section 4.6. Section 4.7 concludes the chapter.

Throughout this chapter, we use superscripts *, ^{T}, and ^{H} to denote complex conjugate, transpose, and hermitian operations respectively. Uppercase symbols denote matrices and bold symbols denote vectors. $\mathbb{E}(.)$ is the expectation operator.

4.2 Cell-Free Massive MIMO System Model and Achievable Uplink Rate Analysis

4.2.1 System Model

Consider uplink transmission in a geographical area that is covered by M single antenna APs and K single antenna users, where $M \gg K$. An example of a cell-free system is

depicted in Figure 2.1 in Chapter 2. All APs are connected to a network controller (NC) via a perfect backhaul network that is error-free. Each user is served by all APs simultaneously. We assume a flat fading channel model for each orthogonal frequency-division multiplexing (OFDM) subcarrier. The channel coefficient between AP m and user k is given by

$$g_{mk} = \sqrt{\beta_{mk}} h_{mk}, \tag{4.1}$$

where β_{mk} and $h_{mk} \sim \mathcal{CN}(0,1)$ are the large scale and small scale fading coefficients respectively that are mutually independent. The OFDM subcarrier index is omitted in (4.1) for simplicity. Large scale fading coefficient β_{mk}, which includes both path loss and shadow fading effects, changes slowly and can be accurately estimated and tracked. We assume small scale fading coefficients $h_{mk},\ m = 1, \cdots, M,\ k = 1, \cdots, K$ are i.i.d random variables that stay constant during a coherent interval and change to independent values at the next coherent interval. For simplicity, we assume that perfect channel state information is available at the NC.

4.2.2 Achievable Uplink Rate Analysis

In the uplink transmission, users transmit data symbols and the mth AP receives

$$y_m = \sum_{i=1}^{K} \sqrt{\rho_r} g_{mi} s_i + w_m, \tag{4.2}$$

where $w_m \sim \mathcal{CN}(0,1)$ is additive noise, ρ_r is the uplink transmit power, and s_i is the data symbol of user i with unit power $\mathbb{E}\left(|s_i|^2\right) = 1$. We assume data symbols of users are mutually independent and all users transmit with equal power $\mathbb{E}\left(\left|\sqrt{\rho_r} s_i\right|^2\right) = \rho_r$. In Chapter 2, we showed that in downlink transmission of cell-fee massive MIMO the ZF precoder, which has a reasonably low complexity, significantly outperforms conjugate beamforming precoder. Hence, here we consider linear ZF detector at the NC to null out interference from other users

and we assume perfect channel state information (CSI) is available at the NC. The mth AP sends y_m to the NC and the NC forms the ZF detector using the pseudo inverse of channel matrix $A_{ZF} = \left(G^H G\right)^{-1} G^H$ (see [8]) as follows

$$\boldsymbol{r} = A_{ZF}\boldsymbol{y} = \sqrt{\rho_r}\boldsymbol{s} + \left(G^H G\right)^{-1} G^H \boldsymbol{w}, \tag{4.3}$$

where $\boldsymbol{y} = [y_1, \cdots, y_M]^T$, $\boldsymbol{s} = [s_1, \cdots, s_K]^T$, and $\boldsymbol{w} = [w_1, \cdots, w_M]^T$. Considering the fact that worst case noise is Gaussian noise [46], the achievable uplink rate of user k is lower bounded by $R_k = \log\left(1 + \text{SNR}_k\right)$, where the signal-to-noise ratio (SNR) is given by

$$\text{SNR}_k = \frac{\rho_r}{\mathbb{E}\left(\left[(G^H G)^{-1}\right]_{kk}\right)}. \tag{4.4}$$

The asymptotic achievable SNR of the kth user, with unlimited number of APs ($M \to \infty$) and finite number of users K is given by [62]

$$\frac{1}{M}\text{SNR}_k \xrightarrow[M\to\infty]{a.s.} \rho_r \bar{\beta}_k, \tag{4.5}$$

where $\bar{\beta}_k$ is defined as

$$\bar{\beta}_k \triangleq \lim_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} \beta_{mk}. \tag{4.6}$$

To simplify the AP location optimization problems in the next section, we will use the SNR expression in (4.5) as an approximation of the achievable SNR in (4.4).

4.3 Access Point Location Optimization in Cell-Free Massive MIMO

In this section, we propose two algorithms to obtain the optimal location of APs in cell-free systems.

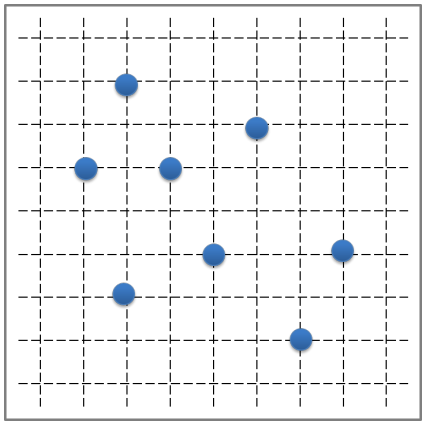

Figure 4.1: APs on grid points.

The optimization of the AP locations with arbitrary AP locations can result in a computationally complex optimization problem. Therefore, to simplify the AP location design problem, we assume APs can only be placed on predefined grid points as shown in Figure 4.1. Let N denote the number of grid points in the area and let us enumerate them in some arbitrary order $n = 1, \cdots, N$ and suppose $N \gg M$. As N increases and the grid gets finer, the accuracy of finding optimal AP locations increases. Now the problem of finding AP locations becomes a combinatorial problem of choosing M out of N locations. Define $\boldsymbol{x} \in \{0, 1\}^N$, where $x_n = 1$ indicates that there is an AP at gird point n. Using the approximation (4.5), the

achievable SNR of users with large number of APs can be expressed as

$$\underbrace{\begin{bmatrix} \text{SNR}_1 \\ \text{SNR}_2 \\ \vdots \\ \text{SNR}_K \end{bmatrix}}_{\boldsymbol{b}} \approx \rho_r \underbrace{\begin{bmatrix} \beta_{11} & \beta_{21} & \cdots & \beta_{N1} \\ \beta_{12} & \beta_{22} & \cdots & \beta_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{1K} & \beta_{2K} & \cdots & \beta_{NK} \end{bmatrix}}_{A} \times \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{x}}, \tag{4.7}$$

where the (n, k)th element of matrix $A \in \mathbb{R}^{K \times N}$, i.e., $[A]_{nk} = \beta_{nk}$, denotes the large scale fading coefficient between the nth grid point and the kth user. Vector $\boldsymbol{b} \in \mathbb{R}^K$ in 4.7 denotes the approximate achievable SNR of all users.

In the next two subsections, we will optimize the AP locations based on two criteria: 1. Max-min criterion, in which we maximize the minimum rate among all users; 2. Max-sum criterion, in which we maximize the sum-throughput of the users. For each criterion, our goal is to find the optimal location of APs by finding location of the M non-zero elements of $\boldsymbol{x}$ in (4.7).

4.3.1 Max-Min Algorithm

Max-min criterion provides equal throughput to all users in the network. Given a target per-user SNR of t, we set $\boldsymbol{b} = t\,[1, \cdots, 1]^T$ and formulate the problem of finding AP locations

on the grid points as follows

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \|\boldsymbol{x}\|_{\ell_0} \\ \text{s.t.} \quad & A\boldsymbol{x} \geq \boldsymbol{b}, \\ & \boldsymbol{x} \in \{0,1\}^N, \end{aligned} \tag{4.8}$$

where $\|\boldsymbol{x}\|_{\ell_0} = |\{n : x_n \neq 0\}|$ is the number of non-zero elements of $\boldsymbol{x}$. The problem in (4.8), with the assumptions of $K < N$ and $N \gg M$, resembles a general sparsity problem in which the aim is to reconstruct the sparse signal $\boldsymbol{x}$ from measurement vector $\boldsymbol{b}$ [63]. In other words, we would like to find locations of M non-zero elements in $\boldsymbol{x}$ for a target SNR vector $\boldsymbol{b}$. The optimization problem (4.8) is non-convex and usually requires combinatorial optimization. Using common compressed sensing techniques, we simplify the problem by replacing ℓ_0-norm with ℓ_1-norm and relaxing the constraints as follows

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \|\boldsymbol{x}\|_{\ell_1} \\ \text{s.t.} \quad & A\boldsymbol{x} \geq \boldsymbol{b}, \\ & 0 \leq \boldsymbol{x} \leq 1, \end{aligned} \tag{4.9}$$

where $\|\boldsymbol{x}\|_{\ell_1} = \sum_{n=1}^{N} |x_n|$ denote that ℓ_1-norm of $\boldsymbol{x}$. The problem (4.9) is a linear programming optimization and can be efficiently solved (see [43, chapter 3.3]). After obtaining the optimal solution of (4.9), we choose the M largest non-zero elements of $\boldsymbol{x}^{\text{opt}}$ as the locations of APs. To capture randomness of the user locations, we place a large number of users ($K \gg M$) in the area when solving (4.9). However, in a cell-free massive MIMO system, the number of served users at each given time is much smaller than the number of APs. Therefore, in the data transmission phase, when we compute the actual achievable rate obtained by the system we choose small values for $K \ll M$.

Note that in (4.9), finding the optimal target per-user SNR t doesn't change the location

of non-zero elements in $\boldsymbol{x}$. In other words, if we replace $\boldsymbol{b}$ with $\alpha\boldsymbol{b}$, the optimal solution will be $\alpha\boldsymbol{x}$. Thus, finding the optimal $\boldsymbol{b}$ is not necessary as long as $\boldsymbol{x}^{\text{opt}}$ contains M non-zero elements.

4.3.2 Max-Sum Algorithm

In this section, our criterion for choosing AP locations is to maximize the sum-throughput. To this end, we will use the following inequity

$$\frac{1}{K}\sum_{i=1}^{K}\log_2\left(1+\text{SNR}_i\right)\overset{(a)}{\leq}\log_2\left(1+\frac{1}{K}\sum_{i=1}^{K}\text{SNR}_i\right) \tag{4.10}$$

to maximize the sum-SNR, which is an upper bound for the actual problem. Note that (a) in (4.10) directly follows from Jensen's inequality. Denote by $\boldsymbol{a}_k^T$ the kth row of matrix A. Then the average achievable SNR of users can be expressed as

$$\frac{1}{K}\sum_{i=1}^{K}\text{SNR}_i=\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{a}_k^T\boldsymbol{x}. \tag{4.11}$$

The AP placement problem with the goal of maximizing the average achievable SNR of users is formulates as

$$\begin{aligned}\max_{\boldsymbol{x}}\quad & \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{a}_k^T\boldsymbol{x}\\ \text{s.t.}\quad & \|\boldsymbol{x}\|_{\ell_0}=M,\\ & \boldsymbol{x}\in\{0,1\}^N.\end{aligned} \tag{4.12}$$

Following the similar procedure as in Section 4.3.1, we relax the AP location optimization problem (4.12) as

$$\begin{aligned}\max_{\boldsymbol{x}}\quad & \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{a}_k^T\boldsymbol{x}\\ \text{s.t.}\quad & M-1<\|\boldsymbol{x}\|_{\ell_1}\leq M,\\ & 0\leq\boldsymbol{x}\leq 1,\end{aligned} \tag{4.13}$$

which is a linear programming optimization and can be efficiently solved. In order to solve (4.13) and to capture the full randomness in simulations, we generate a large number of users ($K \gg M$) according to the density of users in the area, which is similar to Section 4.3.1.

Note that after solving problems (4.9) and (4.13), one can redefine the grid points according to the solution and solve the problems again to further tune the location of APs on a finer grid.

4.4 Small-Cell System Model and Achievable Uplink Rate Analysis

In this section, we provide the small-cell system model and achievable uplink rate of the system. The channel model is the same as (4.1). For large scale fading coefficient between AP m and user k we consider the following model

$$\beta_{mk} = \begin{cases} c_0 & \|\boldsymbol{p}_k - \boldsymbol{q}_m\| \leq r_0 \\ \dfrac{c_1 z_{mk}}{\|\boldsymbol{p}_k - \boldsymbol{q}_m\|^{\gamma}} & \|\boldsymbol{p}_k - \boldsymbol{q}_m\| > r_0 \end{cases}, \tag{4.14}$$

where $\|.\|$ denotes the ℓ_2-norm, $\boldsymbol{q}_m \in \mathbb{R}^2$ and $\boldsymbol{p}_k \in \mathbb{R}^2$ denote the position vectors of AP m and user k respectively, and z_{mk} is the log-normal shadow fading, i.e., $10\log_{10} z_{mk} \sim \mathcal{N}(0, \sigma^2_{\text{shad}})$. In the small-cell system, there is a cell around each AP in which only one user is served by the corresponding AP. This can be viewed as M small cells, each equipped with a single antenna AP. Let k_m be the index of the user that is served by AP m in cell m. In the uplink transmission, users send data symbols and the mth AP receives

$$y_m = \sum_{m'=1}^{M} \sqrt{\rho_r} g_{mk_{m'}} s_{k_{m'}} + w_m, \tag{4.15}$$

where $w_m \sim \mathcal{CN}(0,1)$ is additive noise, ρ_r is the uplink transmit power, and s_i is the data symbol of user i with unit power $\mathbb{E}\left(|s_i|^2\right) = 1$. The mth AP uses MF detector to estimate data

symbol s_{k_m} as

$$\hat{s}_{k_m} = \frac{g^*_{mk_m}}{|g_{mk_m}|} y_m = \underbrace{\sqrt{\rho_r}\,|g_{mk_m}|\,s_{k_m}}_{T_0\text{: desired signal}} + \underbrace{\sum_{m'=1,\,m'\neq m}^{M} \sqrt{\rho_r}\frac{g^*_{mk_m}}{|g_{mk_m}|} g_{mk_{m'}} s_{k_{m'}}}_{T_1\text{: interference}} + v_m, \tag{4.16}$$

where $v_m \sim \mathcal{CN}(0,1)$ is additive Gaussian noise that has the same distribution as w_m. Uplink achievable rate of the user k_m with perfect CSI and MF detector is given by [39]

$$\mathbb{E}\left(\log_2\left(1+\text{SINR}_{k_m}\right)\right) = \frac{1}{\ln 2} e^{\mu_k} \text{Ei}(\mu_k), \tag{4.17}$$

where

$$\text{SINR}_{k_m} = \frac{\rho_r \beta_{mk_m} |h_{mk_m}|^2}{1+\rho_r \sum_{m'\neq m}^{M} \beta_{mk_{m'}}},$$

and

$$\mu_k = \frac{1+\rho_r \sum_{m'\neq m}^{M} \beta_{mk_{m'}}}{\rho_r \beta_{mk_m}},$$

and $\text{Ei}(x) = \int_x^\infty \frac{e^{-t}}{t} \mathrm{d}t$ is the exponential integral. Note that the achievable rate (4.17) is conditioned on the large scale coefficients and depends on the location of APs and users.

4.5 Access Point Location Optimization in Small-Cell Systems

In this section, first, we review the Lloyd algorithm to solve the AP location optimization problem in small-cell systems. Then, we propose an algorithm for AP location design that considers the effect of interference as well.

4.5.1 Lloyd Algorithm

Below, we follow the similar steps as in [60] to show that the location of APs in a small-cell system can be chosen using the Lloyd algorithm. Consider a single user scenario

($K = 1$) in which the user is served by the closest AP denoted by index $m_{\min}$, i.e.,

$$m_{\min} = \underset{n}{\operatorname{argmin}} \ \|\boldsymbol{p} - \boldsymbol{q}_n\|^2 , \tag{4.18}$$

where $\boldsymbol{p}$ is the position vector of the user. In the large scale fading model (4.14), r_0 is much smaller than the dimensions of the area. Hence, the large scale fading coefficient between user and AP $m_{\min}$ can be approximated as

$$\beta_{m_{\min}} \approx \frac{c_1 z_{m_{\min}}}{\left\|\boldsymbol{p} - \boldsymbol{q}_{m_{\min}}\right\|^{\gamma}}. \tag{4.19}$$

The average ergodic rate of the user by averaging over the user position is given by

$$C = \mathbb{E}_{h,\, z,\, \boldsymbol{p}} \left(\log_2 \left(1 + \frac{\rho_r c_1 \left| h_{m_{\min}} \right|^2 z_{m_{\min}}}{\left\|\boldsymbol{p} - \boldsymbol{q}_{m_{\min}}\right\|^{\gamma}} \right) \right), \tag{4.20}$$

where the expected value is taken over small scale fading $h_{m_{\min}}$, shadow fading $z_{m_{\min}}$, and user location $\boldsymbol{p}$. Using Jensen's inequality, the ergodic rate C is lower bounded by

$$C \geq \mathbb{E}_{h,\, z} \left(\log_2 \left(1 + \frac{\rho_r c_1 \left| h_{m_{\min}} \right|^2 z_{m_{\min}}}{\mathbb{E}_{\boldsymbol{p}} \left(\left\|\boldsymbol{p} - \boldsymbol{q}_{m_{\min}}\right\|^2 \right)^{\gamma/2}} \right) \right). \tag{4.21}$$

The AP locations $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_M$ are found by maximizing the lower bound (4.21), which corresponds to minimizing the average minimum distance as follows

$$\min_{\boldsymbol{q}_1, \cdots, \boldsymbol{q}_M} \mathbb{E}_{\boldsymbol{p}} \left(\left\|\boldsymbol{p} - \boldsymbol{q}_{m_{\min}}\right\|^2 \right). \tag{4.22}$$

The average minimum distance in (4.22) is give by

$$\mathbb{E}_{\boldsymbol{p}} \left(\left\|\boldsymbol{p} - \boldsymbol{q}_{m_{\min}}\right\|^2 \right) = \sum_{m=1}^{M} \int_{\boldsymbol{p} \in \mathcal{R}_m} \|\boldsymbol{p} - \boldsymbol{q}_m\|^2 \ f(\boldsymbol{p}) \, , \tag{4.23}$$

where $f(\boldsymbol{p})$ is the probability density function of user location, and region $\mathcal{R}_m$ is defined as

$$\mathcal{R}_m = \left\{ \boldsymbol{p} \,\middle|\, \|\boldsymbol{p} - \boldsymbol{q}_m\| < \|\boldsymbol{p} - \boldsymbol{q}_n\| \, , \ \forall n \neq m \right\}. \tag{4.24}$$

As pointed out in [60], (4.23) is similar to the distortion function in codebook design in vector quantization and Lloyd algorithm is a common approach to solve it [64]. In [65], it has been shown that with infinite number of quantized points, which corresponds to the number APs in our problem $M \to \infty$, the distribution of the quantized points (location of APs) in an r dimensional space is

$$\lambda_r(\boldsymbol{p}) = \frac{f^{r/(r+2)}(\boldsymbol{p})}{\int f^{r/(r+2)}(\boldsymbol{p}')\mathrm{d}\boldsymbol{p}'}, \tag{4.25}$$

where $r = 2$ in our problem.

4.5.2 Proposed Algorithm

As shown in the previous section, Lloyd algorithm is an approach to obtain AP locations in a single user system or in scenarios that interference is negligible. Our goal in this section is to modify the Lloyd algorithm to include the effect of interference. To this end we use the k-means clustering algorithm, which is an iterative clustering algorithm in vector quantization consisting of an assignment (classification) step and an update step [66]. An initial set of random locations is considered for APs and a set of user locations is generated based on the user density function. In the assignment step, the area is partitioned into M cells denoted by $\mathcal{S}_m,\ m = 1, \cdots, M$ by the nearest neighbor rule such that cell S_m contains the users whose distance to AP m, which is defined later, is less than or equal to its distance to any other AP. At the update step, the centroid of each cell is computed which will be the new location of AP at that cell. This procedure is continued until the cell assignments no longer change. In contrast to the Lloyd algorithm, we define a distance function based on the average distance of a user and APs and the average interference that a user experiences. We formulate the AP

location optimization problem in small-cell systems as follows

$$\min_{\substack{\boldsymbol{q}_1,\cdots,\boldsymbol{q}_M \\ \mathcal{S}_1,\cdots,\mathcal{S}_M}} \sum_{m=1}^{M} \int_{\boldsymbol{p}\in\mathcal{S}_m} d(\boldsymbol{p},\boldsymbol{q}_m) f(\boldsymbol{p}) \mathrm{d}\boldsymbol{p}, \tag{4.26}$$

where $d(\boldsymbol{p},\boldsymbol{q}_m)$ is the distance function between AP m and user at location $\boldsymbol{p}$, $f(\boldsymbol{p})$ is the probability density function of user locations, and cell $\mathcal{S}_m$ is defined as

$$\mathcal{S}_m = \{\boldsymbol{p} | d(\boldsymbol{p},\boldsymbol{q}_m) < d(\boldsymbol{p},\boldsymbol{q}_n),\ \forall n \neq m\}. \tag{4.27}$$

We define the heuristic distance function $d(\boldsymbol{p},\boldsymbol{q}_m)$ as follows

$$d(\boldsymbol{p},\boldsymbol{q}_m) \triangleq \frac{\alpha_1}{\mathbb{E}_{s,h,z}\left(|T_0|^2\right)} + \alpha_2 \mathbb{E}_{s,h,z,\boldsymbol{p}}\left(|T_1|^2\right), \tag{4.28}$$

where T_0 and T_1 are, respectively, the desired signal and interference terms given in (4.16). The first term in (4.28) is the inverse of the average desired signal power, where the average is taken over data symbol, small scale fading, and shadow fading coefficients. The second term in (4.28) is the average interference power, where the expectation is taken over small scale fading, shadow fading, data symbols, and location of interfering users $\boldsymbol{p}$. The ratio $\alpha \triangleq \frac{\alpha_2}{\alpha_1}$ is a factor that determines importance of interference over desired signal power. When α is set to zero, (4.26) is the same as Lloyd algorithm. For large values of α, i.e., $\alpha \to \infty$, optimization problem (4.26) only minimizes interference without considering the desired signal power. Thus, α is parameter to be optimized. In the model for large scale fading coefficients (4.14), r_0 is much smaller than the dimensions of the area. Therefore, we can approximate large scale fading coefficient β_{mk} as

$$\beta_{mk} \approx \frac{c_1 z_{mk}}{\|\boldsymbol{p}_k - \boldsymbol{q}_m\|^{\gamma}}. \tag{4.29}$$

After carrying out the expectation operations in (4.28), for $\alpha_1 = \rho_r c_1 \mathbb{E}(z_{mk})$, the distance function can be written as

$$d(\boldsymbol{p},\boldsymbol{q}_m) = \|\boldsymbol{p} - \boldsymbol{q}_m\|^{\gamma} + \alpha \sum_{n=1,n\neq m}^{M} \int_{\boldsymbol{p}'\in\mathcal{S}_n} \frac{1}{\|\boldsymbol{p}' - \boldsymbol{q}_m\|^{\gamma}} f(\boldsymbol{p}')\, \mathrm{d}\boldsymbol{p}'. \tag{4.30}$$

Algorithm 1 AP location design in small-cell systems.

Initialization Step:

Generate N user locations $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_N$ according to the user distribution $f(\boldsymbol{p})$, where $N \gg M$ and initialize AP locations $\boldsymbol{q}_m,\ m = 1, \cdots, M$ using Lloyd algorithm.

Assignment Step:

For all $i = 1, \cdots, N$, assign user i to region $\mathcal{S}_{m_i}$ with the least distance, i.e.,

$$m_i = \underset{n}{\operatorname{argmin}}\, d(\boldsymbol{p}_i, \boldsymbol{q}_n),$$

where $d(\boldsymbol{p}_i, \boldsymbol{q}_n)$ is given in (4.31).

Update Step:

Compute new AP locations $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_M$ as follows

$$\boldsymbol{q}_m^{(i+1)} = \boldsymbol{q}_m^{(i)} - \frac{\delta\,\gamma}{|\mathcal{S}_m|} \sum_{\boldsymbol{p}_i \in \mathcal{S}_m} (\boldsymbol{q}_m - \boldsymbol{p}_i) \left\|\boldsymbol{p}_i - \boldsymbol{q}_m\right\|^{\gamma-2} + \alpha \sum_{n \neq m}^{M} \frac{\delta\,\gamma}{|\mathcal{S}_n|} \sum_{\boldsymbol{p}_j \in \mathcal{S}_n} \frac{(\boldsymbol{q}_m - \boldsymbol{p}_j)}{\left\|\boldsymbol{p}_j - \boldsymbol{q}_m\right\|^{\gamma+2}}.$$

Termination: Repeat assignment and update steps until convergence or until a maximum number of iterations is reached.

For tractability of the AP location optimization problem, we calculate the integrals in (4.30) numerically. To this end, we generate large number of users with locations denoted by $\boldsymbol{p}_i,\ i = 1, \cdots, N$ according to the probability distribution $f(\boldsymbol{p})$ and approximate the distortion function (4.30) as

$$d(\boldsymbol{p}, \boldsymbol{q}_m) \approx \|\boldsymbol{p} - \boldsymbol{q}_m\|^{\gamma} + \alpha \sum_{n=1, n \neq m}^{M} \frac{1}{|\mathcal{S}_n|} \sum_{\boldsymbol{p}_j \in \mathcal{S}_n} \frac{1}{\left\|\boldsymbol{p}_j - \boldsymbol{q}_m\right\|^{\gamma}}. \tag{4.31}$$

The optimization problem (4.26) is solved in the following iterative manner. First, we initialize the AP locations $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_M$ using Lloyd algorithm. In the second step of the algorithm, we keep AP locations $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_M$ constant and assign users to cells $\mathcal{S}_m,\ m =$

$, 1 \cdots, M$ based on the distance function defined in (4.31) such that cell $\mathcal{S}_m$ contains all the users whose distance to AP m is smaller than or equal to its distance to any other AP. In the third step of the algorithm, we hold $\mathcal{S}_1, \cdots, \mathcal{S}_M$ constant and find AP locations that minimize (4.26) for which we employ the steepest descent method as follows

$$\boldsymbol{q}_m^{(i+1)} = \boldsymbol{q}_m^{(i)} - \delta \frac{\partial}{\partial \boldsymbol{q}_m} \int_{\boldsymbol{p} \in \mathcal{S}_m} d(\boldsymbol{p}, \boldsymbol{q}_m) f(\boldsymbol{p}) \mathrm{d}\boldsymbol{p}, \tag{4.32}$$

where δ is the step size and i denotes the iteration index in the steepest descent algorithm. In order to numerically solve for $\boldsymbol{q}_m,\ m = 1, \cdots, M$, we approximate the integral in (4.32) as

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{q}_m} \int_{\boldsymbol{p} \in \mathcal{S}_m} d(\boldsymbol{p}, \boldsymbol{q}_m) f(\boldsymbol{p}) \mathrm{d}\boldsymbol{p} \approx & \frac{\gamma}{|\mathcal{S}_m|} \sum_{\boldsymbol{p}_i \in \mathcal{S}_m} (\boldsymbol{q}_m - \boldsymbol{p}_i) \left\| \boldsymbol{p}_i - \boldsymbol{q}_m \right\|^{\gamma-2} \\ & + \alpha \sum_{n \neq m}^{M} \frac{\gamma}{|\mathcal{S}_n|} \sum_{\boldsymbol{p}_j \in \mathcal{S}_n} \frac{\left(\boldsymbol{p}_j - \boldsymbol{q}_m\right)}{\left\| \boldsymbol{p}_j - \boldsymbol{q}_m \right\|^{\gamma+2}}. \end{aligned} \tag{4.33}$$

The steps of finding AP locations are summarized in Algorithm 1.

4.6 Numerical Results

4.6.1 Cell-Free Massive MIMO

We consider a square area of size 2×2 km^2 with M APs and K users. The area is wrapped around to avoid boundary effects. We select equally spaced grid points in the area. In cell-free massive MIMO, we use three-slope path loss model [52] for these coefficients as follows

$$\beta_{mk} = \begin{cases} 1 & d_{mk} \leq r_0 \\ \dfrac{c_1}{d_{mk}^2} & r_0 \leq d_{mk} \leq r_1 \\ \dfrac{c_2}{d_{mk}^{3.5}} & d_{mk} \geq r_1 \end{cases}, \tag{4.34}$$

where d_{mk} is the distance between AP m and user k in kilometers. We choose $d_0 = 0.01$km and $d_1 = 0.05$km. [1] Path loss parameters c_1 and c_0 in (4.34) are chosen such that path loss remains continuous at boundary points: $c_1 = r_0^2,\ c_2 = r_0^2 r_1^{1.5}$. We use Gaussian mixture model for user locations as

$$f(\boldsymbol{p}) = \sum_{i=1}^{3} p_i \mathcal{N}(\boldsymbol{p}|\boldsymbol{\mu}_i, \sigma_i I), \tag{4.35}$$

with equal weights $p_i = 1/3$ and mean locations $\boldsymbol{\mu}_1 = [-0.5, 0]^T$ km, $\boldsymbol{\mu}_2 = [0, 0.5]^T$ km, and $\boldsymbol{\mu}_3 = [0.5, -0.5]^T$ km, and variances $\sigma_i = 200,\ i = 1, 2, 3$. A realization of this distribution is depicted in Figure 4.2.

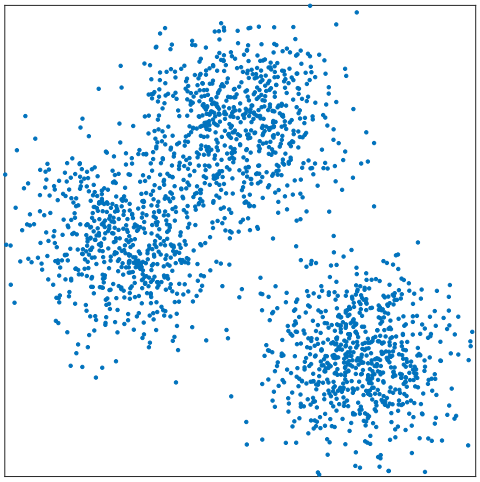

Figure 4.2: A realization of user locations for Gaussian mixture distribution in (4.35).

Note that each run of the AP location optimization algorithms (in cell-free and small-cell systems) results in different AP locations. Therefore, in all experiments, we run each algorithm multiple times and choose the solution that give us the best performance.

Experiment 1: In this experiment, we provide a comparison between achievable rates of cell-free massive MIMO obtained by randomly placed APs, the Lloyd algorithm, and the proposed algorithms in Sections 4.3.1 and 4.3.2. Figures 4.3 and 4.4, respectively, show the

[1] Shadow fading is ignored in large scale fading model for simplicity.

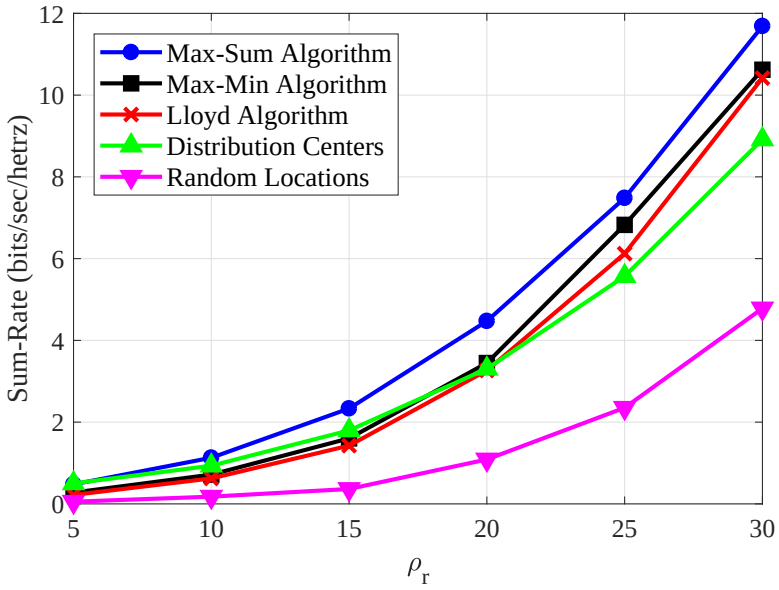

Figure 4.3: Sum-rate versus transmit power ρ_r in cell-free massive MIMO for $M = 32$, $K = 4$, and $N = 50 \times 50$.

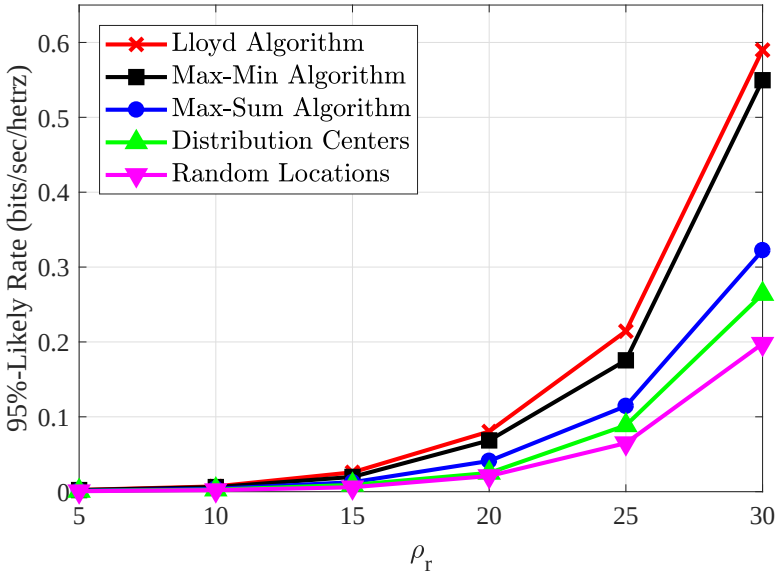

Figure 4.4: 95%-likely per-user rate versus transmit power ρ_r in cell-free massive MIMO for $M = 32$, $K = 4$, and $N = 50 \times 50$.

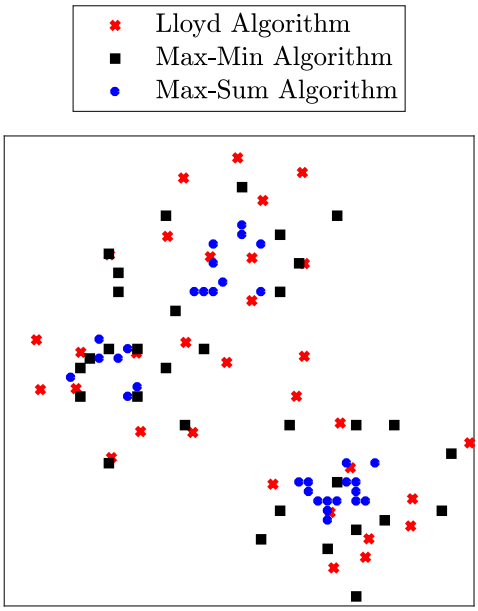

Figure 4.5: AP locations obtained by different AP placement algorithms in cell-free massive MIMO for $M = 32$, $K = 4$, $\rho_r = 20$dB, and $N = 50 \times 50$.

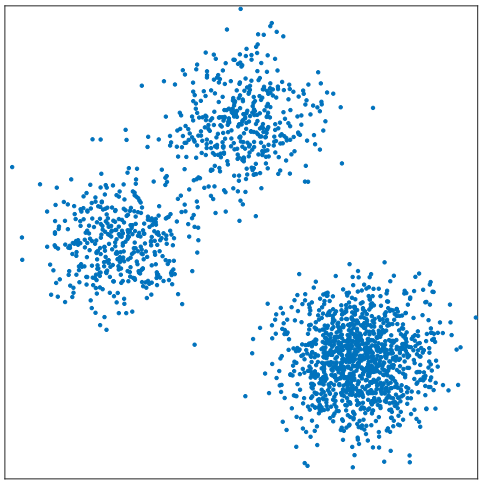

Figure 4.6: A realization of user locations for Gaussian mixture distribution with mixture weights of $0.2,\ 0.2,\ 0.6$.

sum-throughput and 95%-likely per-user throughput versus transmit power in cell-free massive MIMO for $M = 32$, $K = 4$, $N = 50 \times 50$. The 95%-likely throughput is the smallest rate among 95% of the best users. In the figures, the curve "Distribution Centers" is obtained

by placing APs at the centers of the mixture model $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and $\boldsymbol{\mu}_3$. The curve "Random Locations" denotes a cell-free system with randomly placed APs. Rates achieved by the Lloyd algorithm are also plotted in the figures. It is to be noted that the max-sum algorithm provides the best sum-rate among all algorithms in Figure 4.3. Max-min algorithm in Figure 4.4, provides higher 95%-likely rate compared with random AP locations and the case where APs are located at distribution centers. However, the Lloyd algorithm outperforms max-min algorithm even though it is designed for a small-cell system with single user. Part of this loss is due to the fact that in the Lloyd algorithm, APs can be placed in any location in the area, whereas in the max-min algorithm, APs can only be placed on the grid points. It is expected that as the number of grid points N increases, the gap between the two curves decreases.

Figure 4.5 shows the AP locations obtained by max-min and max-sum algorithms in Sections 4.3.1 and 4.3.2 for $M = 32$, $K = 4$, $N = 50 \times 50$, and $\rho_r = 20$dB. It can be observed that with the sum-rate criterion, the APs are located closer to the distribution centers. Whereas, with the min-rate criterion, the APs are more dispersed over the area. This observation is consistent with intuitive reasoning that the systems with co-located APs, in general, can provide higher sum-throughput compared with the systems with distributed APs, while the distributed APs provide larger 95%-likely and minimum rates.

4.6.2 Small-Cell

We consider a wrapped around square area of size 2×2 km^2 with M APs. Each AP serves only one randomly selected user in its cell. Therefore, at any given time, M APs serve M users. For large scale fading coefficient we use model (4.14) with path loss exponent $\gamma = 2$. However, for simplicity, we ignore shadow fading z_{mk} in (4.14). For $\|\boldsymbol{p}_k - \boldsymbol{q}_m\| \leq r_0$ we use

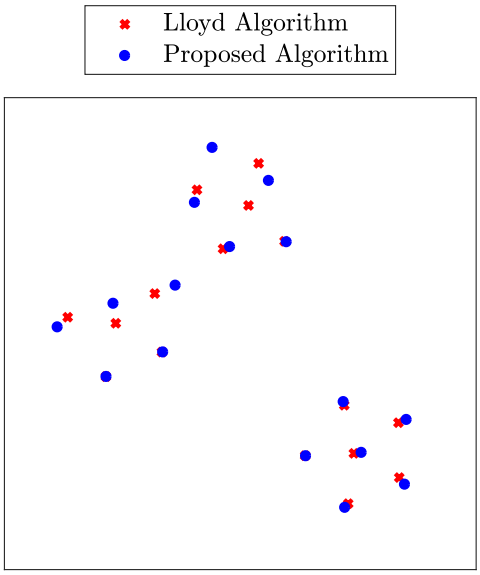

Figure 4.7: AP locations obtained by the proposed and Lloyd algorithms in small-cell system for $M = K = 16$, $\alpha = 1.6 \times 10^8$.

COST-231 Hata propagation model

$$10 \log_{10} c_1 = -46.3 - 33.9 \log_{10} f + 13.82 \log_{10} h_B + (1.1 \log_{10} f - 0.7) h_R - (1.56 \log_{10} f - 0.8), \tag{4.36}$$

where $f = 1900$ MHz is the carrier frequency, $h_B = 15$ m is the BS antenna height, and $h_R = 1.65$ m is the user antenna height. Path loss parameter c_0 in (4.14) is given by

$$10 \log_{10} c_0 = 10 \log_{10} c_1 - 20 \log_{10}(d_0). \tag{4.37}$$

We choose $d_0 = 0.01$km. We use Gaussian mixture model for user locations given in (4.35) with weights $p_1 = 0.2$, $p_2 = 0.2$, $p_3 = 0.6$, and mean locations $\boldsymbol{\mu}_1 = [-0.5, 0]^T$ km, $\boldsymbol{\mu}_2 = [0, 0.5]^T$ km, and $\boldsymbol{\mu}_3 = [0.5, -0.5]^T$ km, and variances $\sigma_i = 150,\ i = 1, 2, 3$. A realization of this distribution is shown in Figure 4.6. User transmit power is set to $\rho_r = 200$ mW. The noise variance at the receiver is $\sigma_w^2 = 290 \times \kappa \times B \times NF$, where κ, B, and NF are Boltzmann constant, bandwidth (20 MHz), and noise figure (9 dB) respectively. The maximum number

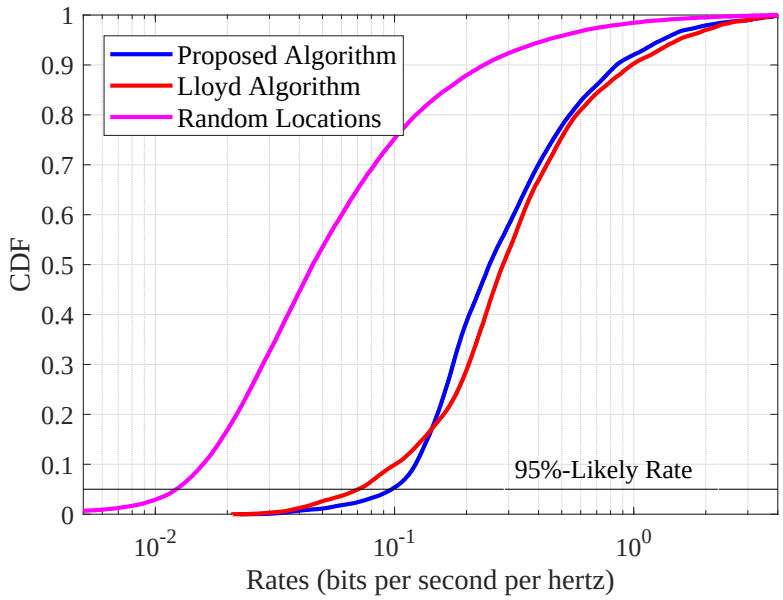

Figure 4.8: CDF of achievable per-user rates in small-cell system for $M = K = 16$, $\alpha = 1.6 \times 10^8$.

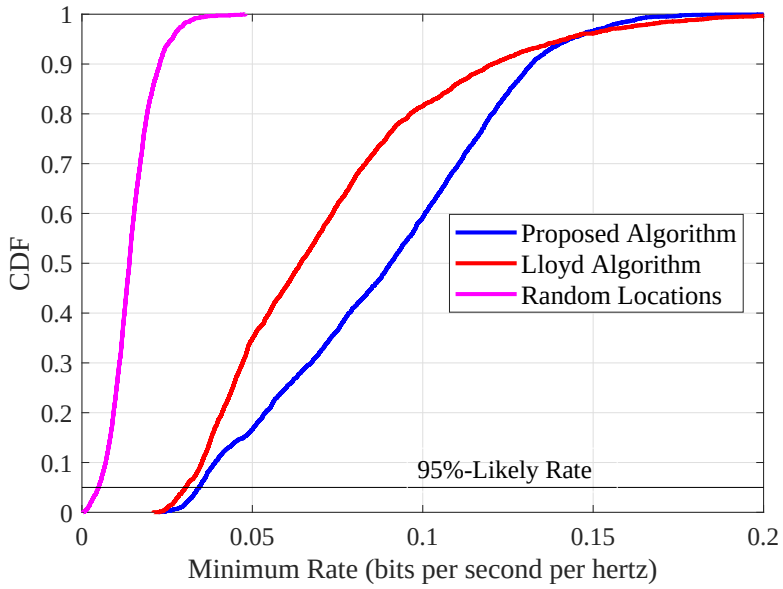

Figure 4.9: CDF of the worst rate in small-cell system for $M = K = 16$, $\alpha = 1.6 \times 10^8$.

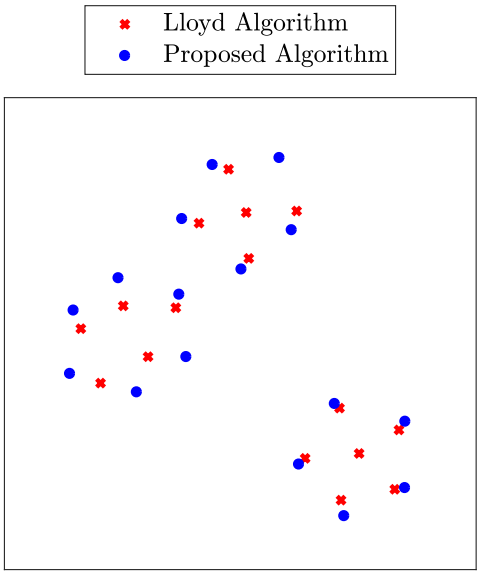

Figure 4.10: AP locations obtained by the proposed and Lloyd algorithms in small-cell system for $M = K = 16$, $\alpha = 5 \times 10^8$.

of iterations in the proposed algorithm is set to 50.

Experiment 2: In this experiment, we compare achievable rates obtained by our proposed method and the Lloyd algorithm in small-cell systems. Figure 4.7 shows the AP locations obtained by the Lloyd algorithm and the proposed algorithm for $M = K = 16$ and $\alpha = 1.6 \times 10^8$. It is to be noted that the distance between APs obtained by the proposed algorithm is more than that of the Lloyd algorithm, which is due to effect of interference in the proposed algorithm.

Figure 4.8 shows the CDF of per-user rates given in (4.17) for $M = K = 16$ and $\alpha = 1.6 \times 10^8$ obtained by the proposed method and the Lloyd algorithm. The achievable rate of a system with randomly located APs is also plotted in the figure. Figure 4.9 presents the similar results for the worst rate (minimum rate) among users for given realization of large-scale coefficients. It can be observed that the proposed algorithm provides about 1.4-fold in 95%-likely rate and improves the worst rate of the system over the Lloyd algorithm and that

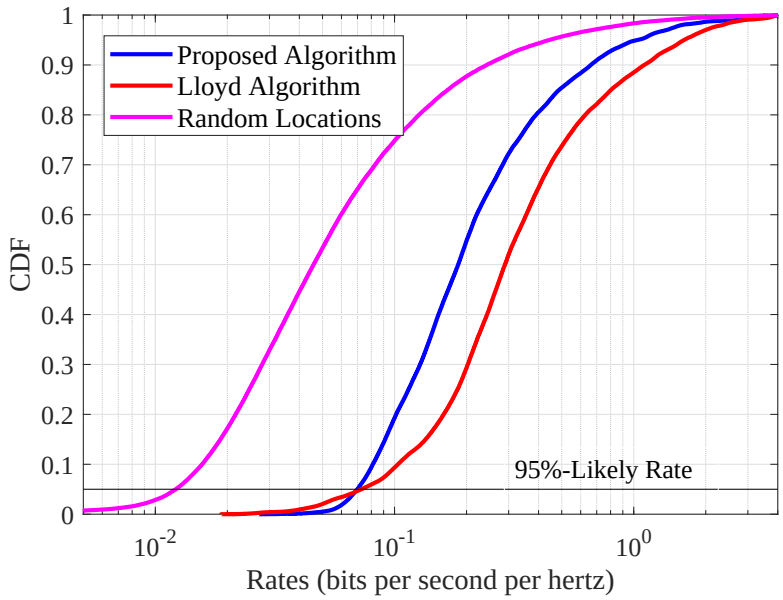

Figure 4.11: CDF of achievable per-user rates in small-cell system for $M = K = 16$, $\alpha = 5 \times 10^8$.

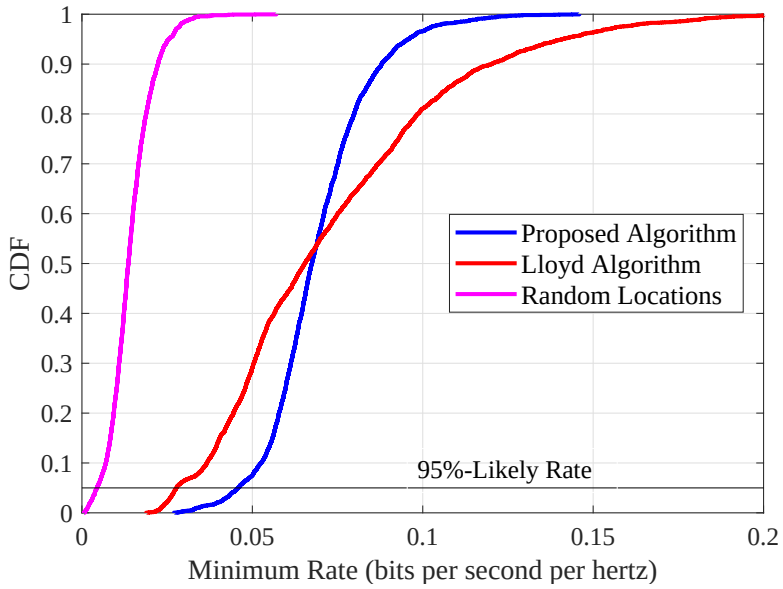

Figure 4.12: CDF of the worst rate in small-cell system for $M = K = 16$, $\alpha = 5 \times 10^8$.

both algorithms outperform the system with randomly located APs.

Figures 4.10, 4.11, and 4.12 show the similar results for $M = K = 16$ and $\alpha = 5\times10^8$. For smaller values of α, location of APs obtained by our algorithm is similar to that of the Lloyd algorithm. However, when we increase α, interference becomes the dominant term in (4.28) and the distance between APs increases. Therefore, α is a parameter that needs to be optimized in the proposed algorithm.

4.7 Conclusion

In this chapter, we investigated AP placement problem in cell-free massive MIMO and small-cell systems. For each system, we provided AP location design algorithms to improve the system throughput. In cell-free massive MIMO, two algorithms are proposed based on two criteria: maximizing the sum-throughput and minimum-throughput. The AP location optimization problems in the cell-free massive MIMO are transformed into linear programming problems that can be solved efficiently. For the sum-throughput criterion in cell-free massive MIMO, the proposed algorithm provides significant improvement over a cell-free system with randomly located users. However, in cell-free system, the Lloyd algorithm can provide higher 95%-likely per-user rates. In small-cell system the proposed algorithm is based on the k-means clustering algorithm with a cost function that is based on the distance between APs and the served users as well as the average interference from other small cells. The proposed algorithm improves 95%-likely and minimum rates over the Lloyd algorithm. However, simplicity and good performance of the Lloyd algorithm makes it an attractive approach to design AP locations in small-cell systems.

4.8 Acknowledgment

This chapter, in part is currently being prepared for submission for publication of the material. The dissertation author is the primary investigator and author of this material and B. D. Rao supervised the research.

Chapter 5

Semi-blind Channel Estimation

Motivated by recent developments in time-division duplex massive multiple-input multiple-output (MIMO) systems, this chapter investigates semi-blind channel estimation for multiuser MIMO systems. A tractable expectation-maximization (EM) algorithm is derived for semi-blind channel estimation by assuming a Gaussian distribution for the unknown data symbols, which improves channel estimates even when the data symbols are drawn from a finite constellation, such as quadrature phase-shift keying. An alternate EM algorithm is also derived by employing suitable priors on the channel coefficients and it is shown to outperform the EM algorithm with no priors in the low signal-to-noise ratio (SNR) regime. To improve the estimation performance for discrete constellations, another EM based channel estimation algorithm is developed based on a Gaussian mixture model (GMM) for the unknown data symbols.

To analytically understand the performance of the semi-blind scheme, Cramer-Rao bounds (CRBs) for semi-blind channel estimation are derived for deterministic and stochastic (Gaussian) data symbol models. To get insight into the behavior of a massive MIMO system, the asymptotic behavior of the CRBs as the number of antennas at the base station (BS) grows is analyzed. The numerical results show the benefits of semi-blind estimation algorithms as measured by the mean squared error. The EM algorithm with a Gaussian prior provides superior channel estimates compared to the EM algorithm with a GMM prior in low SNR regime. However, the latter one outperforms the EM algorithm with Gaussian prior as the SNR or as the number antennas at the BS increases. Furthermore, the performance of the semi-blind estimators become closer to the genie aided maximum likelihood estimator based on known data symbols as the number of antennas increases. This result is consistent with the asymptotic analysis of the two CRBs indicating that semi-blind channel estimation for massive MIMO systems is very promising.

5.1 Introduction

To achieve the expected high capacity gains in multiple-input multiple-output (MIMO) systems, channel state information (CSI) is an important factor and in practical systems is determined by the accuracy of the channel estimates. A simple method to estimate channel coefficients is via training or pilot sequences [40]. As the number of users in time-division duplex (TDD) systems increases, the length of the pilot sequences required to estimate the channel accurately increases, which results in lower spectral efficiency. An approach for improving the quality of the estimates of the channel coefficients is to use the information in the unknown data symbols instead of only using the pilot sequences [67–73]. With this approach, one has the option of obtaining more accurate channel estimates or utilizing smaller number of pilot symbols to estimate the channel coefficients with the same accuracy. In frequency-division duplex (FDD) systems, more accurate channel estimates results in better detection of unknown data symbols at the receiver. In TDD systems, where uplink and downlink physical channels are assumed to be reciprocal [17], better channel estimation not only leads to better uplink detection but it also helps the base station (BS) to form more accurate downlink precoders. In massive MIMO systems the emphasis in transmission protocol is mostly on TDD rather than FDD [15], [10]. Therefore, massive MIMO systems benefit from semi-blind channel estimation in both uplink and downlink transmissions making semi-blind estimation more attractive for next generation wireless systems. This motivates our re-examination of semi-blind channel estimation with an eye towards massive MIMO systems.

Another data aided channel estimation scheme is blind channel estimation based on the received data signal only. With the blind estimation, the channel can only be identified within some ambiguities, whereas in the semi-blind estimation the channel coefficients can be completely identified under certain conditions [67]. In addition, adding a few training

sequences significantly improves channel estimation quality over the blind estimation [69, Figure 1]. Therefore, in this work, we focus on the semi-blind channel estimation using both training and data symbols to estimate the channel.

Semi-blind channel estimation has been investigated in several papers, e.g., [67–72] and references therein. In [67], the authors study the conditions under which the channel and the data signals are blindly and semi-blindly identifiable for an underdetermined MIMO system. They obtain blind and semi-blind channel estimates based on an expectation-maximization (EM) algorithm in the frequency domain and utilize a discrete random variable model for the unknown data. In [68], two iterative channel estimators based on the EM algorithm are proposed. In [69], Cramer-Rao bounds (CRBs) for semi-blind, blind and training sequence based channel estimation for single-input multiple-output (SIMO) systems are studied and compared. In [70], the authors study two semi-blind channel estimators for SIMO systems based on maximum likelihood (ML) estimation with deterministic and Gaussian models. The asymptotic performances of the estimators in [69], [70] are studied when the length of the training sequences and data sequences grow infinitely large. In [72], a semi-blind estimation technique for MIMO systems is introduced, which uses an iterative two-level optimization loop to jointly estimate channel coefficients and data symbols.

In contrast to the previous works, we investigate four different semi-blind channel estimation schemes based on the EM algorithm. In the first algorithm, we use Gaussian distribution for the unknown data symbols which leads to a simple closed form solution in the E-step of the EM algorithm. Fortunately, numerical results illustrate that the performance of the algorithm with symbols drawn from a finite constellation such as quadrature phase-shift keying (QPSK) and Gaussian priors are similar making the assumption practically relevant. An alternate EM algorithm is derived based on utilizing additional channel priors and this leads to enhancement in performance. To improve the performance of channel estimation for

the case when data symbols are drawn from a discrete constellation, we consider a heuristic algorithm by demapping the conditional mean of the data symbols to the nearest constellation point in the EM algorithm. Numerical results illustrate that in the high signal-to-noise ratio (SNR) regime, this heuristic algorithm considerably outperforms the EM algorithm with Gaussian prior. Motivated by this observation, to provide support for the procedure, we pursue deriving an analytically rigorous EM algorithm assuming a Gaussian mixture model (GMM) for data symbols which achieves a performance similar to the heuristic algorithm. We study effects of semi-blind channel estimation in regular MIMO and massive MIMO systems and compare the performance of the semi-blind algorithms with two ML estimators; one based on only the pilot sequences and the other one assuming all the data symbols are known at the receiver. For performance criteria we use *mean squared error* (MSE) and *symbol error rate*. Numerical results indicate that the semi-blind estimation schemes provide better channel estimates compared with channel estimation based on training sequences only. To analytically understand the performance of the methods, we derive the CRBs, *deterministic* and *stochastic*, with two common assumptions on the transmitted data symbols [74–76]. In the deterministic CRB, we assume the data symbols are unknown deterministic values. Whereas in the stochastic CRB, we assume the data symbols are drawn from a Gaussian distribution. In previous works, asymptotic behavior of CRB or estimators is studied when number of data symbols grows infinitely large [69], [70], [74]. In this chapter, we study behavior of CRBs in massive MIMO systems with unlimited number of antennas at the BS. Results indicate that as the number of antennas increases, the deterministic CRB converges to the CRB of a system with known data symbols and the stochastic CRB converges to the CRB of a system with orthogonal pilot sequences of size equal to the whole transmission block (training plus data). Numerical experiments are presented to support the analysis. For TDD systems, the uplink channel estimation also impacts the downlink beamforming performance. In the numerical

experiments, we also demonstrate the benefits of semi-blind channel estimation for downlink beamforming. Numerical results show that semi-blind channel estimation results in a significant improvement in downlink achievable rate compared to the ML training based on pilots alone.

The chapter is organized as follows. In Section 5.2, we describe the system model and two ML estimators. In Section 5.3, four semi-blind channel estimation algorithms based on the EM algorithm are derived followed by a discussion on the effects of channel estimation in downlink beamforming. In Section 5.4, CRBs under two sets of assumptions on the data symbols are obtained and their asymptotic behavior with respect to the number of antennas is analyzed. Finally, numerical results are presented in Section 5.5 and we conclude the chapter in Section 5.6.

Throughout the chapter, we use superscript ^{H} to denote conjugate transpose, ^{T} to denote transpose, * to denote complex conjugate, uppercase symbols to denote matrices, and bold symbols to denote vectors. $\mathbb{E}(.)$ and $\mathrm{tr}(.)$ are the expectation and trace operators respectively. The spectral norm of matrix A is denoted by $\|A\|_2$. We use $\mathrm{diag}(\boldsymbol{a})$ to denote the diagonal matrix whose diagonal entries are entries from the vector $\boldsymbol{a}$.

5.2 System Model and Channel Estimation

5.2.1 System Model

We consider a single cell with a BS equipped with M antennas and randomly located K single antenna users, where $M \geq K$. We study uplink transmission in a communication system with TDD protocol. However, similar estimation techniques can also be applied to a system with FDD protocol. We consider a flat fading channel model for each orthogonal

frequency-division multiplexing (OFDM) subcarrier. The OFDM subcarrier index is omitted for simplicity. The channel matrix between the BS and users is given by

$$G = HB^{1/2}, \tag{5.1}$$

where $H \in \mathbb{C}^{M\times K}$ is a matrix representing small scale fading and $B \in \mathbb{R}^{K\times K}$ is a diagonal matrix that is given by

$$B = \begin{bmatrix} \beta_1 & & 0 \\ & \ddots & \\ 0 & & \beta_K \end{bmatrix}, \tag{5.2}$$

where β_k is the large scale fading coefficient between BS and user k that accounts for the path loss and shadow fading. We assume columns of H are independent from B and are i.i.d circularly-symmetric complex normal vectors $\boldsymbol{h_k} \sim \mathcal{CN}(0, I_M)$. We consider a time block fading model, where channel vectors $\boldsymbol{h}_k,\ k = 1, \cdots, K$ are constant during a block of N symbols and change to independent values at the next coherence block. The model can be simplified down to a point-to-point MIMO system by assuming $\beta_1 = \cdots = \beta_K$.

We further assume perfect channel reciprocity, i.e., the uplink and downlink channel coefficients are the same. Although the propagation channel itself is reciprocal, the hardware chains in the transmit and receive sides are not identical. Thus, in practice, channel calibration is required to enable exploiting channel reciprocity. Since the parameters that result in the mismatch between uplink and downlink channels change slowly, they can be estimated using small overhead signaling [17], [77].

We consider uplink transmission in which users send L known pilot sequences followed by $(N-L)$ unknown data symbols. The uplink signal received by BS at time n is given by

$$\boldsymbol{y}[n] = G\boldsymbol{s}[n] + \boldsymbol{v}[n], \quad n = 0, \cdots, N-1, \tag{5.3}$$

where $\boldsymbol{s}[n] \in \mathbb{C}^{K\times 1}$ for $n = 0, \cdots, L-1$ are known pilot sequences and $\boldsymbol{s}[n] \in \mathbb{C}^{K\times 1}$ for $n = L, \cdots, N-1$ are the unknown data symbols with unit power $\mathbb{E}\left(\boldsymbol{s}[n]\boldsymbol{s}[n]^H\right) = I_K$ and $\boldsymbol{v}[n] \sim \mathcal{CN}(0, \sigma_v^2 I_M)$ is additive Gaussian noise. Let $S_p = [\boldsymbol{s}[0], \cdots, \boldsymbol{s}[L-1]]$ and $S_d = [\boldsymbol{s}[L], \cdots, \boldsymbol{s}[N-1]]$ denote, respectively, the known pilot sequences and data symbols in a channel coherence time. Similarly, let Y_p and Y_d represent the row stacked received training output and received data signals respectively. The complete transmit and received symbols are given by $S = [S_p\ S_d]$ and $Y = [Y_p\ Y_d]$ respectively.

5.2.2 Performance Metric

Since we are estimating a vector, one of the scalar performance criteria we use to demonstrate the improvement in channel estimation, is MSE, i.e., trace of the error covariance matrix. The following Lemma motivates the use of this criterion as our performance measure.

Lemma 5.1 ([71, Lemma 1]). *Let $A, B \in \mathbb{C}^{n\times n}$ be positive definite matrices and let $A \geq B$, i.e., $\boldsymbol{u}^H A\boldsymbol{u} \geq \boldsymbol{u}^H B\boldsymbol{u}, \forall \boldsymbol{u} \in \mathbb{C}^{n\times 1}$. Then tr $(A) =$ tr $(B) \iff A = B$.*

Assume A and B are the error covariance matrix and the CRB matrix respectively. As the MSE of channel matrix becomes closer to the trace of CRB, then the error covariance matrix approaches the CRB. Thus, we will use MSE in most of the numerical results to illustrate the accuracy of the channel estimates.

5.2.3 ML Estimators

In this section, we provide reviews of a commonly used ML estimator utilizing only the training symbols as well as a full data ML estimator with the assumption of perfect data estimation which will serve as an upper bound on the performance of semi-blind estimator.

Training Pilot Sequences

A simple conventional method to estimate channel coefficients is with the aid of known training sequences. The ML estimate of the channel matrix G based on the pilot sequences (S_p) is given by

$$\hat{G}_{\text{ML}}^{\text{tr}} = \left(Y_p S_p^H\right)\left(S_p S_p^H\right)^{-1}. \tag{5.4}$$

In [78], it has been indicated that the training sequences that minimize the MSE subject to the total transmit power, are orthogonal sequences, i.e., $S_p S_p^H = LI_K$. The corresponding MSE is given by

$$\mathbb{E}\left(\|G - \hat{G}_{\text{ML}}^{\text{tr}}\|_F^2\right) = \frac{MK\sigma_v^2}{L}. \tag{5.5}$$

To obtain a reliable channel estimate with this method, large number of training sequences is required which reduces the achievable throughput of the system.

Full Data

The upper bound on the performance of semi-blind channel estimation is the case when all data symbols are known, i.e., genie aided. In this case, all N symbols (S) are assumed to be known at the BS which provides a lower bound on the achievable MSE. This will serve as an estimator that the semi-blind procedure can aspire to imitate. The channel estimate based on all data symbols is denoted by $\hat{G}_{\text{ML}}^{\text{full}}$ and is given by

$$\hat{G}_{\text{ML}}^{\text{full}} = \left(YS^H\right)\left(SS^H\right)^{-1}, \tag{5.6}$$

and the corresponding MSE can be computed as

$$\mathbb{E}\left(\|G - \hat{G}_{\text{ML}}^{\text{full}}\|_F^2\right) = M\sigma_v^2 \,\text{tr}\left(\mathbb{E}\left(\left(SS^H\right)^{-1}\right)\right). \tag{5.7}$$

Remark 5.1. *Note that in multi-cell massive MIMO systems, usually, the length of the training sequences is not long enough to separate the channels of all users which results in pilot contamination effect [20]. The main goal of this chapter, however, is to first understand the phenomenology of semi-blind estimation in single cell scenarios with large scale antenna arrays. Therefore, in numerical results, we consider orthogonal pilot sequences and leave the pilot contamination effect in multi-cell scenarios for future work.*

5.3 Semi-blind Channel Estimation and Downlink Beamforming

In this section, we develop four semi-blind channel estimation schemes based on the EM algorithm. Computational complexities of these estimators are compared with that of the ML estimators in Section 5.2.3. The effect of semi-blind estimation in downlink beamforming is also presented.

5.3.1 Uplink Semi-blind Channel Estimation

The ML estimate of G based on both received training and data signals is given by

$$\hat{G}_{\mathrm{ML}} = \underset{G}{\operatorname{argmax}} \ \log p\left(Y|G\right). \tag{5.8}$$

Obtaining a closed form solution to this incomplete data problem is known to be hard [79]. In [71], an iterative algorithm has been proposed to solve the problem.

EM Algorithm with Gaussian Prior

An alternative way to solve the problem in (5.8) is to use the EM algorithm. The EM algorithm is an iterative algorithm where the channel estimate is updated ($\hat{G}_{\ell+1}$) based on the

old estimate ($\hat{G}_\ell$) in the following manner:

$$\hat{G}_{\ell+1} = \underset{G}{\text{argmax}}\, \mathbb{E}_{p(S_d|Y,\hat{G}_\ell)} \left(\log p\left(Y, S_d|G\right)\right), \tag{5.9}$$

where (Y, S_d) is the complete data. As can be seen from (5.9), the computation involves an expectation evaluation (E-step) and a maximization (M-step). At each iteration of the EM algorithm, the likelihood function increases until a local maximum is achieved [55]. In the E-step, the expectation utilizes $p(S_d|Y, \hat{G}_\ell)$, the conditional density of S_d given Y and the old estimates of the unknown parameters. The expected value of the log likelihood function of the received signal, i.e., $\mathbb{E}_{p(S_d|Y,\hat{G}_\ell)} \left(\log p\left(Y, S_d|G\right)\right)$, is given by

$$\mathcal{L} = \text{const} - \sum_{n=0}^{L-1} \frac{1}{\sigma_v^2} \left\| \boldsymbol{y}[n] - G\boldsymbol{s}[n] \right\|^2 - \sum_{n=L}^{N-1} \frac{1}{\sigma_v^2} \mathbb{E}_{p(S_d|Y,\hat{G}_\ell)} \left(\left\| \boldsymbol{y}[n] - G\boldsymbol{s}[n] \right\|^2 \right). \tag{5.10}$$

Carrying out the maximization (M-step), one can show that the channel estimation at the $(\ell+1)$th iteration is given by [67]

$$\hat{G}_{\ell+1} = \left(Y_p S_p^H + Y_d\, \mathbb{E}\left(S_d \middle| \hat{G}_\ell, Y \right)^H \right) \left(S_p S_p^H + \mathbb{E}\left(S_d S_d^H \middle| \hat{G}_\ell, Y \right) \right)^{-1}. \tag{5.11}$$

To compute the updated estimate (5.11), there are a few expectations that are needed, and the details of the expectation computation needed to complete the E-step of the algorithm has not been mentioned explicitly in [67]. Using a discrete random variable model such as QPSK modulation for data symbols leads to excessively complex E-step which grows exponentially with K (see Appendix 5.8.A for details). Thus, for tractability of the problem, we assume that the data symbols are Gaussian, i.e., $\boldsymbol{s}[n] \sim \mathcal{CN}\left(0, I_K\right),\ n = L, \cdots, N-1$. Given G, S_d and Y are jointly Gaussian. Thus, $\mathbb{E}(S_d|\hat{G}_\ell, Y)$ and $\mathbb{E}(S_d S_d^H|\hat{G}_\ell, Y)$ in (5.11), can be computed from the conditional density of circularly symmetric Gaussian random vectors. The E-step of the EM algorithm based on the estimates at the ℓth iteration is given by

$$\begin{aligned} \boldsymbol{\mu}_n^\ell &= \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 I_K \right)^{-1} \hat{G}_\ell^H \boldsymbol{y}[n], \\ \Sigma^\ell &= \sigma_v^2 \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 I_K \right)^{-1}, \end{aligned} \tag{5.12a}$$

and the M-step at $(\ell+1)$th iteration is given by

$$\hat{G}_{\ell+1} = \left(Y_p S_p^H + \sum_{n=L}^{N-1} \boldsymbol{y}[n](\boldsymbol{\mu}_n^\ell)^H \right) \left(S_p S_p^H + \sum_{n=L}^{N-1} \left(\boldsymbol{\mu}_n^\ell (\boldsymbol{\mu}_n^\ell)^H + \Sigma^\ell \right) \right)^{-1}. \tag{5.12b}$$

Derivation of the E- and M-steps are presented in Appendix 5.8.B.

EM Algorithm with Channel Priors

We now present a variant by using priors on the channel coefficients. Given the large scale fading coefficients B, channel matrix G is Gaussian. We can add Gaussian priors for the channel coefficients to the likelihood function as follows

$$\mathcal{L} = \mathbb{E}_{p(S_d|Y,\hat{G}_\ell)} \left(\log p\left(Y, S_d|G\right) \right) + \log p\left(G|B\right). \tag{5.13}$$

In Appendix 5.8.C, it is shown that the E-step of the EM algorithm with channel priors is the same as in (5.12a) and the M-step is modified, in which the likelihood function is maximized for G and B separately at each iteration. This results in the following update procedure.

$$\hat{G}_{\ell+1} = \left(Y_p S_p^H + \sum_{n=L}^{N-1} \boldsymbol{y}[n](\boldsymbol{\mu}_n^\ell)^H \right) \hat{B}_\ell \left(S_p S_p^H \hat{B}_\ell + \sum_{n=L}^{N-1} \left(\boldsymbol{\mu}_n^\ell (\boldsymbol{\mu}_n^\ell)^H + \Sigma^\ell \right) \hat{B}_\ell + \sigma_v^2 I_K \right)^{-1}, \tag{5.14a}$$

$$\hat{B}_{\ell+1} = \text{diag} \left(\frac{\left\| \hat{\boldsymbol{g}}_1^{\ell+1} \right\|^2}{M}, \cdots, \frac{\left\| \hat{\boldsymbol{g}}_K^{\ell+1} \right\|^2}{M} \right), \tag{5.14b}$$

where $\hat{\boldsymbol{g}}_k^\ell$ is the k-th column of $\hat{G}_\ell$. Note that the resulting algorithm in fact is a generalized EM algorithm [55], in which the likelihood function is increased in the M-step and not necessarily maximized.

In general, the large scale fading coefficients change slower than the small scale fading coefficients and are easier to estimate over a longer period of time and can be fixed beforehand. In that case, B is fixed in (5.14a) from the beginning and (5.14b) is removed from the M-step.

Heuristic Semi-blind Algorithm

We now modify the EM algorithm to improve the estimation performance when data symbols $\mathbf{s}[n],\ n = L, \cdots, N-1$, are drawn from a discrete constellation. A heuristic approach is to assign the conditional mean of data symbols $\mathbb{E}(\mathbf{s}[n]|Y, \hat{G}_\ell),\ n = L, \cdots, N-1$, to the closest constellation point, which results in the following E-step:

$$\begin{aligned} \boldsymbol{\mu}_n^\ell &= F\left(\left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 I_K\right)^{-1} \hat{G}_\ell^H \mathbf{y}[n]\right), \\ \Sigma^\ell &= \sigma_v^2 \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 I_K\right)^{-1}, \end{aligned} \tag{5.15a}$$

where $F(.)$ is the element-wise constellation demmaping function. Note that the M-step remains the same as (5.12b):

$$\hat{G}_{\ell+1} = \left(Y_p S_p^H + \sum_{n=L}^{N-1} \boldsymbol{y}[n](\boldsymbol{\mu}_n^\ell)^H\right)\left(S_p S_p^H + \sum_{n=L}^{N-1}\left(\boldsymbol{\mu}_n^\ell(\boldsymbol{\mu}_n^\ell)^H + \Sigma^\ell\right)\right)^{-1}. \tag{5.15b}$$

EM Algorithm with GMM Prior

Numerical results in Section 5.5 suggest that the modification of the EM algorithm in (5.15) improves the estimation performance for discrete constellations. To provide analytical support for this heuristic approach, we derive a mathematically rigorous algorithm in the following by assuming a GMM distribution for data symbols which has a similar flavor. This algorithm is also based on the EM algorithm and hence its convergence to a local maximum is assured. Suppose data symbols have GMM distribution, i.e.,

$$\mathbf{s}[n] \sim \mathcal{CN}\left(\mathbf{c}_n, \sigma_s^2 I_K\right), \quad n = L, \cdots, N-1, \tag{5.16}$$

where $\mathbf{c}_n \in \mathbb{C}^{K\times 1}$ is the the transmitted constellation vector at time n that will be treated as the unknown parameter in the EM algorithm. The hyperparameter σ_s^2 in (5.16) is the variance of each data symbol around the corresponding constellation point. As σ_s^2 becomes smaller,

the GMM distribution in (5.16) becomes closer to the actual discrete distribution of the data symbols.

Let $\Theta = [G, \mathbf{c}_L, \cdots, \mathbf{c}_{N-1}]$ denote the unknown variables in the EM algorithm. The expected value of the log likelihood of the received signal, i.e., $\mathbb{E}_{p(S_d|Y,\hat{\Theta}_\ell)}\left(\log p(Y, S_d|\Theta)\right)$, is given by

$$\begin{aligned}\mathcal{L} = \text{const } &- \sum_{n=0}^{L-1} \frac{1}{\sigma_v^2} \left\|\boldsymbol{y}[n] - G\boldsymbol{s}[n]\right\|^2 - \sum_{n=L}^{N-1} \frac{1}{\sigma_v^2} \mathbb{E}_{p(S_d|Y,\hat{\Theta}_\ell)} \left(\left\|\boldsymbol{y}[n] - G\boldsymbol{s}[n]\right\|^2\right) \\ &- \sum_{n=L}^{N-1} \frac{1}{\sigma_s^2} \mathbb{E}_{p(S_d|Y,\hat{\Theta}_\ell)} \left(\left\|\boldsymbol{s}[n] - \boldsymbol{c}_n\right\|^2\right). \end{aligned} \tag{5.17}$$

The E-step of the EM algorithm with GMM is are given by

$$\begin{aligned}\boldsymbol{\mu}_n^\ell &= \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 \left(\hat{\mathbf{c}}_n^\ell (\hat{\mathbf{c}}_n^\ell)^H + \sigma_s^2 I_K\right)^{-1}\right)^{-1} \hat{G}_\ell^H \mathbf{y}[n], \\ \Sigma_n^\ell &= \sigma_v^2 \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 \left(\hat{\mathbf{c}}_n^\ell (\hat{\mathbf{c}}_n^\ell)^H + \sigma_s^2 I_K\right)^{-1}\right)^{-1}. \end{aligned} \tag{5.18a}$$

Maximizing the log likelihood function yields in the following M-step:

$$\begin{aligned}\hat{G}_{\ell+1} &= \left(Y_p S_p^H + \sum_{n=L}^{N-1} \mathbf{y}[n](\boldsymbol{\mu}_n^\ell)^H\right) \left(S_p S_p^H + \sum_{n=L}^{N-1} \left(\boldsymbol{\mu}_n^\ell (\boldsymbol{\mu}_n^\ell)^H + \Sigma_n^\ell\right)\right)^{-1}, \\ \hat{\mathbf{c}}_n^{\ell+1} &= F\left(\boldsymbol{\mu}_n^\ell\right), \end{aligned} \tag{5.18b}$$

where $F(.)$ is the element-wise constellation demmaping function.The derivation of the algorithm is presented in Appendix 5.8.D.

Note that all semi-blind algorithms introduce a time delay in estimating channel and uplink data symbols as we need to collect all data symbols before carrying out the estimation, which needs to be accounted for in implementing practical systems.

Table 5.1: Computational complexities of channel estimation using ML training-based estimation and full data. The mathematical operations in each step, e.g., matrix inversion, matrix multiplication, etc., determine the complexity of the corresponding step.

$\hat{G}_{\text{ML}}^{\text{tr}} = Y_p \, S_p^H \, \big(\underbrace{S_p S_p^H}_{\mathcal{O}(K^2L)} \big)^{-1}$ $\underbrace{\qquad\qquad}_{\mathcal{O}(K^3)}$ $\underbrace{\qquad\qquad\qquad}_{\mathcal{O}(K^2L)}$ $\underbrace{\qquad\qquad\qquad\qquad}_{\mathcal{O}(MLK)}$	$\hat{G}_{\text{ML}}^{\text{full}} = Y \, S^H \, \big(\underbrace{S S^H}_{\mathcal{O}(K^2N)} \big)^{-1}$ $\underbrace{\qquad\qquad}_{\mathcal{O}(K^3)}$ $\underbrace{\qquad\qquad\qquad}_{\mathcal{O}(K^2N)}$ $\underbrace{\qquad\qquad\qquad\qquad}_{\mathcal{O}(MNK)}$
Dominant term ($M \geq K, N \geq L \geq K$): $\mathcal{O}(MLK)$	Dominant term ($M \geq K, N \geq L \geq K$): $\mathcal{O}(MNK)$

5.3.2 Computational Complexity

We now compare the computational complexity of semi-blind estimation with ML estimators in Section 5.2.3 under the assumption that

$$M \geq K, \; N \geq L \geq K. \tag{5.19}$$

Calculation of ML training-based channel estimation in (5.4) consists of matrix multiplications with dominant factor of order $\mathcal{O}(MLK)$ and a matrix inversion with complexity $\mathcal{O}(K^3)$. Therefore, the dominant factor in calculation of (5.4) is of order $\mathcal{O}(MLK)$. Similarly, we can show that the computational complexity of channel estimation with full data in (5.6) is of order $\mathcal{O}(MNK)$. The calculation steps of the ML estimators based on pilots only and full data are presented in Table 5.1.

Now we evaluate the computational complexity of semi-blind channel estimation in (5.12). There are multiple ways to compute terms $\sum_{n=L}^{N-1} \boldsymbol{y}[n](\boldsymbol{\mu}_n^\ell)^H$ and $\sum_{n=L}^{N-1} \boldsymbol{\mu}_n^\ell(\boldsymbol{\mu}_n^\ell)^H$ in (5.12b) and depending on the order that multiplications are carried out, some of the already computed expressions can be later reused. Each of these cases will lead to a different overall computational complexity. Let Q denote the number of iterations required for EM algorithm

with Gaussian prior in (5.12) to converge. After investigating all possible orders to carry out the computational steps in the EM algorithm, we arrive at the following complexity order under the assumptions in (5.19).

$$\min\left\{\mathcal{O}\left((N-L+K)MKQ\right), \mathcal{O}\left((N-L+KQ)M^2\right)\right\}. \tag{5.20}$$

The first term in (5.2) corresponds to the computational complexities of calculating Σ^ℓ and multiplying $\boldsymbol{y}[n]$ by $(\boldsymbol{\mu}_n^\ell)^H$ for Q times. The second term in (5.2) corresponds to the case when we separately calculate $\hat{G}_\ell(\hat{G}_\ell^H\hat{G}_\ell + \sigma_v^2 I_K)^{-1}$ and $\sum_{n=L}^{N-1}\boldsymbol{y}[n]\boldsymbol{y}[n]^H$ and then multiply them together for Q times. The computational steps of these cases are presented in Table 5.2.

Suppose $L = K$ and Q is small constant as observed in the numerical results. The computational complexity of the EM algorithm with Gaussian prior is then in the order of $\mathcal{O}(MNKQ)$, which is Q times the complexity of channel estimation with full data. On the contrary, the assumption of $Q \gg \frac{M}{K}$ makes the computational complexity given in (5.2) to be in the order of $\mathcal{O}((N-L+KQ)M^2)$.

The complexity of EM algorithm with channel priors in (5.14) and the complexity of the heuristic semi-blind algorithm in (5.15) are of the same order as the EM algorithm with Gaussian prior. Thus, compared with the conventional training-based estimation in (5.4), semi-blind estimation provides better channel estimates (see Section 5.5) but at the expense of higher computational complexity.

The computational complexity of the EM algorithm with GMM prior in (5.18) also depends on the order that multiplications in terms $\sum_{n=L}^{N-1}\boldsymbol{y}[n](\boldsymbol{\mu}_n^\ell)^H$ and $\sum_{n=L}^{N-1}\boldsymbol{\mu}_n^\ell(\boldsymbol{\mu}_n^\ell)^H$ are carried out. Each scenario, will lead to a different complexity order. After considering all possible scenarios, we present the case that has the smallest complexity in Table 5.3. Thus, the computational complexity of the EM algorithm with GMM prior under the assumptions in

Table 5.2: Computational complexity of the EM algorithm with Gaussian prior using different multiplication orders. The mathematical operations in each step, e.g., matrix inversion, matrix multiplication, etc., determine the complexity of the corresponding step.

Scenario 1	$\sum_{n=L}^{N-1} \boldsymbol{y}[n] \left(\boldsymbol{\mu}_n^\ell\right)^H = \underbrace{\sum_{n=L}^{N-1} \boldsymbol{y}[n]\, \boldsymbol{y}[n]^H}_{\mathcal{O}((N-L)MKQ)} \underbrace{\hat{G}_\ell}_{\mathcal{O}((N-L)MKQ)} \underbrace{\Big(}_{\mathcal{O}(MK^2Q)} \underbrace{\hat{G}_\ell^H \hat{G}_\ell}_{\mathcal{O}(MK^2Q)} + \sigma_v^2 I_K \Big)^{-1}}_{\mathcal{O}(K^3Q)}$ $\sum_{n=L}^{N-1} \boldsymbol{\mu}_n^\ell \left(\boldsymbol{\mu}_n^\ell\right)^H = \underbrace{\underbrace{\left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 I_K\right)^{-1} \hat{G}_\ell^H}_{\text{already computed}} \underbrace{\sum_{n=L}^{N-1} \boldsymbol{y}[n] \boldsymbol{y}[n]^H \hat{G}_\ell \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 I_K\right)^{-1}}_{\text{already computed}}}_{\mathcal{O}(MK^2Q)}$ Dominant term ($M \geq K,\ N \geq L \geq K$): $\mathcal{O}((N-L)MKQ + MK^2Q)$
Scenario 2	$\sum_{n=L}^{N-1} \boldsymbol{y}[n] \left(\boldsymbol{\mu}_n^\ell\right)^H = \underbrace{\underbrace{\sum_{n=L}^{N-1} \boldsymbol{y}[n] \boldsymbol{y}[n]^H}_{\mathcal{O}((N-L)M^2)} \hat{G}_\ell \Big(}_{\mathcal{O}(M^2KQ)} \underbrace{\hat{G}_\ell^H \hat{G}_\ell}_{\mathcal{O}(MK^2Q)} + \sigma_v^2 I_K \Big)^{-1}}_{\mathcal{O}(K^3Q)}$ $\sum_{n=L}^{N-1} \boldsymbol{\mu}_n^\ell \left(\boldsymbol{\mu}_n^\ell\right)^H = \underbrace{\underbrace{\left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 I_K\right)^{-1} \hat{G}_\ell^H}_{\text{already computed}} \underbrace{\sum_{n=L}^{N-1} \boldsymbol{y}[n] \boldsymbol{y}[n]^H \hat{G}_\ell \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 I_K\right)^{-1}}_{\text{already computed}}}_{\mathcal{O}(MK^2Q)}$ Dominant term ($M \geq K,\ N \geq L \geq K$): $\mathcal{O}((N-L)M^2 + M^2KQ)$
	Overall complexity: $\min\left\{\mathcal{O}((N-L+K)MKQ), \mathcal{O}\left((N-L+KQ)M^2\right)\right\}$

Table 5.3: Computational complexity of the EM algorithm with GMM prior. The mathematical operations in each step, e.g., matrix inversion, matrix multiplication, etc., determine the complexity of the corresponding step.

$$\sum_{n=L}^{N-1} \boldsymbol{y}[n]\left(\boldsymbol{\mu}_n^{\ell}\right)^H = \sum_{n=L}^{N-1} \boldsymbol{y}[n] \underbrace{\underbrace{\underbrace{\underbrace{\boldsymbol{y}[n]^H \hat{G}_\ell}_{\mathcal{O}((N-L)MKQ)} \underbrace{\Big(\underbrace{\hat{G}_\ell^H \hat{G}_\ell}_{\mathcal{O}(MK^2Q)} + \underbrace{\sigma_v^2 \big(\hat{\mathbf{c}}_n^{\ell}(\hat{\mathbf{c}}_n^{\ell})^H + \sigma_s^2 I_K\big)^{-1}}_{\mathcal{O}((N-L)K^3Q)} \Big)^{-1}}_{\mathcal{O}((N-L)K^3Q)}}_{\mathcal{O}((N-L)K^2Q)}}_{\mathcal{O}((N-L)MKQ)}}$$

$$\sum_{n=L}^{N-1} \boldsymbol{\mu}_n^{\ell}\left(\boldsymbol{\mu}_n^{\ell}\right)^H = \underbrace{\sum_{n=L}^{N-1} \underbrace{\big(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 \left(\hat{\mathbf{c}}_n^{\ell}(\hat{\mathbf{c}}_n^{\ell})^H + \sigma_s^2 I_K\right)^{-1}\big)^{-1} \hat{G}_\ell^H \boldsymbol{y}[n]}_{\text{already computed}} \times \underbrace{\boldsymbol{y}[n]^H \hat{G}_\ell \big(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 \left(\hat{\mathbf{c}}_n^{\ell}(\hat{\mathbf{c}}_n^{\ell})^H + \sigma_s^2 I_K\right)^{-1}\big)^{-1}}_{\text{already computed}}}_{\mathcal{O}((N-L)K^2Q)}$$

Dominant term ($M \geq K,\ N \geq L \geq K$): $\mathcal{O}((N-L)MKQ + (N-L)K^3Q + MK^2Q)$

(5.19) is given by

$$\mathcal{O}((N-L)MKQ + (N-L)K^3Q + MK^2Q). \tag{5.21}$$

5.3.3 Downlink Beamforming

In this section, we consider the effect of channel estimation in downlink beamforming. In TDD systems, due to channel reciprocity, the uplink channel estimates are used to form the downlink precoders. The better we estimate the channel coefficients in uplink, the more accurate the downlink precoders become. Thus, in TDD systems we benefit from semi-blind channel estimation in both uplink and downlink transmissions. Note that due to the mismatch in hardware chains of transmitter and receiver, the system needs to be calibrated before chan-

nel reciprocity can be exploited [17], [77]. For simplicity, however, we assume uplink and downlink channels are the same. With conjugate beamforming, the BS transmits the signal $\boldsymbol{x} = \sum_{i=1}^{K} \sqrt{p_i} \frac{\hat{\boldsymbol{g}}_i^*}{\|\hat{\boldsymbol{g}}_i\|} s_i$, where p_i and s_i are the transmit power and the data signal intended to user i respectively, and $\hat{\boldsymbol{g}}_i$ is the ith column of channel estimate $\hat{G}$. The time index is omitted in this section to simplify the notations. The k-th user receives

$$y_k = \boldsymbol{g}_k^T \boldsymbol{x} + v_k = \underbrace{\sqrt{p_k} \frac{\boldsymbol{g}_k^T \hat{\boldsymbol{g}}_k^*}{\|\hat{\boldsymbol{g}}_k\|} s_k}_{J_0\text{: desired signal}} + \underbrace{\sum_{\substack{i=1 \\ i \neq k}}^{K} \sqrt{p_i} \frac{\boldsymbol{g}_k^T \hat{\boldsymbol{g}}_i^*}{\|\hat{\boldsymbol{g}}_i\|} s_i}_{J_1\text{: interference}} + v_k, \tag{5.22}$$

where $v_k \sim \mathcal{CN}\left(0, \sigma_v^2\right)$ is additive noise. Note that the gain of desired signal, i.e., $\frac{\sqrt{p_k} \boldsymbol{g}_k^T \hat{\boldsymbol{g}}_k^*}{\|\hat{\boldsymbol{g}}_k\|}$, can be estimated at user device by sending a few downlink pilots. Since data symbols of different users and additive noise are mutually independent, terms J_0, J_1, and v_k in (5.22) are mutually uncorrelated. Suppose $\hat{g}_{mi}$ and g_{mk} are uncorrelated for any $i \neq k$. According to [46], the worst case noise for mutual information is Gaussian additive noise with the variance equal to the variance of $J_1 + v_k$. Hence, the achievable rate is lower bounded by $R_k = \log_2(1 + \text{SINR}_k)$, where

$$\text{SINR}_k = \frac{p_k \frac{\left|\boldsymbol{g}_k^T \hat{\boldsymbol{g}}_k^*\right|^2}{\|\hat{\boldsymbol{g}}_k\|^2}}{\sum\limits_{i \neq k}^{K} p_i \sum\limits_{m=1}^{M} \mathbb{E}\left(|g_{mk}|^2\right) \mathbb{E}\left(\frac{|\hat{g}_{mi}|^2}{\|\hat{\boldsymbol{g}}_i\|^2}\right) + \sigma_v^2}. \tag{5.23}$$

Let us assume channel estimates $\hat{g}_{mi},\ m = 1, \cdots, M$ are i.i.d random variables. The following lemma is used to further simplify the formulation of SINR in (5.23).

Lemma 5.2. *Suppose $\hat{g}_{mi},\ m = 1, \cdots, M$ are i.i.d random variables and let $y_m = \frac{|\hat{g}_{mi}|^2}{\|\hat{\boldsymbol{g}}_i\|^2},\ m = 1, \cdots, M$. Then $y_1, \cdots, y_M$ are identically distributed with mean $\mathbb{E}\left(y_m\right) = \frac{1}{M},\ m = 1, \cdots, M$.*

Proof. See Appendix 5.8.E. □

Based on Lemma 5.2, the achievable rate in high signal-to-noise ratio (SNR) regime

can be written as

$$R_k \approx \log_2\left(\text{SINR}_k\right) = \log_2\left(\gamma\right) + \log_2\left(\frac{p_k \left\|\boldsymbol{g}_k\right\|^2}{\sum_{i\neq k}^{K} \frac{p_i}{M}\mathbb{E}\left(\left\|\boldsymbol{g}_k\right\|^2\right) + \sigma_v^2}\right), \tag{5.24}$$

where

$$\gamma = \frac{\left|\boldsymbol{g}_k^T \hat{\boldsymbol{g}}_k^*\right|^2}{\|\boldsymbol{g}_k\|^2 \|\hat{\boldsymbol{g}}_k\|^2}. \tag{5.25}$$

The second term in (5.24) is the achievable rate with perfect CSI in high SNR regime. Note that $\log_2(\gamma)$ in (5.24) represents the loss in capacity due to channel estimation error. As the channel estimation accuracy increases, parameter γ becomes closer to 1 minimizing the rate loss. Thus, in numerical results, parameter $\log_2(\gamma)$ is used as our performance measure to compare the estimation accuracy of semi-blind and ML estimators for downlink transmission.

5.4 Cramer-Rao Bound

In the following subsections we derive the CRB, covariance lower bound, for semi-blind channel estimation. There are two common assumptions made to obtain the CRB [74], [75], [76]: 1. deterministic model in which the data signal S_d is modeled as an unknown deterministic quantity. 2. Stochastic model where the data signal S_d is modeled as a random sequence. For each model, using results of large random matrix theory we further derive the asymptotic behavior of CRB when M, the number of antennas, grows infinitely large providing insight into the behavior of massive MIMO systems.

5.4.1 Deterministic CRB

Under the deterministic assumption for the data signal S_d, we have

$$\boldsymbol{y}[n] \sim \mathcal{CN}\left(G\boldsymbol{s}[n], \sigma_v^2 I_M\right), \quad n = 0, \cdots, N-1. \tag{5.26}$$

A general deterministic CRB of covariance matrix has been derived in [75]. Define $\bar{x} \triangleq \text{Re}\{x\}$, $\tilde{x} \triangleq \text{Im}\{x\}$, $C = \left(\frac{2}{\sigma_v^2} G^H G\right)^{-1}$ and

$$W_k[n] = \frac{2}{\sigma_v^2} G^H s_k[n], \quad \lambda_{ki} = \frac{2}{\sigma_v^2} \sum_{n=0}^{N-1} s_k[n]^* s_i[n]. \tag{5.27}$$

The deterministic CRB of the covariance matrix of any unbiased semi-blind estimator of G is given by

$$\text{CRB} = \left(\Lambda - \sum_{n=L}^{N-1} \Omega_n^T \begin{bmatrix} \bar{C} & -\tilde{C} \\ \tilde{C} & \bar{C} \end{bmatrix} \Omega_n \right)^{-1}, \tag{5.28}$$

where

$$\Lambda = \begin{bmatrix} \begin{bmatrix} \bar{\lambda}_{11} I_M & -\tilde{\lambda}_{11} I_M \\ \tilde{\lambda}_{11} I_M & \bar{\lambda}_{11} I_M \end{bmatrix} & \cdots & \begin{bmatrix} \bar{\lambda}_{1K} I_M & -\tilde{\lambda}_{1K} I_M \\ \tilde{\lambda}_{1K} I_M & \bar{\lambda}_{1K} I_M \end{bmatrix} \\ \vdots & \ddots & \vdots \\ \begin{bmatrix} \bar{\lambda}_{K1} I_M & -\tilde{\lambda}_{K1} I_M \\ \tilde{\lambda}_{K1} I_M & \bar{\lambda}_{K1} I_M \end{bmatrix} & \cdots & \begin{bmatrix} \bar{\lambda}_{KK} I_M & -\tilde{\lambda}_{KK} I_M \\ \tilde{\lambda}_{KK} I_M & \bar{\lambda}_{KK} I_M \end{bmatrix} \end{bmatrix}, \tag{5.29a}$$

and for $n = L, \cdots, N-1$,

$$\Omega_n = \begin{bmatrix} \begin{bmatrix} \bar{W}_1[n] & -\tilde{W}_1[n] \\ \tilde{W}_1[n] & \bar{W}_1[n] \end{bmatrix} & \cdots & \begin{bmatrix} \bar{W}_K[n] & -\tilde{W}_K[n] \\ \tilde{W}_K[n] & \bar{W}_K[n] \end{bmatrix} \end{bmatrix}. \tag{5.29b}$$

The calculations are given in Appendix 5.8.F.

Theorem 5.1. *The limit of the deterministic CRB as $M \to \infty$, i.e., the number of antennas at the BS increases, is given by*

$$\text{CRB} \xrightarrow[M \to \infty]{a.s.} \Lambda^{-1}. \tag{5.30}$$

Proof. See Appendix 5.8.G. □

Remark 5.2. *Based on the derivations in Appendix 5.8.F, matrix Λ^{-1} in (5.30) corresponds to the CRB of channel matrix G when all data symbols S_d are known. Thus, as the number of antennas at the BS increases, the CRB of the semi-blind estimation converges to the CRB of a genie aided system with full data in which all N symbols (S) are known at the BS.*

5.4.2 Stochastic CRB

With Gaussian assumption for the data symbols S_d, we have

$$\begin{aligned}\boldsymbol{y}[n] &\sim \mathcal{CN}\left(G\boldsymbol{s}[n], \sigma_v^2 I_M\right), \quad n = 0, \cdots, L-1, \\ \boldsymbol{y}[n] &\sim \mathcal{CN}\left(0, R\right), \quad n = L, \cdots, N-1,\end{aligned} \tag{5.31}$$

where $R = GG^H + \sigma_v^2 I_M$. For deriving the result, we use the fact that the CRB on the covariance matrix of observed data vector $\boldsymbol{z} \sim \mathcal{CN}\left(\boldsymbol{\mu}(\boldsymbol{\alpha}), \Sigma(\boldsymbol{\alpha})\right)$ for any unbiased estimate of arbitrary $\boldsymbol{\alpha}$ is given by [80]

$$\left[\text{CRB}^{-1}\right]_{ij} = \text{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \alpha_i}\Sigma^{-1}\frac{\partial \Sigma}{\partial \alpha_j}\right) + 2\text{Re}\left\{\frac{\partial \boldsymbol{\mu}^H}{\partial \alpha_i}\Sigma^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \alpha_j}\right\}. \tag{5.32}$$

Let $\boldsymbol{\alpha} = \left[\bar{\boldsymbol{g}}_1^T, \tilde{\boldsymbol{g}}_1^T, \cdots, \bar{\boldsymbol{g}}_K^T, \tilde{\boldsymbol{g}}_K^T\right]^T$. The stochastic CRB of any unbiased estimator of G with the stochastic model defined in (5.31) and orthogonal pilot sequences, i.e., $S_p S_p^H = LI_K$, is given by

$$\left[\text{CRB}^{-1}\right]_{ij} = \frac{2L}{\sigma_v^2}\delta\left(i-j\right) + (N-L)\text{tr}\left(R^{-1}\frac{\partial R}{\partial \alpha_i}R^{-1}\frac{\partial R}{\partial \alpha_j}\right). \tag{5.33}$$

The calculations are given in Appendix 5.8.H.

Theorem 5.2. *The limit of the stochastic CRB with orthogonal pilot sequences ($S_p S_p^H = LI_K$) as $M \to \infty$, i.e., the number of antennas at the BS increases, is given by*

$$\text{CRB} \xrightarrow[M\to\infty]{a.s.} \frac{\sigma_v^2}{2N} I_{2MK}. \tag{5.34}$$

Proof. See Appendix 5.8.I. □

Remark 5.3. *Based on Theorem 5.2, the CRB of semi-blind channel estimation with unlimited number of BS antennas is equivalent to the CRB of a system where the whole transmission block of length N acts as orthogonal pilot sequences, i.e., $SS^H = NI_K$.*

Remark 5.4. *In massive MIMO systems with an increasing number of users, pilot contamination, which originates from non-orthogonal pilot sequences or the reuse of pilot sequences for neighboring cells, is a severe limiting factor in achievable data rates [15], [20]. Increasing length of the pilot sequences reduces the pilot contamination effect. The asymptotic results of CRB with both deterministic and stochastic assumptions indicate that with sufficiently large number of antennas at BS, the effective pilot length of the system increases. In other words, using information carried in unknown data is analogous to increasing the pilot length when only pilot sequences are used for channel estimation (see Figures 5.4, 5.7, and 5.8) but without the loss in throughput. This property makes semi-blind channel estimation an attractive approach to alleviate the pilot contamination bottleneck in massive MIMO systems which is our future topic of study.*

5.5 Numerical Results

In the numerical experiments, we consider a single cell with radius 500m and a BS located at the center of the cell and uniformly distributed users. We use a three-slope path loss model [52] for large scale fading coefficients as follows

$$\beta_k = \begin{cases} c_0 & d_k \leq d_0 \\ \dfrac{c_1}{d_k^2} & d_0 < d_k \leq d_1 \\ \dfrac{c_2 z_k}{d_k^{3.5}} & d_k > d_1 \end{cases}, \tag{5.35}$$

where d_k is the distance in kilometers between user k and the BS, and z_k is the log-normal shadow fading, i.e., $10 \log_{10} z_k \sim \mathcal{N}(0, \sigma_{\text{shad}}^2)$ with $\sigma_{\text{shad}} = 8$ dB. For $d_k > d_1$ we use COST-

231 Hata propagation model

$$10\log_{10} c_2 = -46.3 - 33.9\log_{10} f + 13.82\log_{10} h_B + (1.1\log_{10} f - 0.7)h_R - (1.56\log_{10} f - 0.8), \quad (5.36)$$

where $f = 1900$ MHz is the carrier frequency, $h_B = 15$m is the BS antenna height, and $h_R = 1.65$m is the user antenna height. Path loss parameters c_1 and c_0 in (5.35) are given by

$$10\log_{10} c_1 = 10\log_{10} c_2 - 15\log_{10}(d_1), \quad 10\log_{10} c_0 = 10\log_{10} c_1 - 20\log_{10}(d_0). \quad (5.37)$$

We choose $d_0 = 0.01$km and $d_1 = 0.05$km. Signal-to-noise ratio in the experiments is considered to be SNR $= \frac{\mathbb{E}(\beta_k)}{\sigma_v^2}$. Pilot sequences ($S_p$) and data symbols ($S_d$) are drawn from a QPSK constellation and pilot sequences are chosen to be orthogonal. We initialize all semi-blind algorithms using the ML training-based estimate in (5.4). In the EM algorithm with GMM prior, we set $\sigma_s = 0.001$.

Experiment 1: In this experiment, it is shown that one can obtain more accurate channel estimates by using semi-blind estimation. We compare MSE of the ML estimates given in Subsection 5.2.3 and the semi-blind algorithms given in (5.12), (5.14), (5.15), and (5.18). In this experiment, $M = 8$, $K = 4$, $L = 16$, and $N = 512$. Figure 5.1 shows the scaled MSE, i.e., $\mathbb{E}(\|G - \hat{G}\|_F^2)/\mathbb{E}(\beta_k)$, of the channel estimates versus SNR for the two ML estimators and the semi-blind algorithms. Scaled MSE of the EM algorithm defined in (5.12) with Gaussian data symbols is also plotted in the Figure 5.1. Even though the EM algorithm with Gaussian prior defined in (5.12) is obtained for the Gaussian data, one can observe that MSE of the QPSK data is virtually indistinguishable from that of the experiment with Gaussian data symbols. We provide the following observation to support the good performance observed. In the Bayesian framework, the prior information is of limited importance and the posterior density of the symbols given the received data, as required in the EM algorithm, is controlled

by the received data [55]. In this case, the posterior is a Gaussian density with a mean close to the actual symbol value. This can be understood by noting that (5.12a) is in fact the linear minimum mean squared error (LMMSE) estimate of symbols. In addition, if the prior is being approximated, a non-informative prior is often used and suggested. Based on the maximum-entropy principle, when only partial information is available on the prior distribution, the distribution with the largest entropy consistent with the partial information is a useful non-informative prior [81], [82]. Based on the zero mean and variance of the symbols, the Gaussian density has the maximum entropy among continuous real valued densities making it a reasonable choice. Moreover, it is not necessary for improved channel estimation to have exact symbol recovery. The quality of channel estimation improves as long as the estimates of the data symbols are in the neighborhood of actual data symbols and are properly weighted by the uncertainty in the estimate. Based on these observations, we believe the Gaussian density is a useful approximation as a prior for data symbols and the EM algorithm obtained with the Gaussian data assumption improves channel estimates even when the data symbols are drawn from a discrete constellation.

One can observe that all semi-blind algorithms provide better channel estimates compared with the ML training-based estimation.Note that in low SNRs, the EM algorithm with channel priors defined in (5.14) outperforms the EM algorithm with Gaussian prior given in (5.12). In low SNRs, the EM algorithm with Gaussian prior outperforms the heuristic semi-blind scheme and the EM algorithm with GMM prior. However, as SNR increases, the heuristic semi-blind estimation and the EM algorithm with GMM prior provide better channel estimates and become closer to the genie-aided ML estimator. To explain this behavior, we point out that the constellation demapping in the heuristic algorithm and the EM algorithm with GMM prior adds to the estimation error when the estimates of $\boldsymbol{\mu}_n$ are uncertain in low SNRs. This phenomenon works in favor of these two algorithms in the high SNR regime by

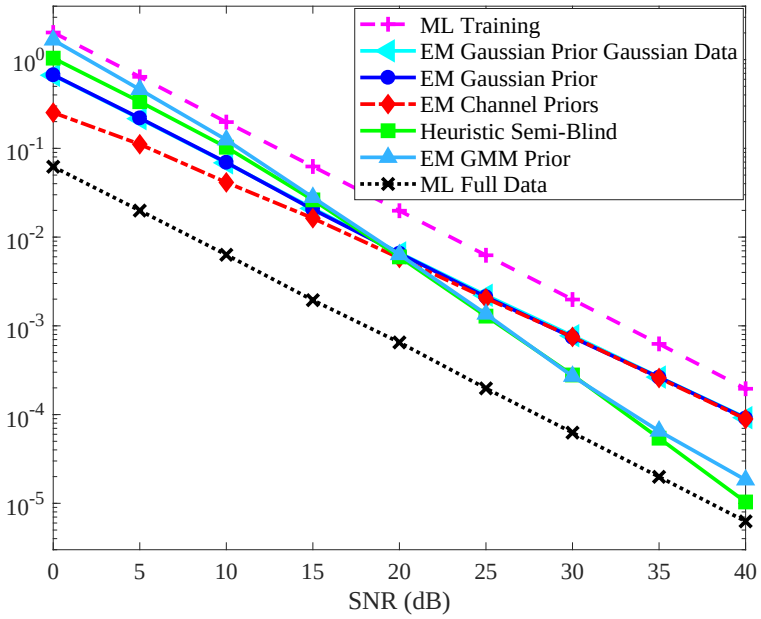

Figure 5.1: Scaled MSE versus SNR with $M = 8$, $K = 4$, $L = 16$, and $N = 512$.

mapping $\boldsymbol{\mu}_n$ to its true value.

In Figures 5.2 and 5.3, the scaled MSE of EM algorithm with Gaussian prior versus number of iterations is plotted for SNR $= 15$dB and SNR $= 30$dB respectively. It can be observed that the EM algorithm with Gaussian prior converges after a few iterations. Therefore, its complexity with small number of iterations (Q) is comparable to that of the ML estimator with full data (see Section 5.3.2).

In the next three experiments we consider a massive MIMO system in which the number of antennas at the BS is much larger than the number of users ($M \gg K$).

Experiment 2: In this experiment, we compare MSE and symbol error rate of the ML estimators described in Subsection 5.2.3 and the EM algorithm with Gaussian prior given in (5.12) for $M = 64$, $K = 8$, $L = 16$, and $N = 512$. We consider symbol error rate of a LMMSE receiver in uplink transmission. The LMMSE receiver is based on the channel estimates obtained by each algorithm. Scaled MSE of the EM algorithm with Gaussian prior given in

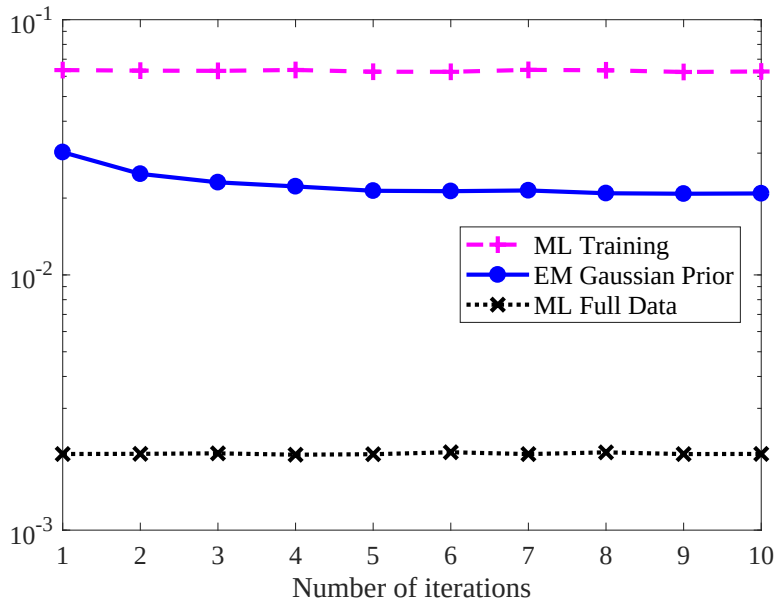

Figure 5.2: Scaled MSE versus number of iterations with $M = 8$, $K = 4$, $L = 16$, $N = 512$, and SNR $= 15$dB.

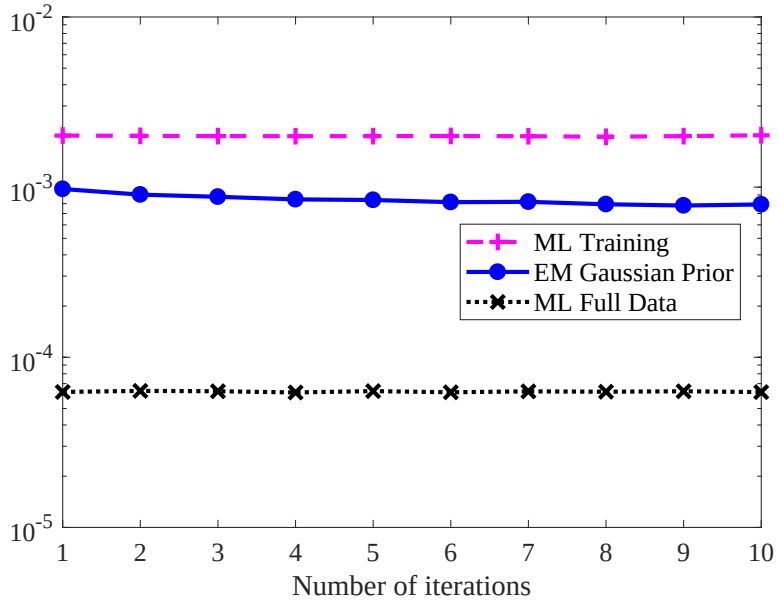

Figure 5.3: Scaled MSE versus number of iterations with $M = 8$, $K = 4$, $L = 16$, $N = 512$, and SNR $= 30$dB.

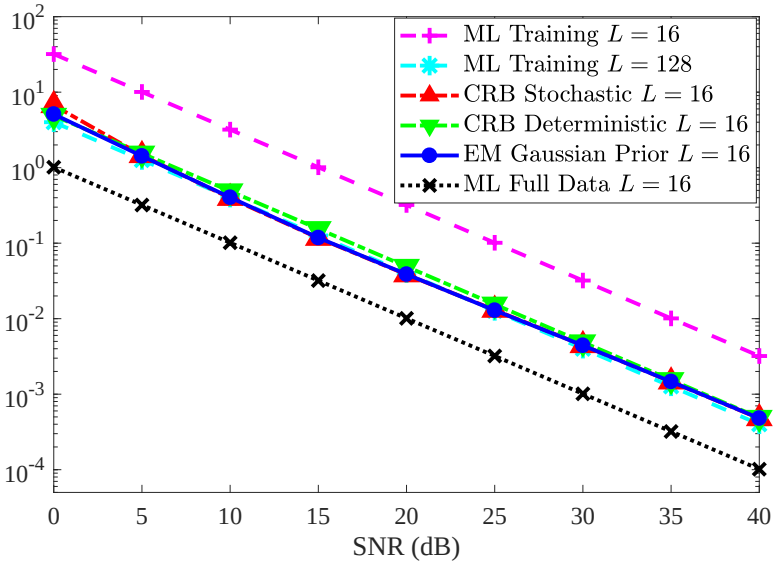

Figure 5.4: Scaled MSE versus SNR with $M = 64$, $K = 8$, and $N = 512$.

(5.12) and the ML estimators are plotted in Figure 5.4. One can see that the EM algorithm with Gaussian prior outperforms ML estimator based on pilot sequences significantly. Scaled MSE of the ML estimator with training sequence of length $L = 128$ is also plotted in this figure. It can be seen that with eight times smaller training sequence, MSE of the EM algorithm with Gaussian prior is very close to the MSE of the ML estimator with $L = 128$ training sequences, indicating a significant benefit from the semi-blind scheme. The deterministic and stochastic CRBs of the semi-blind channel estimation are also plotted in Figure 5.4. Note that the deterministic and stochastic CRBs of semi-blind estimation are defined for any unbiased estimator. In [83], it has been shown that the EM estimator for channel coefficients is biased. For this reason, in Figure 5.4 the EM estimator is not necessarily lower bounded by CRB.

Figure 5.5 shows the symbol error rate of the LMMSE receiver versus SNR. It can be seen that performance of the EM algorithm with Gaussian prior becomes closer to the ML with full data as SNR increases. In the EM algorithm with Gaussian prior, the hidden parameters are more likely to be estimated accurately as SNR increases. Therefore, symbol error rate of

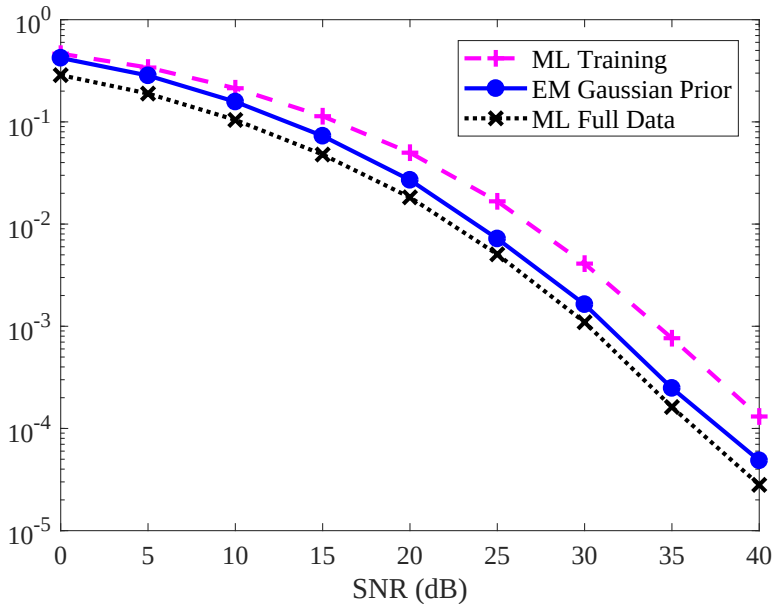

Figure 5.5: Symbol error rate versus SNR with $M = 64$, $K = 8$, $L = 16$, and $N = 512$.

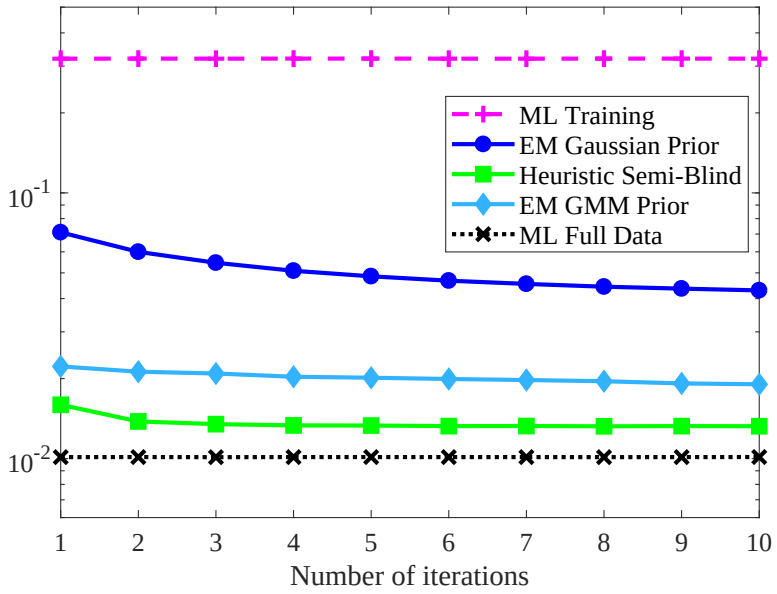

Figure 5.6: Scaled MSE versus number of iterations with $M = 64$, $K = 8$, $L = 16$, $N = 512$, and SNR $= 20$dB.

the EM algorithm with Gaussian prior is closer to that of the ML estimator with full data in high SNRs.

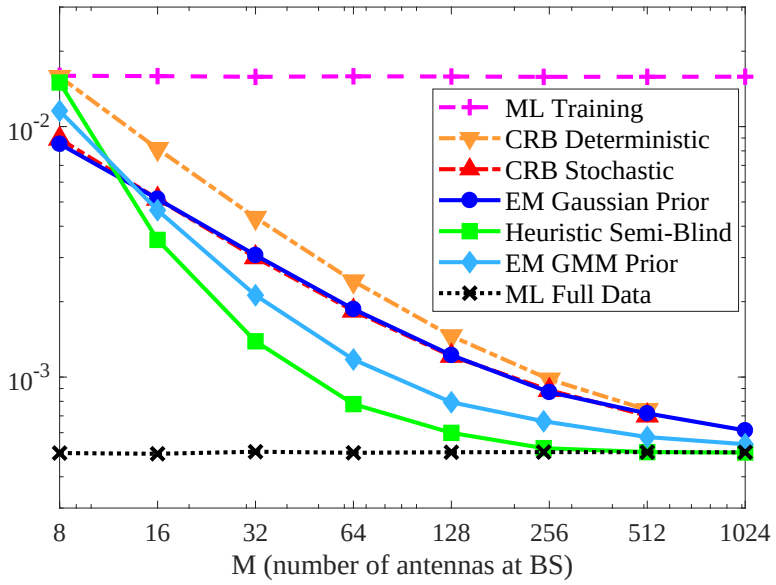

Figure 5.7: $\frac{1}{M}$MSE versus M with $K = 8$, $L = 16$, $N = 512$, and SNR$= 15$dB.

In Figure 5.6, we plot the scaled MSE of the semi-blind algorithms versus number of iterations for $M = 64$, $K = 8$, $L = 16$, $N = 512$, and SNR $= 20$dB. One can see that all semi-blind algorithms converge after a few iterations and that the heuristic semi-blind algorithm and the EM algorithm with GMM prior show faster convergence compared to the EM algorithm with Gaussian prior.

Experiment 3: In this experiment, we study effect of increasing the number of antennas at BS with $K = 8$, $L = 16$, and $N = 512$. Figure 5.7 shows the scaled MSE of the ML estimators described in Subsection 5.2.3 and the semi-blind algorithms in (5.12), (5.15), and (5.18) versus number of antennas M for SNR $= 15$dB. The CRBs given in Section 5.4 are also plotted in the figure. It can be observed that as the number of antennas increases, the performance of the semi-blind algorithms becomes closer to the ML estimate with full data which depicts the effectiveness of using the information carried in unknown data to estimate channel coefficients in massive MIMO systems. The figure also confirms results obtained in Theorems 5.1 and 5.2. Figure 5.8 shows the same results for SNR $= 30$dB.

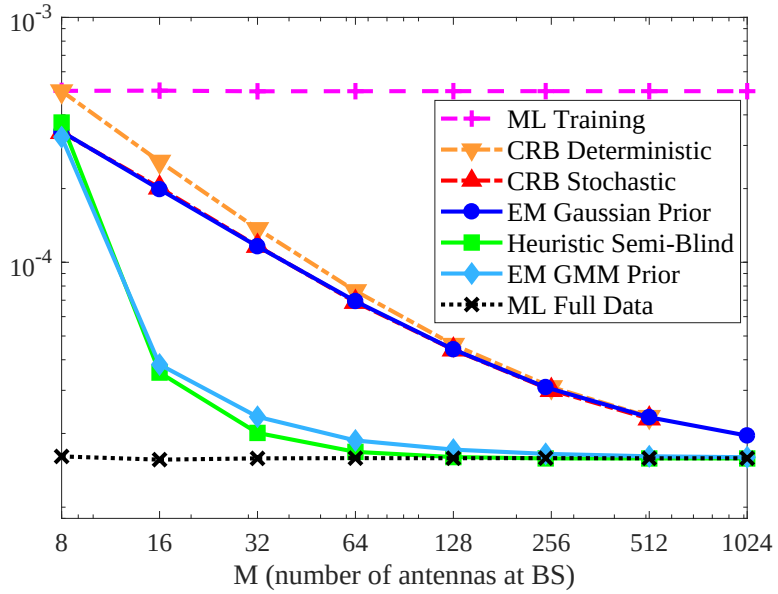

Figure 5.8: $\frac{1}{M}$MSE versus M with $K = 8$, $L = 16$, $N = 512$, and SNR$=$ 30dB.

Experiment 4: Finally, we consider the effect of semi-blind channel estimation in downlink beamforming. Figure 5.9 shows the cumulative distribution function (CDF) of $\log_2(\gamma)$ defined in (5.25) for all users with ML estimators and the EM algorithm with Gaussian prior for SNR $=$ 15dB. Parameter $\log_2(\gamma)$ represents the loss in capacity (in high SNR regime) due to channel estimation error. It can be seen that the CDF of $\log_2(\gamma)$ with semi-blind channel estimation defined in (5.12) is closer to that of the perfect CSI compared to the ML training-based estimation defined in (5.4). The horizontal line corresponds to the 10th percentile. The loss in capacity in terms of the 10th percentile (due to channel estimation error) for ML training-based estimation and the EM algorithm with Gaussian prior is 1.84 bits per channel use and 0.48 bits per channel use respectively. Thus, the semi-blind channel estimation provides 1.36 bits per channel use improvement in the 10th percentile of the achievable rate over the ML training-based estimation. Figure 5.10 presents the CDF of $\log_2(\gamma)$ for SNR $=$ 30dB. In this scenario, semi-blind estimation defined in (5.12) provides 0.1 bits per channel use improvement in the 10th percentile of the achievable rate over ML training-based estimation.

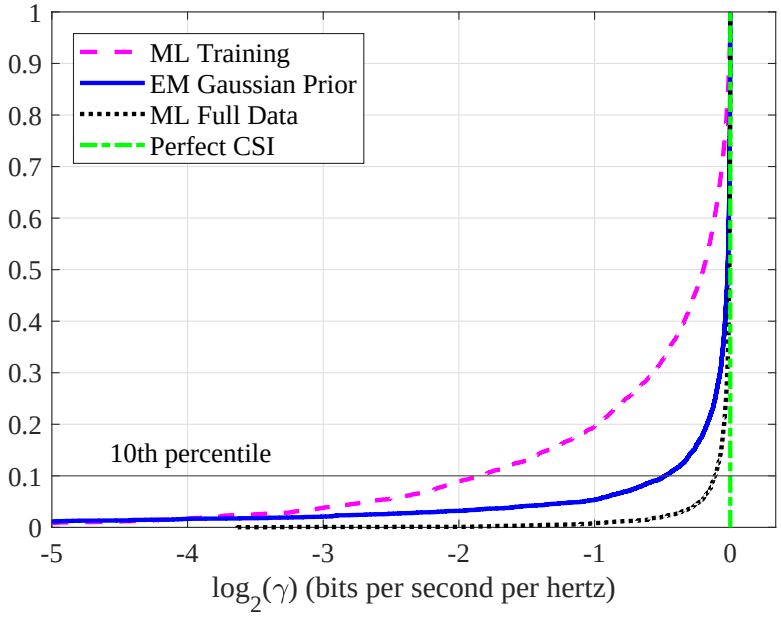

Figure 5.9: CDF of $\log_2(\gamma) = \log_2\left(\frac{|\boldsymbol{g}_k^T \hat{\boldsymbol{g}}_k^*|^2}{\|\boldsymbol{g}_k\|^2 \|\hat{\boldsymbol{g}}_k\|^2}\right)$ for different channel estimation schemes with $M = 64$, $K = 8$, $L = 16$, $N = 512$, and SNR$= 15$dB.

In high SNR regime, the quality of channel estimation with ML training-based estimator in (5.4) is close to the perfect CSI case. Hence, in this case, using semi-blind channel estimation slightly improves the performance in terms of the 10th percentile of the achievable rate over training-based estimator.

5.6 Conclusion

Motivated by TDD massive MIMO systems, we developed EM based algorithms for semi-blind channel estimation by employing Gaussian and GMM priors on data symbols and a suitable prior on channel coefficients. To understand the limits of estimation accuracy, we obtained CRBs for semi-blind channel estimation under two sets of commonly used assumptions. The behavior of CRBs for massive MIMO systems when the number of antennas at the BS grows without bound is investigated leading to conclusions that provide theoretical support

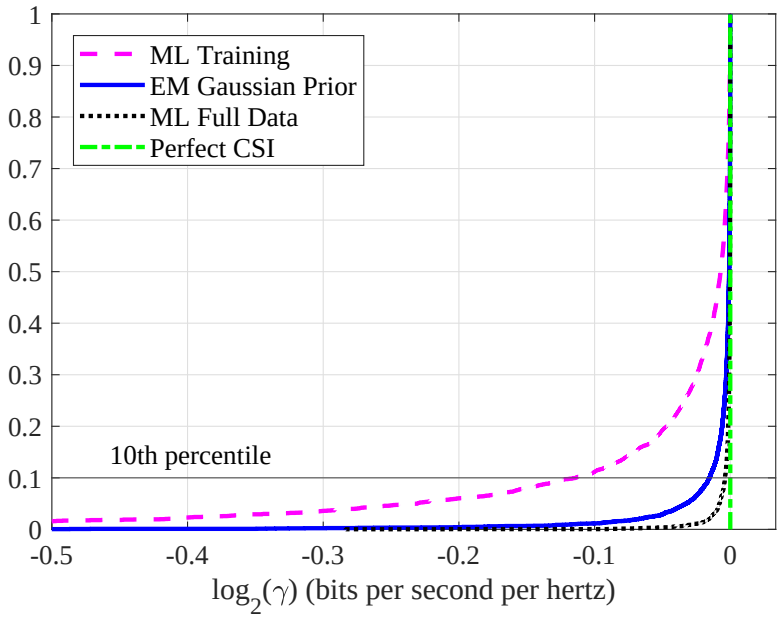

Figure 5.10: CDF of $\log_2(\gamma) = \log_2\left(\frac{|\boldsymbol{g}_k^T \hat{\boldsymbol{g}}_k^*|^2}{\|\boldsymbol{g}_k\|^2 \|\hat{\boldsymbol{g}}_k\|^2}\right)$ for different channel estimation schemes with $M = 64$, $K = 8$, $L = 16$, $N = 512$, and SNR$= 30$dB.

for the use of semi-blind algorithms for TDD based massive MIMO systems. We compared semi-blind estimation with known ML estimators. Numerical results indicate effectiveness of semi-blind channel estimation (compared to the estimation based on pilot sequences only) for both uplink and downlink transmissions in MIMO and specially in massive MIMO systems. In particular, the performance of the semi-blind channel estimation algorithms becomes closer to the genie aided ML estimator based on full data as the number of BS antennas increases. Numerical results show that the EM algorithm with Gaussian prior has superior performance compared with the EM algorithm with GMM prior in the low SNR regime. However, as the SNR or as the number BS antennas increases, the performance of the EM algorithm with GMM prior improves compared to the EM algorithm with Gaussian prior and becomes closer to the genie aided ML estimator.

5.7 Acknowledgment

This chapter, in part, is a reprint of the material as it appears in [4], E. Nayebi, and B. D. Rao, “Semi-blind Channel Estimation for Multiuser Massive MIMO Systems," *IEEE Transactions on Signal Processing*, vol. 66, no.2, pp. 540-553, Jan. 2018, and is also based on the material as it appears in [5], E. Nayebi, and B. D. Rao, “Semi-blind Channel Estimation in Massive MIMO Systems with Different Priors on Data Symbols” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2018*, to appear. The dissertation author is the primary investigator and author of these papers.

5.8 Appendices

5.8.A E-Step of the EM Algorithm with Discrete Constellation

Suppose data symbols are chosen from a discrete constellation such as QPSK. Denote by $p(x)$ the probability mass function of discrete random variable x and let $\mathcal{S}$ be a set of size $|\mathcal{S}| = 4^K$, which contains constellation points for all users. Then the conditional mean of data symbols, $\mathbb{E}(\boldsymbol{s}[n]|\hat{G}_\ell, Y)$, in the E-step of the EM algorithm based on the estimates at the ℓth iteration can be computed as follows

$$\begin{aligned}\mathbb{E}\left(\boldsymbol{s}[n]\middle|\hat{G}_\ell, Y\right) &= \frac{\sum_{\boldsymbol{s}[n]\in\mathcal{S}} \boldsymbol{s}[n] p\left(\boldsymbol{y}[n]\middle|\boldsymbol{s}[n], \hat{G}_\ell\right) p\left(\boldsymbol{s}[n]\right)}{\sum_{\boldsymbol{s}[n]\in\mathcal{S}} p\left(\boldsymbol{y}[n]\middle|\boldsymbol{s}[n], \hat{G}_\ell\right) p\left(\boldsymbol{s}[n]\right)} \\ &= \frac{\sum_{\boldsymbol{s}[n]\in\mathcal{S}} \boldsymbol{s}[n] e^{-\frac{1}{\sigma_v^2}\left\|\boldsymbol{y}[n]-\hat{G}_\ell \boldsymbol{s}[n]\right\|^2} p\left(\boldsymbol{s}[n]\right)}{\sum_{\boldsymbol{s}[n]\in\mathcal{S}} e^{-\frac{1}{\sigma_v^2}\left\|\boldsymbol{y}[n]-\hat{G}_\ell \boldsymbol{s}[n]\right\|^2} p\left(\boldsymbol{s}[n]\right)}.\end{aligned} \tag{5.38}$$

Similarly $\mathbb{E}(\boldsymbol{s}[n]\boldsymbol{s}[n]^H|\hat{G}_\ell, Y)$ in the E-step is given by

$$\mathbb{E}\left(\boldsymbol{s}[n]\boldsymbol{s}[n]^H\middle|\hat{G}_\ell, Y\right) = \frac{\sum_{\boldsymbol{s}[n]\in\mathcal{S}} \boldsymbol{s}[n]\boldsymbol{s}[n]^H e^{-\frac{1}{\sigma_v^2}\left\|\boldsymbol{y}[n]-\hat{G}_\ell \boldsymbol{s}[n]\right\|^2} p\left(\boldsymbol{s}[n]\right)}{\sum_{\boldsymbol{s}[n]\in\mathcal{S}} e^{-\frac{1}{\sigma_v^2}\left\|\boldsymbol{y}[n]-\hat{G}_\ell \boldsymbol{s}[n]\right\|^2} p\left(\boldsymbol{s}[n]\right)}. \tag{5.39}$$

In (5.38) and (5.39), term $\left\|\boldsymbol{y}[n]-\hat{G}_\ell \boldsymbol{s}[n]\right\|^2$ needs to be computed for all constellation points in $\mathcal{S}$. Therefore, complexity of the E-step with a discrete model for data symbols grows exponentially with the number of users K.

5.8.B Derivation of the EM Algorithm with Gaussian Prior

Proof. Let $\boldsymbol{g}_{[m]}^T \in \mathbb{C}^{1\times K}$ denote the mth row of channel matrix G. The expected value of the log likelihood in (5.10) can be written as

$$\begin{aligned}\mathcal{L} = \text{const} &- \sum_{n=0}^{L-1}\sum_{m=1}^{M}\frac{1}{\sigma_v^2}\left\|y_m[n]-\boldsymbol{g}_{[m]}^T\boldsymbol{s}[n]\right\|^2 \\ &- \sum_{n=L}^{N-1}\sum_{m=1}^{M}\frac{1}{\sigma_v^2}\mathbb{E}_{p(S_d|Y,\hat{G}_\ell)}\left(\left\|y_m[n]-\boldsymbol{g}_{[m]}^T\boldsymbol{s}[n]\right\|^2\right),\end{aligned} \tag{5.40}$$

where $y_m[n]$ is the mth element of $\boldsymbol{y}[n]$. The likelihood function in (5.40) is a concave function of $\boldsymbol{g}_{[m]}$ and its maximum at the $(\ell+1)$th iteration is obtained by taking the complex gradient of it with respect to $\boldsymbol{g}_{[m]}$ [84], and setting the result to zero as follows

$$\begin{aligned}\hat{\boldsymbol{g}}_{[m]}^T = &\left(\sum_{n=0}^{L-1} y_m[n]\boldsymbol{s}[n]^H + \sum_{n=L}^{N-1} y_m[n]\mathbb{E}\left(\boldsymbol{s}[n]\middle|\hat{G}_\ell, Y\right)^H\right) \\ &\times\left(S_p S_p^H + \sum_{n=L}^{N-1}\mathbb{E}\left(\boldsymbol{s}[n]\boldsymbol{s}[n]^H\middle|\hat{G}_\ell, Y\right)\right)^{-1}.\end{aligned} \tag{5.41}$$

Substituting all values in $\hat{G}_{\ell+1} = \left[\hat{\boldsymbol{g}}_{[1]},\cdots,\hat{\boldsymbol{g}}_{[M]}\right]^T$ gives us the M-step in (5.12b). To compute the E-step, data symbols are assumed to be Gaussian, i.e., $\boldsymbol{s}[n] \sim \mathcal{CN}(0, I_K)$, $n = L,\cdots,N-1$. Given G, S_d and Y are jointly Gaussian with conditional mean and covariance matrix that are given below [85]:

$$\begin{aligned}\mathbb{E}\left(\boldsymbol{s}[n]\middle|\hat{G}_\ell, Y\right) &= \mathbb{E}\left(\boldsymbol{s}[n]\boldsymbol{y}[n]^H\middle|\hat{G}_\ell\right)\mathbb{E}\left(\boldsymbol{y}[n]\boldsymbol{y}[n]^H\middle|\hat{G}_\ell\right)^{-1}\boldsymbol{y}[n] \\ &= \hat{G}_\ell^H\left(\hat{G}_\ell\hat{G}_\ell^H + \sigma_v^2 I_M\right)^{-1}\boldsymbol{y}[n] \\ &\overset{(a)}{=} \left(\hat{G}_\ell^H\hat{G}_\ell + \sigma_v^2 I_K\right)^{-1}\hat{G}_\ell^H\boldsymbol{y}[n],\end{aligned} \tag{5.42}$$

and

$$
\begin{aligned}
&\mathrm{Cov}\left(\boldsymbol{s}[n]\middle|\hat{G}_\ell, Y\right) \\
&= \mathbb{E}\left(\boldsymbol{s}[n]\boldsymbol{s}[n]^H\middle|\hat{G}_\ell\right) - \mathbb{E}\left(\boldsymbol{s}[n]\boldsymbol{y}[n]^H\middle|\hat{G}_\ell\right)\mathbb{E}\left(\boldsymbol{y}[n]\boldsymbol{y}[n]^H\middle|\hat{G}_\ell\right)^{-1}\mathbb{E}\left(\boldsymbol{y}[n]\boldsymbol{s}[n]^H\middle|\hat{G}_\ell\right) \\
&= I_K - \hat{G}_\ell^H\left(\hat{G}_\ell\hat{G}_\ell^H + \sigma_v^2 I_M\right)^{-1}\hat{G}_\ell \\
&\overset{(b)}{=} \sigma_v^2\left(\hat{G}_\ell^H\hat{G}_\ell + \sigma_v^2 I_K\right)^{-1},
\end{aligned} \tag{5.43}
$$

where (a) and (b) follow from matrix inversion lemma [55, Eq. C.7]. □

5.8.C Derivation of the EM Algorithm with Channel Priors

Proof. Let $\boldsymbol{g}_{[m]}^T \in \mathbb{C}^{1\times K}$ denote the mth row of channel matrix G. The expected value of likelihood function with the Gaussian channel priors in (5.13) can be written as

$$
\begin{aligned}
\mathcal{L}(G,B) = \text{const} &- \sum_{n=0}^{L-1}\sum_{m=1}^{M}\frac{1}{\sigma_v^2}\left\|y_m[n] - \boldsymbol{g}_{[m]}^T\boldsymbol{s}[n]\right\|^2 \\
&- \sum_{n=L}^{N-1}\sum_{m=1}^{M}\frac{1}{\sigma_v^2}\mathbb{E}_{p(S_d|Y,\hat{G}_\ell)}\left(\left\|y_m[n] - \boldsymbol{g}_{[m]}^T\boldsymbol{s}[n]\right\|^2\right) \\
&- \sum_{m=1}^{M}\boldsymbol{g}_{[m]}^T B^{-1}\boldsymbol{g}_{[m]}^* - \sum_{m=1}^{M}\sum_{i=1}^{K}\log\beta_i.
\end{aligned} \tag{5.44}
$$

It can be observed that, again, terms $\mathbb{E}(\boldsymbol{s}[n]|\hat{G}_\ell, Y)$ and $\mathbb{E}(\boldsymbol{s}[n]\boldsymbol{s}[n]^H|\hat{G}_\ell, Y)$ appear in (5.44). Therefore, the E-step remains the same as in (5.12a). In the M-step, for simplicity, we maximize $\mathcal{L}(G,B)$ with respect to G and B separately, which doesn't necessarily maximize the likelihood function $\mathcal{L}(G,B)$ but increases it. Holding B constant and taking the complex gradient of $\mathcal{L}(G,B)$ with respect to $\boldsymbol{g}_{[m]}$ and setting it to zero gives us

$$
\begin{aligned}
\hat{\boldsymbol{g}}_{[m]}^T = &\left(\sum_{n=0}^{L-1} y_m[n]\boldsymbol{s}[n]^H + \sum_{n=L}^{N-1} y_m[n]\left(\boldsymbol{\mu}_n^\ell\right)^H\right)\times\hat{B}_\ell \\
&\times\left(S_p S_p^H\hat{B}_\ell + \sum_{n=L}^{N-1}\left(\boldsymbol{\mu}_n^\ell(\boldsymbol{\mu}_n^\ell)^H + \Sigma^\ell\right)\hat{B}_\ell + \sigma_v^2 I_K\right)^{-1}.
\end{aligned} \tag{5.45}
$$

Substituting all values in $\hat{G}_{\ell+1} = \left[\hat{\boldsymbol{g}}_{[1]}, \cdots, \hat{\boldsymbol{g}}_{[M]}\right]^T$ gives (5.14a) and results in a larger likelihood function, i.e., $\mathcal{L}(\hat{G}_{\ell+1}, B) \geq \mathcal{L}(G, B)$. Holding $\hat{G}_{\ell+1}$ constant we solve for $B = \text{diag}\,(\beta_1, \cdots, \beta_K)$ by taking the derivative of $\mathcal{L}(\hat{G}_{\ell+1}, B)$ with respect to $\beta_1, \cdots, \beta_K$ and setting it to zero as follows

$$\beta_i^{\ell+1} = \frac{\left\|\boldsymbol{g}_i^{\ell+1}\right\|^2}{M},\ i = 1, \cdots, K, \tag{5.46}$$

which completes the M step. This step ensures that $\mathcal{L}(\hat{G}_{\ell+1}, \hat{B}_{\ell+1}) \geq \mathcal{L}(\hat{G}_{\ell+1}, B)$, and hence the overall likelihood function increases in the M-step, i.e., $\mathcal{L}(\hat{G}_{\ell+1}, \hat{B}_{\ell+1}) \geq \mathcal{L}(G, B)$. Therefore, this procedure results in a generalized EM algorithm [55]. □

5.8.D Derivation of the EM Algorithm with GMM Prior

Proof. The expected value of likelihood function in (5.17) can be written as

$$\begin{aligned}\mathcal{L} = \text{const} &- \sum_{n=0}^{L-1}\sum_{m=1}^{M} \frac{1}{\sigma_v^2}\left\|y_m[n] - \boldsymbol{g}_{[m]}^T \boldsymbol{s}[n]\right\|^2 \\ &- \sum_{n=L}^{N-1}\sum_{m=1}^{M} \frac{1}{\sigma_v^2}\mathbb{E}_{p(S_d|Y,\hat{\Theta}_\ell)}\left(\left\|y_m[n] - \boldsymbol{g}_{[m]}^T \boldsymbol{s}[n]\right\|^2\right) \\ &- \sum_{n=L}^{N-1}\sum_{k=1}^{K} \frac{1}{\sigma_s^2}\mathbb{E}_{p(S_d|Y,\hat{\Theta}_\ell)}\left(\left|s_k[n] - c_{nk}\right|^2\right),\end{aligned} \tag{5.47}$$

where $\boldsymbol{g}_{[m]}^T \in \mathbb{C}^{1\times K}$ denotes the mth row of G, and c_{nk} is the kth element of $\boldsymbol{c}_n$. Similar to Appendix 5.8.B, the channel estimate at the $(\ell+1)$th iteration is obtained by taking the derivative of the log likelihood $\mathcal{L}$ and setting it to zero, which gives us

$$\begin{aligned}\hat{\boldsymbol{g}}_{[m]}^T = &\left(\sum_{n=0}^{L-1} y_m[n]\boldsymbol{s}[n]^H + \sum_{n=L}^{N-1} y_m[n]\mathbb{E}\left(\boldsymbol{s}[n]\middle|\hat{\Theta}_\ell, Y\right)^H\right) \\ &\times \left(S_p S_p^H + \sum_{n=L}^{N-1}\mathbb{E}\left(\boldsymbol{s}[n]\boldsymbol{s}[n]^H\middle|\hat{\Theta}_\ell, Y\right)\right)^{-1}.\end{aligned} \tag{5.48}$$

By stacking vectors $\hat{\boldsymbol{g}}_{[1]}^T, \cdots, \hat{\boldsymbol{g}}_{[M]}^T$ in $\hat{G}_{\ell+1}$ we get (5.18b). Let $\mathcal{C}$ be a set containing constellation points. The optimal c_{nk} in (5.47) is obtained as follows

$$\begin{aligned} c_{nk}^{\ell+1} &= \underset{c_{nk}\in\mathcal{C}}{\operatorname{argmin}}\, \mathbb{E}_{p(S_d|Y,\hat{\Theta}_\ell)} \left(\left| s_k[n] - c_{nk} \right|^2 \right) \\ &= \underset{c_{nk}\in\mathcal{C}}{\operatorname{argmin}}\, |c_{nk}|^2 - 2\mathrm{Re}\left\{ \mathbb{E}_{p(S_d|Y,\hat{\Theta}_\ell)} \left(s_k[n] \right) c_{nk}^* \right\} \end{aligned} \tag{5.49}$$

which is equivalent to demapping $\mathbb{E}\left(s_k[n] \middle| Y, \hat{\Theta}_\ell \right)$ to the closest constellation point:

$$c_{nk}^{\ell+1} = F\left(\mathbb{E}\left(s_k[n] \middle| Y, \hat{\Theta}_\ell \right) \right). \tag{5.50}$$

In the vector form, $\boldsymbol{c}_n$ can be written as

$$\hat{\boldsymbol{c}}_n^{\ell+1} = F\left(\mathbb{E}\left(\boldsymbol{s}[n] \middle| Y, \hat{\Theta}_\ell \right) \right). \tag{5.51}$$

Given Θ, S_d and Y are jointly Gaussian. Similar to Appendix 5.8.B, $\mathbb{E}(\mathbf{s}[n]|\hat{\Theta}_\ell, Y)$ and $\mathbb{E}(\mathbf{s}[n]\mathbf{s}[n]^H|\hat{\Theta}_\ell, Y)$ that appear in the likelihood function (5.47) can be computed from the conditional density of circularly symmetric Gaussian random vectors as follows [85]

$$\begin{aligned} \mathbb{E}\left(\boldsymbol{s}[n] \middle| \hat{\Theta}_\ell, Y \right) &= \mathbb{E}\left(\boldsymbol{s}[n]\boldsymbol{y}[n]^H \middle| \hat{\Theta}_\ell \right) \mathbb{E}\left(\boldsymbol{y}[n]\boldsymbol{y}[n]^H \middle| \hat{\Theta}_\ell \right)^{-1} \boldsymbol{y}[n] \\ &= \left(\hat{\mathbf{c}}_n^\ell (\hat{\mathbf{c}}_n^\ell)^H + \sigma_s^2 I_K \right) \hat{G}_\ell^H \left(\hat{G}_\ell \left(\hat{\mathbf{c}}_n^\ell (\hat{\mathbf{c}}_n^\ell)^H + \sigma_s^2 I_K \right) \hat{G}_\ell^H + \sigma_v^2 I_M \right)^{-1} \boldsymbol{y}[n] \\ &\overset{(a)}{=} \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 \left(\hat{\mathbf{c}}_n^\ell (\hat{\mathbf{c}}_n^\ell)^H + \sigma_s^2 I_K \right)^{-1} \right)^{-1} \hat{G}_\ell^H \boldsymbol{y}[n], \end{aligned} \tag{5.52}$$

and

$$\begin{aligned} &\mathrm{Cov}\left(\boldsymbol{s}[n] \middle| \hat{\Theta}_\ell, Y \right) \\ &= \mathbb{E}\left(\boldsymbol{s}[n]\boldsymbol{s}[n]^H \middle| \hat{\Theta}_\ell \right) - \mathbb{E}\left(\boldsymbol{s}[n]\boldsymbol{y}[n]^H \middle| \hat{\Theta}_\ell \right) \mathbb{E}\left(\boldsymbol{y}[n]\boldsymbol{y}[n]^H \middle| \hat{\Theta}_\ell \right)^{-1} \mathbb{E}\left(\boldsymbol{y}[n]\boldsymbol{s}[n]^H \middle| \hat{\Theta}_\ell \right) \\ &= \left(\hat{\mathbf{c}}_n^\ell (\hat{\mathbf{c}}_n^\ell)^H + \sigma_s^2 I_K \right) \\ &\quad - \left(\hat{\mathbf{c}}_n^\ell (\hat{\mathbf{c}}_n^\ell)^H + \sigma_s^2 I_K \right) \hat{G}_\ell^H \left(\hat{G}_\ell \left(\hat{\mathbf{c}}_n^\ell (\hat{\mathbf{c}}_n^\ell)^H + \sigma_s^2 I_K \right) \hat{G}_\ell^H + \sigma_v^2 I_M \right)^{-1} \hat{G}_\ell \left(\hat{\mathbf{c}}_n^\ell (\hat{\mathbf{c}}_n^\ell)^H + \sigma_s^2 I_K \right) \\ &\overset{(b)}{=} \sigma_v^2 \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 I_K \left(\hat{\mathbf{c}}_n^\ell (\hat{\mathbf{c}}_n^\ell)^H + \sigma_s^2 I_K \right)^{-1} \right)^{-1}, \end{aligned} \tag{5.53}$$

where (a) and (b) are the result of matrix inversion lemma [55, Eq. C.7].

□

5.8.E Proof of Lemma 5.2

Proof. Define $x_m = |\hat{g}_{mi}|^2$ with probability distribution function $f(x_m)$ for $m = 1, \cdots, M$. Then, the cumulative distribution function of $y_1 = \frac{x_1}{x_1+\cdots+x_M}$ for $0 \leq a < 1$ is given by

$$\begin{aligned} Pr\left\{y_1 \leq a\right\} &= Pr\left\{x_1 \leq \frac{a}{1-a}\left(x_2 + x_3 + \cdots + x_M\right)\right\} \\ &= \int_0^1 \cdots \int_0^1 \int_0^{\frac{a}{1-a}(x_2+x_3+\cdots+x_M)} f(x_1)\cdots f(x_M) dx_1 \cdots dx_M. \end{aligned} \tag{5.54}$$

A change of variables between x_1 and x_2 gives us

$$\begin{aligned} Pr\left\{y_1 \leq a\right\} &= \int_0^1 \cdots \int_0^1 \int_0^{\frac{a}{1-a}(x_1+x_3+\cdots+x_M)} f(x_1)\cdots f(x_M) dx_1 \cdots dx_M \\ &= Pr\left\{y_2 \leq a\right\}, \end{aligned} \tag{5.55}$$

which indicates that $y_1, \cdots, y_M$ are identically distributed and hence have the same mean. The fact that $\sum_{m=1}^{M} y_m = 1$, results in $\mathbb{E}\left(y_m\right) = \frac{1}{M},\ m = 1, \cdots, M$. □

5.8.F Derivation of the Deterministic CRB

Define $\bar{x} \triangleq \text{Re}\left\{x\right\}$ and $\tilde{x} \triangleq \text{Im}\left\{x\right\}$. The log likelihood function of the received signal is given by

$$\mathcal{L} = \text{const} - \sum_{n=0}^{N-1} \frac{1}{\sigma_v^2} \left\| \boldsymbol{y}[n] - G\boldsymbol{s}[n] \right\|^2. \tag{5.56}$$

The CRB of both channel coefficients and unknown data symbols (S_d) is given by

$$\text{CRB}\left(S_d, G\right) = \mathbb{E}\left(\mathcal{I}\mathcal{I}^T\right)^{-1}, \tag{5.57}$$

where

$$\mathcal{I} = \partial\mathcal{L}/\partial\big[\bar{\boldsymbol{s}}[L]^T, \tilde{\boldsymbol{s}}[L]^T, \cdots, \bar{\boldsymbol{s}}[N-1]^T, \tilde{\boldsymbol{s}}[N-1]^T, \bar{\boldsymbol{g}}_1^T, \tilde{\boldsymbol{g}}_1^T, \cdots, \bar{\boldsymbol{g}}_K^T, \tilde{\boldsymbol{g}}_K^T\big]^T. \tag{5.58}$$

Define $A = \frac{2}{\sigma_v^2} G^H G$. Following similar steps as in the proof of [75, Theorem 4.1], for $n = L, \cdots, N-1$, and $k = 1, \cdots, K$, we can show that

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial\bar{\boldsymbol{s}}[n]} &= \frac{2}{\sigma_v^2}\text{Re}\left\{G^H\boldsymbol{v}[n]\right\}, & \frac{\partial\mathcal{L}}{\partial\tilde{\boldsymbol{s}}[n]} &= \frac{2}{\sigma_v^2}\text{Im}\left\{G^H\boldsymbol{v}[n]\right\}, \\
\frac{\partial\mathcal{L}}{\partial\bar{\boldsymbol{g}}_k} &= \frac{2}{\sigma_v^2}\sum_{n=0}^{N-1}\text{Re}\left\{\boldsymbol{v}[n]s_k[n]^*\right\}, & \frac{\partial\mathcal{L}}{\partial\tilde{\boldsymbol{g}}_k} &= \frac{2}{\sigma_v^2}\sum_{n=0}^{N-1}\text{Im}\left\{\boldsymbol{v}[n]s_k[n]^*\right\}.
\end{aligned} \tag{5.59}$$

Using $\mathbb{E}\left(\boldsymbol{v}[n]\boldsymbol{v}[p]^H\right) = \sigma_v^2 I_M \delta(n-p)$ and R3 in [75, Theorem 4.1], for $n, p = L, \cdots, N-1$ and $k = 1, \cdots, K$, we get

$$\begin{aligned}
&\mathbb{E}\left(\frac{\partial\mathcal{L}}{\partial\bar{\boldsymbol{s}}[n]}\Big(\frac{\partial\mathcal{L}}{\partial\bar{\boldsymbol{s}}[p]}\Big)^T\right) = \mathbb{E}\left(\frac{\partial\mathcal{L}}{\partial\tilde{\boldsymbol{s}}[n]}\Big(\frac{\partial\mathcal{L}}{\partial\tilde{\boldsymbol{s}}[p]}\Big)^T\right) = \bar{A}\,\delta(n-p), \\
&\mathbb{E}\left(\frac{\partial\mathcal{L}}{\partial\bar{\boldsymbol{s}}[n]}\Big(\frac{\partial\mathcal{L}}{\partial\tilde{\boldsymbol{s}}[p]}\Big)^T\right) = -\tilde{A}\,\delta(n-p), \\
&\mathbb{E}\left(\frac{\partial\mathcal{L}}{\partial\bar{\boldsymbol{s}}[n]}\Big(\frac{\partial\mathcal{L}}{\partial\bar{\boldsymbol{g}}_k}\Big)^T\right) = \mathbb{E}\left(\frac{\partial\mathcal{L}}{\partial\tilde{\boldsymbol{s}}[n]}\Big(\frac{\partial\mathcal{L}}{\partial\tilde{\boldsymbol{g}}_k}\Big)^T\right) = \bar{W}_k[n], \\
&\mathbb{E}\left(\frac{\partial\mathcal{L}}{\partial\tilde{\boldsymbol{s}}[n]}\Big(\frac{\partial\mathcal{L}}{\partial\bar{\boldsymbol{g}}_k}\Big)^T\right) = -\mathbb{E}\left(\frac{\partial\mathcal{L}}{\partial\bar{\boldsymbol{s}}[n]}\Big(\frac{\partial\mathcal{L}}{\partial\tilde{\boldsymbol{g}}_k}\Big)^T\right) = \tilde{W}_k[n], \\
&\mathbb{E}\left(\frac{\partial\mathcal{L}}{\partial\bar{\boldsymbol{g}}_k}\Big(\frac{\partial\mathcal{L}}{\partial\bar{\boldsymbol{g}}_i}\Big)^T\right) = \mathbb{E}\left(\frac{\partial\mathcal{L}}{\partial\tilde{\boldsymbol{g}}_k}\Big(\frac{\partial\mathcal{L}}{\partial\tilde{\boldsymbol{g}}_i}\Big)^T\right) = \bar{\lambda}_{ki} I_M, \\
&\mathbb{E}\left(\frac{\partial\mathcal{L}}{\partial\tilde{\boldsymbol{g}}_k}\Big(\frac{\partial\mathcal{L}}{\partial\bar{\boldsymbol{g}}_i}\Big)^T\right) = -\mathbb{E}\left(\frac{\partial\mathcal{L}}{\partial\bar{\boldsymbol{g}}_k}\Big(\frac{\partial\mathcal{L}}{\partial\tilde{\boldsymbol{g}}_i}\Big)^T\right) = \tilde{\lambda}_{ki} I_M.
\end{aligned} \tag{5.60}$$

Substituting (5.60) in (5.57), gives us

$$\text{CRB}\,(S_d, G)^{-1} = \left[\begin{array}{ccc|c}
\begin{bmatrix} \bar{A} & -\tilde{A} \\ \tilde{A} & \bar{A} \end{bmatrix} & & 0 & \Omega_L \\
& \ddots & & \vdots \\
0 & & \begin{bmatrix} \bar{A} & -\tilde{A} \\ \tilde{A} & \bar{A} \end{bmatrix} & \Omega_{N-1} \\
\hline
\Omega_L^T & \cdots & \Omega_{N-1}^T & \Lambda
\end{array}\right], \tag{5.61}$$

where Λ and $\Omega_n,\ n = L, \cdots, N-1$ are defined in (5.29). The CRB of channel matrix G is the $2MK \times 2MK$ submatrix formed of intersection of the last $2MK$ rows and $2MK$ columns of CRB (S_d, G). By results on the inverse of a block matrix, we can show that

$$\text{CRB}\,(G) = \left(\Lambda - \sum_{n=L}^{N-1} \Omega_n^T \begin{bmatrix} \bar{C} & -\tilde{C} \\ \tilde{C} & \bar{C} \end{bmatrix} \Omega_n \right)^{-1}. \tag{5.62}$$

5.8.G Proof of Theorem 5.1

We will need the following preparatory lemma to prove Theorem 5.1.

Lemma 5.3. *Denote by $\boldsymbol{g}_{[m]}^T \in \mathbb{C}^{1\times K}$ the mth row of channel matrix $G \in \mathbb{C}^{M\times K}$ defined in (5.1). Then, for any $m, m' \in \{1, \cdots, M\}$, we have*

$$\boldsymbol{g}_{[m]}^T \left(G^H G\right)^{-1} \boldsymbol{g}_{[m']}^* \xrightarrow[M\to\infty]{a.s.} 0. \tag{5.63}$$

Proof. Let $\boldsymbol{h}_{[m]}^T$ be the mth row of small scale fading matrix H in (5.1) and define $\Psi_{(m)} = \left(\frac{1}{M} H^H H - \frac{1}{M} \boldsymbol{h}_{[m]}^* \boldsymbol{h}_{[m]}^T\right)^{-1}$ with elements $\left[\Psi_{(m)}\right]_{ij} = \psi_{ij}$. Note that $\Psi_{(m)}$ and $\boldsymbol{h}_{[m]}$ are statistically independent. For any $A \in \mathbb{C}^{M\times M}$ and $B \in \mathbb{C}^{M\times M}$ by Cauchy-Schwarz inequality, we have

$$\left|\text{tr}\,(AB)\right| \le \sqrt{\text{tr}\,(AA^H)\,\text{tr}\,(BB^H)} \le M \left\|A\right\|_2 \left\|B\right\|_2. \tag{5.64}$$

By applying matrix inversion lemma [53, Lemma 6.2], for $m = m'$, we obtain

$$\boldsymbol{g}_{[m]}^T \left(G^H G\right)^{-1} \boldsymbol{g}_{[m']}^* = \frac{\frac{1}{M} \boldsymbol{h}_{[m]}^T \Psi_{(m)} \boldsymbol{h}_{[m]}^*}{1 + \frac{1}{M} \boldsymbol{h}_{[m]}^T \Psi_{(m)} \boldsymbol{h}_{[m]}^*}. \tag{5.65}$$

In order to prove (5.63), for $m = m'$, we will show that $\mathbb{E}\left(\left|\frac{1}{M} \boldsymbol{h}_{[m]}^T \Psi_{(m)} \boldsymbol{h}_{[m]}^*\right|^2\right) \le \frac{c}{M^2}$ for

some constant c as follows

$$\begin{aligned}
\mathbb{E}\left(\left|\frac{1}{M}\boldsymbol{h}_{[m]}^{T}\Psi_{(m)}\boldsymbol{h}_{[m]}^{*}\right|^{2}\right) &= \frac{1}{M^{2}}\sum_{i,i'=1}^{K}\sum_{j,j'=1}^{K}\mathbb{E}\left(h_{mi}h_{mi'}^{*}\right)\mathbb{E}\left(h_{mj}^{*}h_{mj'}\right)\mathbb{E}\left(\psi_{ij}\psi_{i'j'}^{*}\right) \\
&\quad + \frac{1}{M^{2}}\sum_{i,i'=1}^{K}\sum_{j,j'=1}^{K}\mathbb{E}\left(h_{mi}h_{mj}^{*}\right)\mathbb{E}\left(h_{mi'}^{*}h_{mj'}\right)\mathbb{E}\left(\psi_{ij}\psi_{i'j'}^{*}\right) \\
&= \frac{1}{M^{2}}\mathbb{E}\left(\mathrm{tr}\left(\Psi_{(m)}\Psi_{(m)}^{H}\right)\right) + \frac{1}{M^{2}}\mathbb{E}\left(\left|\mathrm{tr}\left(\Psi_{(m)}\right)\right|^{2}\right) \\
&\leq \frac{K}{M^{2}}\mathbb{E}\left(\left\|\Psi_{(m)}\right\|_{2}^{2}\right) + \frac{K^{2}}{M^{2}}\mathbb{E}\left(\left\|\Psi_{(m)}\right\|_{2}^{2}\right), \qquad (5.66)
\end{aligned}$$

where the last inequality follows from (5.64). Since $\frac{1}{M}H^{H}H$ is non-singular with probability one for all large M [53], the maximum eigenvalue of matrix $\Psi_{(m)}$ is bounded for all large M almost surely. By dominated convergence theorem [86], $\mathbb{E}\left(\left\|\Psi_{(m)}\right\|_{2}^{2}\right)$ is uniformly bounded. Therefore, $\mathbb{E}\Big(\left|\frac{1}{M}\boldsymbol{h}_{[m]}^{T}\Psi_{(m)}\boldsymbol{h}_{[m]}^{*}\right|^{2}\Big)$ is of order $\mathcal{O}(M^{-2})$ and by Borel-Cantelli lemma it follows that

$$\frac{1}{M}\boldsymbol{h}_{[m]}^{T}\Psi_{(m)}\boldsymbol{h}_{[m]}^{*} \xrightarrow[M\to\infty]{a.s.} 0. \qquad (5.67)$$

By matrix inversion lemma, for $m \neq m'$, term $\boldsymbol{g}_{[m]}^{T}(G^{H}G)^{-1}\boldsymbol{g}_{[m']}^{*}$ is equal to

$$\frac{\frac{1}{M}\boldsymbol{h}_{[m]}^{T}\Psi_{(mm')}\boldsymbol{h}_{[m']}^{*}}{\left(1+\frac{1}{M}\boldsymbol{h}_{[m]}^{T}\Psi_{(m)}\boldsymbol{h}_{[m]}^{*}\right)\left(1+\frac{1}{M}\boldsymbol{h}_{[m']}^{T}\Psi_{(mm')}\boldsymbol{h}_{[m']}^{*}\right)}, \qquad (5.68)$$

where $\Psi_{(mm')} = \left(\frac{1}{M}H^{H}H - \frac{1}{M}\boldsymbol{h}_{[m]}^{*}\boldsymbol{h}_{[m]}^{T} - \frac{1}{M}\boldsymbol{h}_{[m']}^{*}\boldsymbol{h}_{[m']}^{T}\right)^{-1}$. Note that $\boldsymbol{h}_{[m]}$, $\boldsymbol{h}_{[m']}$, and $\Psi_{(mm')}$ are independent for $m \neq m'$. Similarly, we can show that $\mathbb{E}\Big(\left|\frac{1}{M}\boldsymbol{h}_{[m]}^{T}\Psi_{(mm')}\boldsymbol{h}_{[m']}^{*}\right|^{2}\Big)$ is of order $\mathcal{O}(M^{-2})$ and hence

$$\frac{1}{M}\boldsymbol{h}_{[m]}^{T}\Psi_{(mm')}\boldsymbol{h}_{[m']}^{*} \xrightarrow[M\to\infty]{a.s.} 0. \qquad (5.69)$$

Combining all the results yields

$$\boldsymbol{g}_{[m]}^{T}\left(G^{H}G\right)^{-1}\boldsymbol{g}_{[m']}^{*} \xrightarrow[M\to\infty]{a.s.} 0. \qquad (5.70)$$

□

Proof of Theorem 5.1. For notational simplicity, we define $D_{kl}[n] = W_k[n]^H C W_l[n]$. After some manipulation, submatrices in $\Omega_n^T \begin{bmatrix} \bar{C} & -\tilde{C} \\ \tilde{C} & \bar{C} \end{bmatrix} \Omega_n$ can be simplified as follows

$$\begin{bmatrix} \bar{W}_k[n]^T & \tilde{W}_k[n]^T \\ -\tilde{W}_k[n]^T & \bar{W}_k[n]^T \end{bmatrix} \begin{bmatrix} \bar{C} & -\tilde{C} \\ \tilde{C} & \bar{C} \end{bmatrix} \begin{bmatrix} \bar{W}_l[n] & -\tilde{W}_l[n] \\ \tilde{W}_l[n] & \bar{W}_l[n] \end{bmatrix} = \begin{bmatrix} \bar{D}_{kl}[n] & -\tilde{D}_{kl}[n] \\ \tilde{D}_{kl}[n] & \bar{D}_{kl}[n] \end{bmatrix}. \tag{5.71}$$

Denote by $\boldsymbol{g}_{[m]}^T \in \mathbb{C}^{1\times K}$ the mth row of channel matrix G defined in (5.1). The (m, m')th component of $D_{kl}[n]$ is then given by

$$[D_{kl}[n]]_{mm'} = \frac{2 s_k[n] s_l[n]^*}{\sigma_v^2} \boldsymbol{g}_{[m]}^T \left(G^H G\right)^{-1} \boldsymbol{g}_{[m']}^* \xrightarrow[M\to\infty]{(a)} 0, \tag{5.72}$$

where (a) follows from Lemma 5.3, which shows that each element of matrix $\Omega_n^T \begin{bmatrix} \bar{C} & -\tilde{C} \\ \tilde{C} & \bar{C} \end{bmatrix} \Omega_n$ converges to zero almost surely. Therefore, by continuous mapping theorem [56] we have

$$\text{CRB}(G) \xrightarrow[M\to\infty]{a.s.} \Lambda^{-1}. \tag{5.73}$$

□

5.8.H Derivation of the Stochastic CRB

The log likelihood function of the received signal with Gaussian data symbols is given by

$$\mathcal{L} = \text{const} - \sum_{n=0}^{L-1} \underbrace{\frac{1}{\sigma_v^2} \left\| \boldsymbol{y}[n] - G\boldsymbol{s}[n] \right\|^2}_{\mathcal{L}_1[n]} - \sum_{n=L}^{N-1} \underbrace{\left(\boldsymbol{y}[n]^H R^{-1} \boldsymbol{y}[n] - \log\det R \right)}_{\mathcal{L}_2[n]}. \tag{5.74}$$

Terms $\mathcal{L}_1[0], \cdots, \mathcal{L}_1[L-1], \mathcal{L}_2[L], \cdots, \mathcal{L}_2[N-1]$ are independent. Thus, the CRB of covariance matrix is given by

$$\left[\text{CRB}^{-1}\right]_{ij} = \sum_{n=0}^{L-1} \mathbb{E}\left(\frac{\partial \mathcal{L}_1[n]}{\partial \alpha_i} \frac{\partial \mathcal{L}_1[n]}{\partial \alpha_j} \right) + \sum_{n=L}^{N-1} \mathbb{E}\left(\frac{\partial \mathcal{L}_2[n]}{\partial \alpha_i} \frac{\partial \mathcal{L}_2[n]}{\partial \alpha_j} \right). \tag{5.75}$$

From (5.32), for $n = 0, \cdots, L-1$, we have

$$\mathbb{E}\left(\frac{\partial \mathcal{L}_1[n]}{\partial \bar{g}_{mk}}\frac{\partial \mathcal{L}_1[n]}{\partial \bar{g}_{rl}}\right) = \mathbb{E}\left(\frac{\partial \mathcal{L}_1[n]}{\partial \tilde{g}_{mk}}\frac{\partial \mathcal{L}_1[n]}{\partial \tilde{g}_{rl}}\right) = \frac{2}{\sigma_v^2}\text{Re}\left\{s_k[n]^* s_l[n]\right\}\delta(m-r),$$
$$\mathbb{E}\left(\frac{\partial \mathcal{L}_1[n]}{\partial \bar{g}_{mk}}\frac{\partial \mathcal{L}_1[n]}{\partial \tilde{g}_{rl}}\right) = -\frac{2}{\sigma_v^2}\text{Im}\left\{s_k[n]^* s_l[n]\right\}\delta(m-r). \tag{5.76}$$

Suppose pilot sequences are orthogonal, i.e., $\sum_{n=0}^{L-1} s_k[n]^* s_l[n] = L\delta(k-l)$. We then have

$$\mathbb{E}\left(\frac{\partial \mathcal{L}_1[n]}{\partial \alpha_i}\frac{\partial \mathcal{L}_1[n]}{\partial \alpha_j}\right) = \frac{2L}{\sigma_v^2}\delta(i-j). \tag{5.77}$$

From (5.32), it follows that

$$\mathbb{E}\left(\frac{\partial \mathcal{L}_2[n]}{\partial \alpha_i}\frac{\partial \mathcal{L}_2[n]}{\partial \alpha_j}\right) = \text{tr}\left(R^{-1}\frac{\partial R}{\partial \alpha_i}R^{-1}\frac{\partial R}{\partial \alpha_j}\right). \tag{5.78}$$

Substituting (5.77) and (5.78) in (5.75) yields

$$\left[\text{CRB}^{-1}\right]_{ij} = \frac{2L}{\sigma_v^2}\delta(i-j) + (N-L)\text{tr}\left(R^{-1}\frac{\partial R}{\partial \alpha_i}R^{-1}\frac{\partial R}{\partial \alpha_j}\right). \tag{5.79}$$

5.8.I Proof of Theorem 5.2

We first provide the following preparatory results which are used in proving Theorem 5.2.

Lemma 5.4. *Let $R = GG^H + \sigma_v^2 I_M$, where $G \in \mathbb{C}^{M\times K}$ is the channel matrix defined in (5.1). Assume columns of matrix G, denoted by $\boldsymbol{g}_k \sim \mathcal{CN}(0, \beta_k I_M)$, are independent and $P \in \mathcal{C}^{M\times M}$ is a deterministic matrix with uniformly bounded spectral norm. Then,*

$$\text{tr}\left(R^{-1}P\right) - \frac{\text{tr}\left(P\right)}{\sigma_v^2} \xrightarrow[M\to\infty]{a.s.} 0. \tag{5.80}$$

Proof. When M and K grow large such that $0 < \liminf_M \frac{M}{K} \leq \limsup_M \frac{M}{K} < \infty$, we have [54, Theorem 1]

$$\frac{1}{M}\text{tr}\left((\frac{1}{M}R)^{-1}P\right) - \frac{1}{M}\text{tr}\left(TP\right) \xrightarrow[M\to\infty]{a.s.} 0, \tag{5.81}$$

where

$$T = \frac{M}{\sum_{j=1}^{K} \frac{\beta_j}{1+e_{M,j}} + \sigma_v^2} I_M, \tag{5.82}$$

and $e_{M,1}, \cdots, e_{M,K}$ form the unique solution of

$$e_{M,i} = \frac{M\beta_i}{\sum_{j=1}^{K} \frac{\beta_j}{1+e_{M,j}} + \sigma_v^2}. \tag{5.83}$$

Suppose $M \to \infty$ such that $\frac{K}{M} \to 0$. We can show that $e_{M,i}$ is of order $\mathcal{O}(M)$ and consequently $T \to \frac{M}{\sigma_v^2} I_M$. Therefore,

$$\text{tr}\left(R^{-1}P\right) - \frac{\text{tr}\left(P\right)}{\sigma_v^2} \xrightarrow[M\to\infty]{a.s.} 0. \tag{5.84}$$

□

Corollary 5.1. *Let* $\boldsymbol{e}_m = \left[\mathbf{0}_{1\times(m-1)},\ 1,\ \mathbf{0}_{1\times(M-m)}\right]^T$*, where* $\mathbf{0}_{1\times(m-1)}$ *is a zero vector of length* $(m-1)$*. We then have*

$$\boldsymbol{e}_m^H R^{-1} \boldsymbol{e}_n \xrightarrow[M\to\infty]{a.s.} \frac{\delta(m-n)}{\sigma_v^2}. \tag{5.85}$$

Proof. Substituting $P = \boldsymbol{e}_n \boldsymbol{e}_m^H$ in Lemma 5.4 completes the proof. □

Corollary 5.2. *Let* $R_{(k)} = GG^H + \sigma_v^2 I_M - \boldsymbol{g}_k \boldsymbol{g}_k^H$*, where* G *is the channel matrix defined in (5.1), and suppose* $\boldsymbol{g}_k \sim \mathcal{CN}\left(0, \beta_k I_M\right)$ *is the* k*th column of* G*. Then,*

$$\frac{1}{M} \boldsymbol{g}_k^H R_{(k)}^{-1} \boldsymbol{g}_k \xrightarrow[M\to\infty]{a.s.} \frac{\beta_k}{\sigma_v^2}. \tag{5.86}$$

Proof. Let $\boldsymbol{h}_k$ denote the kth column of small scale fading matrix H in (5.1). Note that $\boldsymbol{h}_k$ and $R_{(k)}$ are independent, and matrix $R_{(k)}^{-1}$ has uniformly bounded spectral norm for all large M, i.e., $\limsup_M \left\|R_{(k)}^{-1}\right\|_2 \leq \frac{1}{\sigma_v^2} < \infty$. Thus, we have

$$\frac{1}{M} \boldsymbol{g}_k^H R_{(k)}^{-1} \boldsymbol{g}_k = \frac{1}{M} \beta_k \boldsymbol{h}_k^H R_{(k)}^{-1} \boldsymbol{h}_k \xrightarrow[M\to\infty]{(a)} \frac{\beta_k}{M} \text{tr}\left(R_{(k)}^{-1}\right) \xrightarrow[M\to\infty]{(b)} \frac{\beta_k}{\sigma_v^2}, \tag{5.87}$$

where (a) and (b) follow from [53, Theorem 3.4] and Lemma 5.4 respectively. □

Lemma 5.5. *Define* $\boldsymbol{e}_m = \left[\mathbf{0}_{1\times(m-1)},\ 1,\ \mathbf{0}_{1\times(M-m)}\right]^T$ *and* $R_{(k)} = GG^H + \sigma_v^2 I_M - \boldsymbol{g}_k\boldsymbol{g}_k^H$, *where* G *is the channel matrix defined in (5.1) and* $\boldsymbol{g}_k \sim \mathcal{CN}\left(0, \beta_k I_M\right)$ *is the* k*th column of* G. *Then,*

$$\frac{1}{M}\boldsymbol{e}_m^H R_{(k)}^{-1}\boldsymbol{g}_k \xrightarrow[M\to\infty]{a.s.} 0. \tag{5.88}$$

Proof. Define $A = R_{(k)}^{-1}$ with elements $[A]_{ij} = a_{ij}$ and let $h_{mk},\ m = 1,\cdots,M,\ k = 1,\cdots,K$ denote the elements of small scale fading matrix H in (5.1). It is clear that A and $\boldsymbol{g}_k$ are independent. In order to prove (5.88), we will show that $\mathbb{E}\big(\left|\frac{1}{M}\boldsymbol{e}_m^H A\boldsymbol{g}_k\right|^4\big) \le \frac{c}{M^2}$ for some constant c as follows

$$\begin{aligned}
&\mathbb{E}\left(\left|\frac{1}{M}\boldsymbol{e}_m^H A\boldsymbol{g}_k\right|^4\right) = \mathbb{E}\left(\frac{\beta_k^2}{M^4}\left|\sum_{i=1}^{M} a_{mi}h_{ik}\right|^4\right)\\
&= \frac{\beta_k^2}{M^4}\sum_{\substack{i_1,i_2\\ i_3,i_4}}^{M}\mathbb{E}\left(a_{mi_1}a_{mi_2}^* a_{mi_3}a_{mi_4}^*\right)\Big[\mathbb{E}\left(h_{i_1k}h_{i_2k}^*\right)\mathbb{E}\left(h_{i_3k}h_{i_4k}^*\right) + \mathbb{E}\left(h_{i_1k}h_{i_4k}^*\right)\mathbb{E}\left(h_{i_2k}h_{i_3k}^*\right)\Big]\\
&= \frac{2\beta_k^2}{M^4}\mathbb{E}\left(\Big(\sum_{i=1}^{M}\left|a_{mi}\right|^2\Big)^2\right) \le \frac{2\beta_k^2}{M^4}\mathbb{E}\left(\mathrm{tr}\left(AA^H\right)^2\right)\\
&\overset{(a)}{\le} \frac{2\beta_k^2}{M^2}\mathbb{E}\left(\|A\|_2^4\right),
\end{aligned} \tag{5.89}$$

where (a) comes from (5.64).

Matrix A has uniformly bounded spectral norm, i.e., $\limsup_M \|A\|_2 \le \frac{1}{\sigma_v^2} < \infty$. By dominated convergence theorem [86], $\mathbb{E}\big(\|A\|_2^4\big)$ is also uniformly bounded and hence $\frac{1}{M^2}\mathbb{E}\big(\|A\|_2^4\big)$ is of order $\mathcal{O}(M^{-2})$. Therefore, $\mathbb{E}\big(\left|\frac{1}{M}\boldsymbol{e}_m^H A\boldsymbol{g}_k\right|^4\big)$ is summable and by Borel-Cantelli lemma it follows that

$$\frac{1}{M}\boldsymbol{e}_m^H R_{(k)}^{-1}\boldsymbol{g}_k \xrightarrow[M\to\infty]{a.s.} 0. \tag{5.90}$$

□

Proof of Theorem 5.2. Define the following quantities:

$$R_{(k)} = R - \boldsymbol{g}_k \boldsymbol{g}_k^H, \qquad R_{(kl)} = R_{(k)} - \boldsymbol{g}_l \boldsymbol{g}_l^H,$$
$$T_{mk} = \left[\mathbf{0}_{M\times(m-1)},\ \boldsymbol{g}_k,\ \mathbf{0}_{M\times(M-m)}\right], \qquad \boldsymbol{e}_m = \left[\mathbf{0}_{1\times(m-1)},\ 1,\ \mathbf{0}_{1\times(M-m)}\right]^T, \tag{5.91}$$

where $\mathbf{0}_{M\times(m-1)}$ is a zero matrix of size $M \times (m-1)$. Note that, for $l \neq k$, quantities $\boldsymbol{g}_k$, $\boldsymbol{g}_l$, $R_{(k)}$, and $R_{(kl)}$ are mutually independent. The second term in (5.33) for different values of α_i and α_j can be rewritten as follows

$$\begin{aligned}
\text{tr}\left(R^{-1}\frac{\partial R}{\partial \bar{g}_{mk}} R^{-1}\frac{\partial R}{\partial \bar{g}_{nl}}\right) &= 2\text{Re}\left\{\text{tr}\left(R^{-1}T_{mk}R^{-1}T_{nl}\right)\right\} + 2\text{Re}\left\{\text{tr}\left(R^{-1}T_{mk}R^{-1}T_{nl}^H\right)\right\},\\
\text{tr}\left(R^{-1}\frac{\partial R}{\partial \tilde{g}_{mk}} R^{-1}\frac{\partial R}{\partial \tilde{g}_{nl}}\right) &= -2\text{Re}\left\{\text{tr}\left(R^{-1}T_{mk}R^{-1}T_{nl}\right)\right\} + 2\text{Re}\left\{\text{tr}\left(R^{-1}T_{mk}R^{-1}T_{nl}^H\right)\right\},\\
\text{tr}\left(R^{-1}\frac{\partial R}{\partial \bar{g}_{mk}} R^{-1}\frac{\partial R}{\partial \tilde{g}_{nl}}\right) &= 2\text{Im}\left\{\text{tr}\left(R^{-1}T_{mk}R^{-1}T_{nl}\right)\right\} - 2\text{Im}\left\{\text{tr}\left(R^{-1}T_{mk}R^{-1}T_{nl}^H\right)\right\},\\
\text{tr}\left(R^{-1}\frac{\partial R}{\partial \tilde{g}_{mk}} R^{-1}\frac{\partial R}{\partial \bar{g}_{nl}}\right) &= 2\text{Im}\left\{\text{tr}\left(R^{-1}T_{mk}R^{-1}T_{nl}\right)\right\} + 2\text{Im}\left\{\text{tr}\left(R^{-1}T_{mk}R^{-1}T_{nl}^H\right)\right\}.
\end{aligned} \tag{5.92}$$

By matrix inversion lemma [53, Lemma 6.2], we have

$$\begin{aligned}
\text{tr}\left(R^{-1}T_{mk}R^{-1}T_{nl}\right) &= \left(\boldsymbol{e}_n^H R^{-1}\boldsymbol{g}_k\right)\left(\boldsymbol{e}_m^H R^{-1}\boldsymbol{g}_l\right)\\
&= \frac{\boldsymbol{e}_n^H R_{(k)}^{-1}\boldsymbol{g}_k}{1+\boldsymbol{g}_k^H R_{(k)}^{-1}\boldsymbol{g}_k}\frac{\boldsymbol{e}_m^H R_{(l)}^{-1}\boldsymbol{g}_l}{1+\boldsymbol{g}_l^H R_{(l)}^{-1}\boldsymbol{g}_l},
\end{aligned} \tag{5.93}$$

By Corollary 5.2, $\frac{1}{M}\boldsymbol{g}_k^H R_{(k)}^{-1}\boldsymbol{g}_k \to \frac{\beta_k}{\sigma_v^2}$, and $\frac{1}{M}\boldsymbol{g}_l^H R_{(l)}^{-1}\boldsymbol{g}_l \to \frac{\beta_l}{\sigma_v^2}$, almost surely. By Lemma 5.5, it follows that $\frac{1}{M}\boldsymbol{e}_n^H R_{(k)}^{-1}\boldsymbol{g}_k \to 0$ and $\frac{1}{M}\boldsymbol{e}_m^H R_{(l)}^{-1}\boldsymbol{g}_l \to 0$. We therefore have

$$\text{tr}\left(R^{-1}T_{mk}R^{-1}T_{nl}\right) \xrightarrow[M\to\infty]{a.s.} 0. \tag{5.94}$$

By applying matrix inversion lemma to term $\text{tr}\left(R^{-1}T_{mk}R^{-1}T_{nl}^H\right)$ in (5.92), we obtain

$$\text{tr}\left(R^{-1}T_{mk}R^{-1}T_{nl}^H\right) = \frac{\boldsymbol{e}_m^H R^{-1}\boldsymbol{e}_n}{1+\boldsymbol{g}_k^H R_{(k)}^{-1}\boldsymbol{g}_k}\begin{cases}\boldsymbol{g}_k^H R_{(k)}^{-1}\boldsymbol{g}_k & l=k\\ \dfrac{\boldsymbol{g}_l^H R_{(kl)}^{-1}\boldsymbol{g}_k}{1+\boldsymbol{g}_l^H R_{(kl)}^{-1}\boldsymbol{g}_l} & l\neq k\end{cases} \tag{5.95}$$

It follows from Corollaries 5.1 and 5.2 that, for $l = k$, tr $\left(R^{-1}T_{mk}R^{-1}T_{nl}^H\right) \rightarrow \frac{\delta(n-m)}{\sigma_v^2}$ almost surely. From [53, Theorem 3.7], for $l \neq k$, we know that $\frac{1}{M}\boldsymbol{g}_l^H R_{(kl)}^{-1}\boldsymbol{g}_k \rightarrow 0$ almost surely. Thus, for $l \neq k$, tr $\left(R^{-1}T_{mk}R^{-1}T_{nl}^H\right) \rightarrow 0$ almost surely. Therefore,

$$\text{tr}\left(R^{-1}T_{mk}R^{-1}T_{nl}^H\right) \xrightarrow[M\rightarrow\infty]{a.s.} \frac{\delta(n-m)\delta(k-l)}{\sigma_v^2}. \tag{5.96}$$

Substituting (5.94) and (5.96) in (5.92) yields

$$\text{tr}\left(R^{-1}\frac{\partial R}{\partial \alpha_i}R^{-1}\frac{\partial R}{\partial \alpha_j}\right) \xrightarrow[M\rightarrow\infty]{a.s.} \frac{2\delta(i-j)}{\sigma_v^2}. \tag{5.97}$$

Thus, $\left[\text{CRB}^{-1}\right]_{ij} \xrightarrow[M\rightarrow\infty]{a.s.} \frac{2N}{\sigma_v^2}\delta\left(i-j\right)$. By the continuous mapping theorem [56], we therefore have

$$\text{CRB} \xrightarrow[M\rightarrow\infty]{a.s.} \frac{\sigma_v^2}{2N}I_{2MK}. \tag{5.98}$$

□

Bibliography

[1] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, July 2017.

[2] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and H. Yang, "Cell-free massive MIMO system/s," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2015, pp. 695–699.

[3] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and B. D. Rao, "Performance of cell-free massive MIMO systems with MMSE and LSFD receivers," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2016, pp. 203–207.

[4] E. Nayebi and B. D. Rao, "Semi-blind channel estimation for multiuser massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 540–553, Jan. 2018.

[5] E. Nayebi and B. D. Rao, "Semi-blind channel estimation in massive MIMO systems with different priors on data symbols," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP) 2018*, to appear.

[6] E. Telatar, "Capacity of multi-antenna gaussian channels," *Transactions on Emerging Telecommunications Technologies*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[7] 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, Mar. 1998.

[8] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[9] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, UK: Cambridge University Press, 2016.

[10] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE*

Signal Process. Mag., vol. 30, no. 1, pp. 40–60, Jan. 2013.

[11] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[12] D. Gesbert, M. Kountouris, R. W. Heath Jr, C.-B. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sept. 2007.

[13] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[14] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

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

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

[17] C. Shepard, H. Yu, N. Anand, L. E. Li, T. L. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proc. of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 53–64.

[18] Y. Ding and B. D. Rao, "Channel estimation using joint dictionary learning in (fdd) massive mimo systems," in *Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP),*, 2015, pp. 185–189.

[19] X. Rao, V. K. Lau, and X. Kong, "CSIT estimation and feedback for FDD multi-user massive MIMO systems," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2014, pp. 3157–3161.

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

[21] A. Ashikhmin and T. L. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. IEEE Symp. Inf. Theory (ISIT)*, 2012, pp. 1137–1141.

[22] A. Adhikary, A. Ashikhmin, and T. L. Marzetta, "Uplink interference reduction in large scale antenna systems," in *Proc. IEEE Symp. Inf. Theory (ISIT)*, 2014, pp. 2529–2533.

[23] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Mar. 2012, pp. 3249–3252.

[24] H. Yang and T. L. Marzetta, "A macro cellular wireless network with uniformly high user throughputs," in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2014.

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

[26] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[27] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[28] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated multipoint transmission/reception techniques for LTE-advanced [Coordinated and Distributed MIMO]," *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 26–34, Jun. 2010.

[29] J. Joung, Y. K. Chia, and S. Sun, "Energy-efficient, large-scale distributed-antenna system (L-DAS) for multiple users," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 954–965, Oct. 2014.

[30] A. Lozano, R. W. Heath, and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5213–5226, Sep. 2013.

[31] Y. Huang, G. Zheng, M. Bengtsson, K. K. Wong, L. Yang, and B. Ottersten, "Distributed multicell beamforming with limited intercell coordination," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 728–738, Feb. 2011.

[32] O. Somekh, O. Simeone, Y. Bar-Ness, A. M. Haimovich, and S. Shamai, "Cooperative multicell zero-forcing beamforming in cellular downlink channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3206–3219, Jul. 2009.

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

[34] K. T. Truong and R. W. Heath, "The viability of distributed antennas for massive MIMO systems," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 2013, pp. 1318–1323.

[35] H. Yang and T. L. Marzetta, "Capacity performance of multicell large-scale antenna systems," in *Proc. 51st Allerton Conf. Commun., Control, Comput.*, Urbana-Champaign, Illinois, Oct. 2013.

[36] J. Wang and L. Dai, "Asymptotic rate analysis of downlink multi-user systems with co-located and distributed antennas," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3046–3058, Jun. 2015.

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

[38] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO: Uniformly great service for everyone," in *Proc. 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 201–205.

[39] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[40] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Oct. 2006, pp. 359–363.

[41] S. M. Kay, "Fundamentals of statistical signal processing, volume i: estimation theory," 1993.

[42] A. Ashikhmin, T. L. Marzetta, and L. Li, "Interference reduction in multi-cell massive MIMO systems I: large-scale fading precoding and decoding," *arXiv preprint arXiv:1411.4182*, 2014.

[43] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[44] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.

[45] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Veh. Technol. Mag.*, vol. 6, no. 1, pp. 37–43, Mar. 2011.

[46] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[47] R. J. Plemmons, "M-matrix characterizations. iâĂŤnonsingular M-matrices," *Linear Algebra Appl.*, vol. 18, no. 2, pp. 175–188, 1977.

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

[49] S. Wagner, R. Couillet, M. Debbah, and D. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.

[50] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge, U.K.: Cambridge university press, 1999.

[51] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.

[52] A. Tang, J. Sun, and K. Gong, "Mobile propagation loss with a low base station antenna for NLOS street microcells in urban area," in *Proc. IEEE Veh. Technol. Conf.*, May 2001, pp. 333–336.

[53] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[54] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.

[55] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[56] A. W. Van der Vaart, *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, New York, 2000.

[57] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.

[58] J. Zhang, C. K. Wen, S. Jin, X. Gao, and K. K. Wong, "On capacity of large-scale MIMO multiple access channels with distributed sets of correlated antennas," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 133–148, Feb. 2013.

[59] S. Zhou, M. Zhao, X. Xu, J. Wang, and Y. Yao, "Distributed wireless communication system: a new architecture for future public wireless access," *IEEE Commun. Mag.*, vol. 41, no. 3, pp. 108–113, Mar. 2003.

[60] X. Wang, P. Zhu, and M. Chen, "Antenna location design for generalized distributed antenna systems," *IEEE Commun. Lett.*, vol. 13, no. 5, pp. 315–317, May 2009.

[61] J. Guo, E. Koyuncu, and H. Jafarkhani, "Energy efficiency in two-tiered wireless sensor networks," in *Proc. IEEE International Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[62] A. Yang, Y. Jing, C. Xing, Z. Fei, and J. Kuang, "Performance analysis and location optimization for massive MIMO systems with circularly distributed antennas," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5659–5671, Oct. 2015.

[63] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[64] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. New York, NY, USA: Springer Science & Business Media, 2012, vol. 159.

[65] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[66] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding," *Proc. IEEE*, vol. 73, no. 11, pp. 1551–1588, Nov. 1985.

[67] C. H. Aldana, E. de Carvalho, and J. M. Cioffi, "Channel estimation for multicarrier multiple input single output systems using the EM algorithm," *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3280–3292, Dec. 2003.

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

[69] E. de Carvalho and D. T. M. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *Proc. IEEE Signal Process. Adv. Wireless Commun. Workshop*, Apr. 1997, pp. 129–132.

[70] E. de Carvalho and D. T. M. Slock, "Asymptotic performance of ML methods for semi-blind channel estimation," in *Proc. 31st Asilomar Conf. Signals, Syst., Comput.*, vol. 2, 1997, pp. 1624–1628.

[71] A. K. Jagannatham and B. D. Rao, "Whitening-rotation-based semi-blind MIMO channel estimation," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 861–869, Mar. 2006.

[72] M. Abuthinien, S. Chen, and L. Hanzo, "Semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems," *IEEE Signal Process. Lett.*, vol. 15, pp. 202–205, Nov. 2008.

[73] J. Choi, *Adaptive and iterative signal processing in communications*. Cambridge, U.K.: Cambridge University Press, 2006.

[74] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.

[75] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.

[76] U. Sandkühler and J. J. F. Bohme, "Accuracy of maximum-likelihood estimates for array processing," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, vol. 12, Apr. 1987, pp. 2015–2018.

[77] P. Zetterberg, "Experimental investigation of TDD reciprocity-based zero-forcing transmit precoding," *EURASIP Journal on Advances in Signal Processing*, Jan. 2011.

[78] S. D. Silverstein, "Application of orthogonal codes to the calibration of active phased array antennas for communication satellites," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 206–218, Jan. 1997.

[79] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B*, vol. 39, no. 1, pp. 1–38, 1977.

[80] H. Hung and M. Kaveh, "On the statistical sufficiency of the coherently averaged covariance matrix for the estimation of the parameters of wideband sources," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 12, 1987, pp. 33–36.

[81] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[82] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review*, vol. 106, no. 4, pp. 620 – 630, 1957.

[83] M. Kobayashi, J. Boutros, and G. Caire, "Successive interference cancellation with SISO decoding and EM channel estimation," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1450–1460, Aug. 2001.

[84] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," in *Proc. Inst. Elect. Eng.*, vol. 130, no. 1. IET, Feb. 1983, pp. 11–16.

[85] R. G. Gallager, *Stochastic processes: theory for applications*. Cambridge University Press, 2013.

[86] P. Billingsley, *Probability and Measure*, 3rd ed. John Wiley & Sons Inc., 1995.