Summer 7-1-2022

Analysis of Wireless Networks With Massive Connectivity

Mangqing Guo
Syracuse University

Abstract

Recent years have witnessed unprecedented growth in wireless networks in terms of both data traffic and number of connected devices. How to support this fast increasing demand for high data traffic and connectivity is a key consideration in the design of future wireless communication systems. With this motivation, in this thesis, we focus on the analysis of wireless networks with massive connectivity.

In the first part of the thesis, we seek to improve the energy efficiency (EE) of single-cell massive multiple-input multiple-output (MIMO) networks with joint antenna selection and user scheduling. We propose a two-step iterative procedure to maximize the EE. In each iteration, bisection search and random selection are used first to determine a subset of antennas with the users selected before, and then identify the EE-optimal subset of users with the selected antennas via cross entropy algorithm. Subsequently, we focus on the joint uplink and downlink EE maximization, under a limitation on the number of available radio frequency (RF) chains. With the Jensen's inequality and the power consumption model, the original problem is converted into a combinatorial optimization problem. Utilizing the learning-based stochastic gradient descent framework and the rare event simulation method, we propose an efficient learning-based stochastic gradient descent algorithm to solve the corresponding combinatorial optimization problem.

In the second part of the thesis, we focus on the joint activity detection and channel estimation in cell-free massive MIMO systems with massive connectivity. At first, we conduct an asymptotic analysis of single measurement vector (SMV) based minimum mean square error (MMSE) estimation in cell-free massive MIMO systems with massive connectivity. We establish a decoupling principle of SMV based MMSE estimation for sparse signal vectors with independent and non-identically distributed (i.n.i.d.) non-zero components. Subsequently,

using the decoupling principle, likelihood ratio test and the optimal fusion rule, we obtain detection rules for the activity of users based on the received pilot signals at only one access point (AP), and also based on the cooperation of the received pilot signals from the entire set of APs for centralized and distributed detection. Moreover, we study the achievable uplink rates with zero-forcing (ZF) detector at the central processing unit (CPU) of the cell-free massive MIMO systems.

In the third part, we focus on the performance analysis of intelligent reflecting surface (IRS) assisted wireless networks. Initially, we investigate the MMSE channel estimation for IRS assisted wireless communication systems. Then, we study the sparse activity detection problem in IRS assisted wireless networks. Specifically, employing the generalized approximate message passing (GAMP) algorithm, we obtain the MMSE estimates of the equivalent effective channel coefficients from the base station (BS) to all users, and transform the received pilot signals into additive Gaussian noise corrupted versions of the equivalent effective channel coefficients. Likelihood ratio test is used to acquire decisions on the activity of each user based on the Gaussian noise corrupted equivalent effective channel coefficients, and the optimal fusion rule is used to obtain the final decisions on the activity of all users based on the previous decisions on the activity of each user and the corresponding reliabilities. Finally, we conduct an asymptotic analysis of maximizing the weighted sum rate by joint beamforming and power allocation under transmit power and quality-of-service (QoS) constraints in IRS assisted wireless networks.

ANALYSIS OF WIRELESS NETWORKS WITH MASSIVE CONNECTIVITY

By

Mangqing Guo

B.S., Jilin University, 2013

M.S., Beijing University of Posts and Telecommunications, 2016

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical and Computer Engineering

Syracuse University

July 2022

Acknowledgements

First of all, I would like to sincerely thank my advisor, professor Mustafa Cenk Gursoy. With his guide and help, I was able to successfully complete my Ph.D. study. During this period, I gained a lot of valuable scientific research experience, which will be a valuable asset in my life and guide my future work and study.

Secondly, I would like to thank all my lab mates, Xueyuan Wang, Chen Zhong, Yang Yang, Ziyang Lu and Mohamad Hani Sulieman. Thank you for their tolerance and help, making my doctoral life more colorful.

Finally, I would like to thank my family. I am indebted to them for their continued support in nurturing me into an adult and encouraging me to complete my studies. I will always remember their kindness in my heart, and repay it the rest of my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Wireless Networks With Massive Connectivity

Wireless networks have experienced significant growth recently in terms of both data traffic and the number of connected devices. Supporting such fast increasing demand for data traffic and connectivity is one of the core tasks in future wireless networks, and several advanced technologies, including massive multiple-input multiple-output (MIMO), millimeter wave transmissions, full-duplex operation, and device-to-device (D2D) communication, have been proposed and analyzed intensively in recent years. In particular, massive MIMO systems have large number of antennas at the base station (BS). Thanks to such large number of antennas, space multiplicity and diversity can be fully used. Indeed, the channels of massive MIMO networks are orthogonal to each other, which can lead to significant growth in the data rates per cell by hundreds of times [2, 3]. By utilizing large number of antennas at the BS, massive MIMO network could support large number of mobile devices simultaneously. With such benefits, massive MIMO has become one of the key technologies for 5G cellular systems, and it has been intensively studied in recent years [4, 5].

Cell-free massive MIMO is a potentially key technology for next generation wireless systems, and it has also been intensively studied in recent years. In these systems, a large

number of access points (APs) are spatially distributed in the coverage area, and all the users are jointly and simultaneously served by all of the APs. Compared to the traditional cellular massive MIMO systems, the APs of cell-free massive MIMO networks could be placed at anywhere among a large area, which could achieve larger energy efficiency (EE) and more cost-efficient deployment [6]. Moreover, since there are no cell-edge users, cell-free massive MIMO system could achieve more uniform quality-of-service (QoS) among all users [7].

Internet-of-Things (IoT) is another potentially key technology for future wireless communications. There will be 29.3 billion networked devices by 2023, and IoT connections will be half of the global connected devices and connections [8]. Massive connectivity and sporadic activity are two main characteristics of the future IoT networks [9, 10]. Large number of devices (such as vehicles, sensors, smart furniture, etc.) are potential nodes of the IoT network, while only a small fraction of the nodes actively connect to the network at a given time instant. Active users initially send pilots and then connect to the network to exchange information within the network, and then disconnect from the network after information exchange is completed. In this way, throughput of the network can be significantly improved, and many more users could be served by the network.

Channel condition is one of the decisive factors affecting the performance of wireless networks. However, most of the wireless network technologies can only control the transmitter and receiver, not the environment in between [11]. By introducing large number of passive scattering elements, intelligent reflecting surface (IRS) makes it possible to tune the wireless environments between transmitter and receiver to promote the spectral and energy efficiency of wireless systems [12–15].

1.2 Literature Overview

1.2.1 Antenna Selection and User Scheduling in Massive MIMO Networks

As also noted above, massive MIMO represents a system model in which the BS equipped with a large number of antennas simultaneously serves a large number of users. Owing to the high degrees of freedom available for each user, massive MIMO can increase the capacity ten times or more while improving the EE on the order of one hundred times [16]. Due to such benefits, there has been significant interest in massive MIMO in recent years, and it is considered as one of the key technologies for 5G [17].

The large number of antennas and radio frequency (RF) chains needed at the BS in massive MIMO systems brings forth challenges regarding the system complexity and hardware energy consumption. Therefore, how to achieve the maximum EE is one of the core tasks in the analysis and design of massive MIMO networks. For instance, if uplink and downlink are considered together, improvements in EE could be achieved [1]. Also, in order to lower the hardware cost and complexity, the number of RF chains available at the BS can be reduced to be less than the number of antennas. In such cases, how to improve the EE of massive MIMO systems under limitations on the number of RF chains becomes a critical consideration. In addition to just addressing EE, one may also need to consider the tradeoff between the EE and spectral efficiency (SE) in order to achieve the best performance in the system while ensuring a target data transmission level.

While there are various approaches to improve the EE of massive MIMO systems, antenna selection and user scheduling have been commonly considered and studied in this respect. For instance, multi-objective optimization [18], principal component analysis [19], successive removal [20], $L_{1/2}$-regularity based methods [21], norm-and-correlation-based selection algorithms [22], average absolute value of the channel coefficients based methods [23] and convex optimization [24] have been employed in antenna selection to maximize the downlink EE or

SE of massive MIMO systems. We note that the number of available RF chains at the BS is limited in [21] and [22]. Both uplink and downlink are considered and an antenna selection method that searches for the channels with the strongest absolute channel coefficients is applied to maximize the EE of single-cell massive MIMO systems in [25]. Reference [26] shows that simple random antenna selection can lead to significant EE gains. Moreover, when the EE-optimal number of antennas at the BS is larger than a certain threshold, then the performance of random antenna selection is already very close to that of the optimum antenna selection. Convex relaxation and greedy approach have been used for antenna selection in [27] to maximize the sum rate of an uplink single-cell massive MIMO system.

K^*-random user selection, K^*-location-dependent user selection [28], greedy user selection with linear precoding scheme [29], and joint antenna selection and user scheduling method [30] have been used in downlink massive MIMO systems to maximize the sum rate. Greedy two-step joint antenna selection and user scheduling have been used in [31] to maximize the sum rate of uplink massive MIMO systems, while [32] has the objective to maximize the downlink sum rate. Norm-based, greedy-based and TCB (throughput and complexity balanced) based sub-optimal iterative joint antenna selection and user scheduling algorithms have been proposed to improve the downlink channel capacity of distributed massive MIMO systems in [33]. Semidefinite programming (SDP) approach is used in [34] for state estimation in smart grids. Besides, the authors in [1] work on the EE maximization problem with joint antenna selection and user scheduling. By utilizing the power allocation technique from [35], each user could achieve the same rate. Then, the antenna selection and user scheduling problem becomes finding the optimal number of antennas and users which could maximize the EE of the system.

1.2.2 Sparse Activity Detection and Channel Estimation in Cell-Free Massive MIMO Systems With Massive Connectivity

IoT is one of the promising technologies for next generation wireless networks, and massive connectivity and sporadic activity are two key characteristics of future IoT networks [10]. Specifically, large amount of devices are served by such network. However, only a small partial of the devices are active at a given time instant, while the remaining devices keep inactive. For instance, a device is inactive most of the time, turns to active only when it needs to exchange messages with the network, and converts to inactive once the information exchange is completed. Therefore, at a given time instant, only a small unknown subset out of the entire set of users is active, while most of the users are inactive. Then, identifying the active users is a key requirement in IoT networks with massive connectivity, and this is also referred to as the activity detection problem. In addition to activity detection, it is also critical to estimate the channel coefficients between the active users and APs.

In the literature, IoT networks with massive connectivity have been intensively studied. For instance, massive connectivity for industrial IoT networks has been addressed in [36], where the signal-to-interference-plus-noise ratio (SINR) and SE under the maximum ratio (MR) or zero-forcing (ZF) decoding at the receiver are investigated for the uplink network. Massive connectivity for IoT in distributed massive MIMO networks is analyzed in [37], where the activity detection problem is addressed via the alternating direction method of multipliers (ADMM) algorithm in light-load traffic scenarios and via the dynamic clustering method in heavy-load traffic scenarios. The orthogonal matching pursuit (OMP) algorithm is used in [38] to deal with the activity detection problem, and the authors in [39] have proposed a compressed sensing Neyman-Pearson based activity detection algorithm for IoT networks with massive connectivity. Activity detection using the maximum likelihood method is investigated in [40].

Moreover, several recent studies have analyzed the joint activity detection and effective channel estimation (JADECE) in IoT networks with massive connectivity by employing the

approximate message passing (AMP) algorithm. AMP algorithm can obtain the minimum mean square error (MMSE) estimate of sparse signals with tolerable complexity, and it breaks down the received signal into Gaussian noise corrupted scalar versions. Besides, the variance of the Gaussian noise in the decoupled signal can be described with the state evolution equations, facilitating the performance analysis of the AMP algorithm. With these benefits, AMP algorithm has been widely used in solving the JADECE problem in IoT networks with massive connectivity. For instance, the AMP algorithm is used in [9, 10, 41] to solve the JADECE problem in single-cell massive MIMO networks with massive connectivity, and multi-cell massive MIMO cooperative networks are analyzed with the AMP algorithm in [42]. Authors in [43–45] focus on solving the JADECE problem via the distributed AMP algorithm. The authors in [46] have proposed a structured group sparsity estimation approach to solve the JADECE problem in IoT networks with massive connectivity. Furthermore, sparse matrix factorization and AMP algorithms are used in [47] and [48] to solve the JADECE problem in IRS assisted massive MIMO networks with massive connectivity. In addition to the AMP algorithm, sparsity learning based methods are used in [49] to deal with the JADECE problem in grant-free massive-device multiple access system. The authors in [50] employ a compressed sensing algorithm to solve the JADECE problem for massive connectivity with 1-bit digital-to-analog converter. A turbo receiver based on the bilinear generalized approximate message passing algorithm is proposed to solve the JADECE problem for grant-free massive random access in [51].

Although AMP algorithm is efficient in solving the JADECE problem, it is sensitive to the mean and condition number of the measurement matrix, which impairs the convergence property of the algorithm [52–54]. In order to overcome this, various modified AMP algorithms have been proposed, including mean removal AMP [54], swept AMP [55], generalized approximate message passing (GAMP) [56] and complex Bayesian approximate message passing (CB-AMP) algorithm [57]. Damping and mean removal techniques are employed in [53, 58] to further promote the convergence of the GAMP algorithm.

However, the state evolution equations of the above algorithms depend on the asymptotic mean square error of the estimated signal. Therefore, in order to theoretically analyze the performance of these algorithms, one needs to identify the corresponding asymptotic mean square error first. Since we consider MMSE estimation in chapter 4 and 5, how to find the asymptotic mean square error of the MMSE estimate of the sparse signal becomes a core problem.

Replica method from statistical physics has been widely used to find the asymptotic mean square error of the AMP algorithm [59, 60], and it is shown to work well empirically. With the replica method, the authors in [61] have provided a decoupling principle for the asymptotic analysis of MMSE estimation in large system limit, and the replica analysis for MMSE estimation of multimeasurement vector (MMV) problem is performed in [62]. Moreover, the asymptotic behavior of MMSE estimation for sparse signals with random matrix theory and large deviations is analyzed in [63, 64].

Cell-free massive MIMO is another promising technology for future wireless communications, in which a large number of APs are distributed in the coverage area, and all the APs are connected with the fusion center (FC). All the users are jointly and simultaneously served by all the APs. The channel hardening and favorable propagation properties of cell-free massive MIMO networks are studied in [65]. The achievable rates of cell-free massive MIMO systems are investigated in [7, 66–69]. Power control in cell-free massive MIMO networks is studied in [70–72].Spectral and energy efficiency of cell-free massive MIMO networks are analyzed in [73–77]. Pilot assignment and power control in cell-free massive MIMO systems are analyzed in [69].

In chapter 4 and 5, we focus on the asymptotic analysis of single measurement vector (SMV) based MMSE estimation in cell-free massive MIMO systems with massive connectivity. We address the JADECE problem using the frameworks of MMSE estimation and likelihood ratio test. Prior studies on the MMSE estimation of sparse signals have developed methodologies, including a closed-form method in [78], an approximation method in [79],

and AMP and CB-AMP algorithms. Since we consider cell-free massive MIMO here, the non-zero components of the sparse signal vectors to be estimated are independent and non-identically distributed (i.n.i.d.). However, to the best of our knowledge, all the asymptotic analyses of the MMSE estimation with replica method in the literature focus on the sparse signal vectors with independent and identically distributed (i.i.d.) non-zero components. With i.n.i.d. non-zero components, analysis becomes more challenging. Although the authors in [78] have obtained closed-form expressions for the MMSE estimates of the sparse signals with heteroscedastic non-zero entries, there is a requirement for the measurement matrix to be unitary, while we in our setting have an underdetermined system as a result of the facts that limited number of pilots can be used, and condition of having a unitary measurement matrix cannot be satisfied. An approximation method for the MMSE estimation of sparse signals is provided in [79]. However, when the number of variables becomes large, the computational complexity will grow quickly and the accumulated error will be high.

1.2.3 Intelligent Reflecting Surface Assisted Wireless Networks

Wireless propagation environment is one of the key factors affecting the performance of wireless networks. Typically, the wireless network technologies can only control the transmitter and receiver, not the environment in between [11]. By introducing large amount of passive scattering elements, IRSs make it possible to tune the wireless environments between transmitter and receiver to promote higher spectral and energy efficiency in wireless communication systems [12–15].

Recently, various aspects of IRS assisted wireless networks have been studied, e.g., channel capacity, beamforming optimization, power allocation, channel estimation, etc. In [80], the capacity limit of IRS-aided point-to-point MIMO communication system is considered by jointly optimizing the IRS reflection coefficients and the MIMO transmit covariance matrix. Coverage probability and ergodic capacity of IRS assisted single-input single-output (SISO) system is considered in [81]. Sum rate maximization for IRS based multi-user communication

is achieved in [82] with hybrid beamforming, and joint design of the reflecting coefficients and transmit beamforming to maximize the users' worst rate is studied in [83]. The total transmit power of IRS aided wireless network is minimized by jointly design the transmit beamforming and the passive beamforming at IRS, subject to the users' individual SINR constraints in [84–86]. The energy efficiency of IRS enhanced wireless network is considered by jointly optimize the transmit power and the phase shifts of the scattering elements in [87]. The rate and energy efficiency of wireless communication systems with classic decode-and-forward (DF) relaying or IRS are compared in [11]. In [88], the authors investigated the performance of energy harvesting for IRS enhanced simultaneous wireless information and power transfer (SWIPT) aided systems. Latency minimization for IRS Aided mobile edge computing is considered in [89], and IRS carried by unmanned aerial vehicles (UAVs) to support cellular communication networks is studied in [90]. A stochastic geometry analysis is performed in [91] for large intelligent surface (LIS) assisted millimeter wave networks, and it is demonstrated that the LISs dramatically improve the average rate and area spectral efficiency of millimeter wave networks when the BS density is lower than the LIS density. In [92], weighted sum-rate optimization has been performed by jointly optimizing the active beamforming at the BS and the passive beamforming at the IRS in IRS enhanced wireless networks.

Channel estimation plays a key role in the performance enhancement of IRS assisted wireless networks, and certain challenges exist in estimating the channel since the passive reflecting elements do not have active transmitting/receiving or signal processing abilities. For instance, the existing channel estimation methods for IRS assisted wireless networks are typically costly in terms of either pilot signal usage or computational complexity. The channel estimation process in [93, 94] is divided into several sub-frames, and the number of symbols transmitted in each sub-frame is greater than or equal to the number of users. The channel coefficients of different scattering element assisted links are serially estimated by the control loop of a micro-controller connected to the IRS in [95]. A two-timescale channel

estimation scheme for IRS assisted wireless networks is proposed in [96] by considering the properties that the BS-IRS channel is high dimensional but quasi-static, while the IRS-user channel is mobile but low-dimensional. Minimum variance unbiased estimator is utilized for channel estimation in IRS aided networks in [97]. The channel estimation in an IRS assisted millimeter wave MIMO system is converted into a fixed-rank constrained non-convex optimization problem in [98]. In [94] and [99], the cascade channel, which consists the channels from the BS to the IRS and from the IRS to the users, is estimated using compressed sensing techniques. Sparse matrix factorization technique is employed in [100] and [101] for channel estimation in IRS enhanced wireless communications. The authors in [47] perform joint activity detection and channel estimation for IRS assisted wireless networks with massive connectivity in three stages, namely, by employing sparse matrix factorization, matrix completion and MMV method.

Moreover, channel capacity could be improved dramatically by improving the number of antennas at the BS or the number of users, i.e., using massive MIMO technology. Asymptotic analysis is one of the key tools for the performance analysis of massive MIMO networks. Several recent studies in the literature have focused on the asymptotic analysis of large IRS aided wireless networks. The rate distribution and outage probability of large IRS assisted wireless networks are asymptotically analyzed in [102]. Asymptotic analysis of the uplink data rate in an IRS-based large antenna array system is performed in [103]. Deterministic equivalents of the SINR and the sum rate of IRS assisted MIMO downlink systems under imperfect CSI are analyzed in [104].

Furthermore, maximizing the weighted sum rate is another key approach to improve the channel capacity of wireless networks. Successive zero-forcing dirty paper coding is designed for maximizing the weighted sum rate of a single cell MIMO network in [105]. The weighted sum rate of all users are maximized by jointly designing the beamforming at APs and phase vector of IRS elements for IRS aided wireless networks in [106]. Large system analysis of weighted sum rate maximization of wireless networks are performed in [107, 108]. Motivated

by the relationship between mutual information and MMSE, several studies address the weighted sum rate maximization problem using its relationship with the weighted MMSE approach [109, 110].

In addition, max-min weighted SINR is an effective way to explore the achievable SINR region to facilitate the study of weighted sum rate and weighted MMSE in wireless networks [111, 112]. Max-min weighted SINR under a weighted sum transmit power constraint in MIMO downlink systems is analyzed in [113]. Asymptotic analysis of max-min SINR subject to a given transmit power constraint in a large-scale single-cell multi-user system is performed in [114]. Asymptotic analysis of max-min SINR under a given transmit power constraint in IRS assisted single-cell multi-user networks is conducted in [115].

1.3 Outline of the Thesis

In this thesis, we focus on the performance analysis of wireless networks with massive connectivity. First, higher EE is achieved by joint antenna selection and user scheduling in massive MIMO systems. Then, we study the joint activity detection and channel estimation problem in cell-free massive MIMO systems with massive connectivity. Finally, different types of performance analysis on IRS assisted wireless networks are performed, including channel estimation, asymptotic analysis and joint activity detection and channel estimation.

An energy-efficient joint antenna and user selection algorithm in single-cell massive MIMO communication systems is proposed in Chapter 2. The proposed algorithm involves a two-step iterative procedure. At each time, we first obtain a subset of antennas for the given set of users via bisection search and random selection, and then obtain the optimally energy efficient subset of users with the selected antennas using cross-entropy algorithm. This two-step procedure is demonstrated to improve the EE at each iteration. Simulation results show that the EE could be improved by 71.16% with the maximum-ratio combining (MRC) receiver when the total number of users is 60.

In Chapter 3, we propose a learning-based stochastic gradient descent algorithm to obtain the optimal joint uplink and downlink EE with joint antenna selection and user scheduling in single-cell massive MIMO systems, under a limitation on the number of available RF chains. Using Jensen's inequality and the characteristics of wireless channels, a lower bound on the system throughput is obtained. Subsequently, incorporating the power consumption model, the corresponding lower bound on the EE of the system is identified. Finally, learning-based stochastic gradient descent method is used to solve the joint antenna selection and user scheduling problem, which is a combinatorial optimization problem. Rare event simulation is embedded in the learning-based stochastic gradient descent method to generate samples with very small probabilities. In the analysis, both perfect and imperfect channel state information (CSI) at the BS are considered. MMSE channel estimation is employed in the study of the imperfect CSI case. In addition, the effect of a constraint on the number of available RF chains in massive MIMO system is investigated considering both perfect and imperfect CSI at the BS.

In Chapter 4, we investigate the MMSE estimation of effective channel coefficients in cell-free massive MIMO systems with massive connectivity. To facilitate the theoretical analysis, only SMV based MMSE estimation is considered, i.e., the MMSE estimation is performed based on the received pilot signals at each AP separately. Inspired by the decoupling principle of replica symmetric postulated MMSE estimation of sparse signal vectors with i.i.d. non-zero components, we develop the corresponding decoupling principle for the SMV based MMSE estimation of sparse signal vectors with i.n.i.d. non-zero components, which plays a key role in the theoretical analysis of SMV based MMSE estimation of the effective channel coefficients in cell-free massive MIMO systems with massive connectivity. Subsequently, based on the obtained decoupling principle of MMSE estimation, likelihood ratio test and the optimal fusion rule, we perform user activity detection based on the received pilot signals at only one AP, or cooperation among the entire set of APs for centralized or distributed detection. Via theoretical analysis, we show that the error probabilities of both centralized

and distributed detection tend to zero when the number of APs tends to infinity while the asymptotic ratio between the number of users and pilots is kept constant. We also investigate the asymptotic behavior of oracle estimation in cell-free massive MIMO systems with massive connectivity via random matrix theory. Moreover, in order to demonstrate the potential performance loss of SMV based MMSE estimation, which does not employ the correlation between the received pilot signals at different APs, the MMV based MMSE estimation, i.e., joint MMSE estimation with pilot signals from all APs, is analyzed via numerical results. Numerical analysis shows that the theoretical analyze with our decoupling principle for the SMV based MMSE estimation of sparse signal vectors with i.n.i.d. non-zero components matches well with the numerical results.

In Chapter 5, we focus on the achievable rate of cell-free massive MIMO networks with massive connectivity. With the decoupling properties of MMSE estimation for large linear systems [61], the received pilot signals are decomposed into scalar additive complex Gaussian noise corrupted versions. We determine the variance of the scalar additive complex Gaussian noise with the help of the state evolution equations of the GAMP algorithm, and then identify the variances of both the estimated channel and the corresponding channel estimation error. Subsequently, with the results in [76], we analyze the achievable uplink data transmission rates in cell-free massive MIMO systems with massive connectivity when ZF detector is employed. Moreover, we demonstrate the impact of the accuracy of activity detection on the achievable rates via numerical results.

In Chapter 6, we investigate the MMSE channel estimation of IRS assisted wireless networks. By considering the equivalent channel from the BS to the users, the MMSE channel estimation process can be achieved in one stage via transmitting orthogonal pilots from the users. With the help of the Gaussian scale mixture (GSM) model, we obtain the MMSE estimate of the equivalent channel, and identify analytical upper and lower bounds on the mean square error. Applying the central limit theorem, we perform an asymptotic analysis of the channel estimation, through which we show that the upper bound on the

mean square error of the MMSE estimation equals the asymptotic mean square error of the MMSE estimation when the number of reflecting elements at the IRS tends to infinity.

In Chapter 7, we consider the sparse activity detection in IRS assisted wireless networks. The equivalent channel coefficients from users to the BS are approximated with a Gaussian approximation (GA) model. With GAMP algorithm, the received pilot signals at the BS are decoupled into scalar Gaussian noise corrupted versions of the effective channel coefficients. Subsequently, MMSE estimate of the effective channel coefficients and threshold detection rules are acquired. Finally, the optimal fusion rule is used to obtain the activity detection results of each user. Numerical results show that the average error probability of the sparse activity detection method proposed in this chapter diminishes as the SNR, number of pilots, number of antennas at the BS or number of elements at the IRS increases.

In Chapter 8, we investigate the asymptotic analysis of maximizing weighted sum rate under transmit power and QoS constraints in IRS assisted wireless networks. With the help of the asymptotic results of the max-min SINR problem, the original problems are converted into the optimization over SINR. After a transformation of the optimization variables, the weighted sum rate maximization problem is converted into a concave-convex optimization problem and it can be solved with the iterative approach of the concave-convex fractional programming algorithm proposed in [116]. Besides, we show that the optimal power allocation scheme always makes full use of the available transmit power.

1.4 Bibliographic Note

- The results in Chapter 2 appeared in the conference paper:

  M. Guo and M. C. Gursoy, "Energy-efficient joint antenna and user selection in single-cell massive MIMO systems," in IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2018, pp. 838–842.

- The results in Chapter 3 appeared in the journal paper:

M. Guo and M. C. Gursoy, "Statistical learning based joint antenna selection and user scheduling for single-cell massive MIMO systems," IEEE Transactions on Green Communications and Networking, vol. 5, no. 1, pp. 471–483, 2021.

- The results in Chapter 4 appeared in the journal paper:

  M. Guo and M. C. Gursoy, "Joint activity detection and channel estimation in cell-free massive MIMO networks with massive connectivity," IEEE Transactions on Communications, vol. 70, no. 1, pp. 317–331, 2022.

  and in the conference papers:

  M. Guo and M. C. Gursoy, "Distributed sparse activity detection in cell-free massive MIMO systems," in IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2019, pp. 1–5.

  M. Guo, M. C. Gursoy, and P. K. Varshney, "Sparse activity detection in cell-free massive MIMO systems," in IEEE International Symposium on Information Theory (ISIT), 2020, pp. 1177–1182.

- The results in Chapter 5 appeared in the conference paper:

  M. Guo and M. C. Gursoy, "Performance analysis of cell-free massive MIMO systems with massive connectivity," in IEEE 18th Annual Consumer Communications Networking Conference (CCNC), 2021, pp. 1–6.

- The results in Chapter 6 appeared in the conference paper:

  M. Guo and M. C. Gursoy, "Channel estimation for intelligent reflecting surface assisted wireless communications," in IEEE Wireless Communications and Networking Conference (WCNC), 2021, pp. 1–6.

- The results in Chapter 7 appeared in the conference paper:

  M. Guo and M. C. Gursoy, "Sparse activity detection in intelligent reflecting surface as-

sisted wireless networks," in IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2021, pp. 543–548.

Chapter 2

Energy-Efficient Joint Antenna and User Selection in Single-Cell Massive MIMO Systems

In this chapter, we propose a two-step iterative procedure to maximize the EE of single-cell massive MIMO communication systems. At each time, we first use bisection search and random selection to determine a subset of antennas with the users selected before, and then identify the EE-optimal subset of users with the selected antennas via cross entropy algorithm. Theoretical analysis shows that this two-step procedure improves the EE at each iteration. The performance is evaluated through numerical results.

2.1 System Model

We consider the uplink of a single-cell massive MIMO system, which consists of a BS with M antennas, and K single-antenna users. We assume that the system operates over flat Rayleigh fading channels, the transmission bandwidth is B Hz, and the BS has perfect CSI. The received signal at the BS is

$$\mathbf{y} = \sqrt{p_u}\mathbf{G}\mathbf{x} + \mathbf{n} \tag{2.1}$$

where p_u is the user transmit power, $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_K] \in \mathcal{C}^{M \times K}$ is the channel matrix from all users to the BS, and $\mathbf{n}$ is i.i.d. Gaussian noise vector with zero mean and covariance matrix $\sigma^2\mathbf{I}$. $\mathbf{x}$ is the signal vector whose components are the transmitted signals from the users, and satisfies $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}_K$. $\mathbf{g}_k$ is the uplink channel gain from the kth user to all the antennas at the BS. For flat fading channels, $\mathbf{g}_k$ can be written as $\mathbf{g}_k = \beta_k \mathbf{h}_k$, in which the fast fading coefficients $\mathbf{h}_k$ have i.i.d. complex Gaussian $\mathcal{CN}(0, 1)$ elements. β_k describes the path loss and shadow fading, and can be expressed as, $\beta_k = \frac{z_k}{r_k^\alpha}$, where r_k is the distance between the kth user and the BS, α is the path loss decay exponent, and the shadow fading coefficient z_k is a log-normal random variable, i.e., the quantity $10\log_{10}(z_k)$ is zero-mean Gaussian distributed with standard deviation σ_{shad} [2].

We use linear combination at the BS. Then, the received signal, after linear combination, can be expressed as

$$\hat{\mathbf{y}} = \sqrt{p_u}\mathbf{A}\mathbf{G}\mathbf{x} + \mathbf{A}\mathbf{n} \tag{2.2}$$

where $\mathbf{A} = \mathbf{G}^H$ is the MRC matrix at the BS. Then, the kth user's received signal at the BS is

$$\hat{\mathbf{y}}_k = \sqrt{p_u}\mathbf{a}_k\mathbf{g}_k x_k + \sqrt{p_u}\sum_{i=1, i\neq k}^{K} \mathbf{a}_k\mathbf{g}_i x_i + \mathbf{a}_k\mathbf{n} \tag{2.3}$$

where $\mathbf{a}_k$ is the kth row of $\mathbf{A}$. Therefore, the received SINR of the kth user's signal at the BS is

$$\gamma_k = \frac{p_u \left|\mathbf{a}_k\mathbf{g}_k\right|^2}{p_u \sum\limits_{i=1, i\neq k}^{K} \left|\mathbf{a}_k\mathbf{g}_i\right|^2 + \left\|\mathbf{a}_k\right\|^2 \sigma^2}. \tag{2.4}$$

Then, the total achievable uplink rate in bits per second (bps) is

$$R = B\sum\nolimits_{k=1}^{K} E\{\log_2(1 + \gamma_k)\}. \tag{2.5}$$

By Jensen's inequality and the properties of the Wishart matrix, we can obtain the following

lower bound on R [117]:

$$\widetilde{R} = B \sum\nolimits_{k=1}^{K} \log_2 \left(1 + \frac{p_u (M-1) \beta_k}{p_u \sum_{i=1, i \neq k}^{K} \beta_i + \sigma^2} \right). \tag{2.6}$$

2.2 Energy Efficiency

2.2.1 Power Consumption

We use the power consumption model proposed in [26]. For completeness, we provide a brief discussion of the model here. The total power consumption can be expressed as

$$P_{sum} = K P_{PA} + M P_C \tag{2.7}$$

where $P_{PA} = \frac{p_u}{\eta_e}$ is the power consumed by the power amplifier at a user, $P_C = \frac{P_{BB}}{M} + P_{RF}$ is the circuit power per RF chain at the BS. η_e is the power amplifier efficiency at the user terminal, $P_{BB} = \frac{\chi_p}{\varrho_e}$ is the baseband power consumption at the BS, and P_{RF} is the BS RF front-end power consumption which includes the power consumptions of the mixer, filter and digital-to-analog converter power consumption. ϱ_e is the VLSI processing efficiency, and

$$\chi_p = MB \left[\left(\frac{T_u}{T_s} \right) \log_2 (T_u B) + \left(\frac{T_u}{T_s} \right) \left(1 - \frac{T_p}{T_{sl}} \right) K + \left(\frac{T_u T_p}{T_s T_{sl}} \right) \log_2 \left(\frac{T_u T_p}{T_s T_d} \right) + \left(\frac{T_d}{T_{sl}} \right) K^2 \right]. \tag{2.8}$$

The description of each parameter in (2.8) is shown in Table 2.1.

Table 2.1: Baseband power consumption parameters in (2.8)

Parameters	**Description**	**Value**
T_{sl}	slot length	$0.5ms$
T_p	Pilot length in one slot	$0.214ms$
T_s	Symbol duration	$71.4us$
T_g	Guard Interval (GI)	$4.7us$
T_u	Symbol without GI	$66.7us$
T_d	Delay spread	$4.7us$

2.2.2 Joint Antenna and User Selection with MRC Receiver

A lower bound on the EE of a single-cell massive MIMO communication system with an MRC receiver is given by

$$EE(\mathcal{M},\mathcal{K}) = \frac{B}{P_{sum}} \sum_{i=1}^{|\mathcal{K}|} \log_2 \left(1 + \frac{p_u \left(|\mathcal{M}| - 1 \right) \beta_k}{p_u \sum\limits_{i=1, i \neq k}^{|\mathcal{K}|} \beta_i + \sigma^2} \right). \tag{2.9}$$

The goal is to maximize the EE and hence solve the following optimization problem:

$$\underset{\mathcal{M},\mathcal{K}}{\text{maximize}} \; EE(\mathcal{M},\mathcal{K}) \tag{2.10}$$

where $\mathcal{M}$, $\mathcal{K}$ are the subsets of selected antennas and users, respectively. We address this problem in two steps. First, we get a suboptimum subset of antennas $\mathcal{M}^*$ considering the entire set of users. Then, we determine the EE-optimal subset of users with $\mathcal{M}^*$.

As the number of antennas at the BS is very large in massive MIMO systems, it is difficult to determine the EE-optimal subset of antennas. Because random antenna selection can achieve significant EE gains in massive MIMO systems [26], we first obtain the EE optimal number of antennas M^*, and then randomly select a subset of antennas $\mathcal{M}^*$, i.e., $|\mathcal{M}^*| = M^*$. The derivative of $EE(\mathcal{M},\mathcal{K})$ with respect to the number of antennas can be

expressed as[1]

$$\frac{\partial}{\partial |\mathcal{M}|}(EE(\mathcal{M},\mathcal{K})) = \frac{B\log_2 e}{(|\mathcal{K}|\, P_{PA} + |\mathcal{M}|\, P_c)^2} \left(\sum_{k=1}^{|\mathcal{K}|} \frac{p_u \beta_k (|\mathcal{K}|\, P_{PA} + |\mathcal{M}|\, P_c)}{p_u(|\mathcal{M}|-1)\beta_k + p_u \sum\limits_{i=1, i\neq k}^{|\mathcal{K}|} \beta_i + \sigma^2} \right.$$
$$\left. - P_c \ln \left(1 + \frac{p_u(|\mathcal{M}|-1)\beta_k}{p_u \sum\limits_{i=1, i\neq k}^{|\mathcal{K}|} \beta_i + \sigma^2} \right) \right). \qquad (2.11)$$

Setting $\frac{\partial}{\partial |\mathcal{M}|}(EE(\mathcal{M},\mathcal{K})) = 0$, we get

$$(\frac{|\mathcal{K}|\, P_{PA}}{P_C} + |\mathcal{M}|) \sum_{k=1}^{|\mathcal{K}|} \frac{p_u \beta_k}{p_u(|\mathcal{M}|-1)\beta_k + p_u \sum_{i=1, i\neq k}^{|\mathcal{K}|} \beta_i + \sigma^2}$$
$$- \sum_{k=1}^{|\mathcal{K}|} \ln \left(1 + \frac{p_u(|\mathcal{M}|-1)\beta_k}{p_u \sum_{i=1, i\neq k}^{|\mathcal{K}|} \beta_i + \sigma^2} \right) = 0. \qquad (2.12)$$

We define the left-hand side of (2.12) as $\Gamma_1(\mathcal{M})$. Then, the derivative of $\Gamma_1(\mathcal{M})$ with respect to the number of antennas can be expressed as

$$\frac{\partial}{\partial |\mathcal{M}|}(\Gamma_1(\mathcal{M})) = -\left(\frac{|\mathcal{K}|\, P_{PA}}{P_C} + |\mathcal{M}| \right) \sum_{k=1}^{|\mathcal{K}|} \frac{p_u^2 \beta_k^2}{[p_u(|\mathcal{M}|-1)\beta_k + p_u \sum_{i=1, i\neq k}^{|\mathcal{K}|} \beta_i + \sigma^2]^2}. \quad (2.13)$$

It is obvious that $\frac{\partial}{\partial |\mathcal{M}|}(\Gamma_1(\mathcal{M})) < 0$ when $|\mathcal{M}| \geq 0$. Therefore, $\Gamma_1(\mathcal{M})$ is monotone decreasing in $[0, +\infty)$. As $\Gamma_1(0) > 0$ and $\Gamma_1(\infty)$ tends to $-\infty$, there exists one M^* which satisfies $\Gamma_1(M^*) = 0$. Because $\Gamma_1(\mathcal{M})$ is monotone decreasing in $[0, +\infty)$, we can use bisection search to determine M^*. Now, the EE can be expressed as

$$EE(\mathcal{M}^*, \mathcal{K}) = \frac{B \log_2 e}{P_C} \sum_{k=1}^{|\mathcal{K}|} \frac{p_u \beta_k}{p_u(M^*-1)\beta_k + p_u \sum\limits_{i=1, i\neq k}^{|\mathcal{K}|} \beta_i + 1}. \qquad (2.14)$$

[1]Note that we have applied a relaxation here so integer valued $|\mathcal{M}|$ is assumed to be real valued.

Then the optimization problem in (2.10) becomes

$$\underset{\mathcal{K}}{\text{maximize}}\ EE(\mathcal{M}^*, \mathcal{K}), \tag{2.15}$$

and hence we would like to choose a subset of users $\mathcal{K}_1$ from the total user set $\mathcal{K}$ so that the EE of the system is maximized. The above problem is an NP-hard problem, and couldn't be solved analytically. We assume $\mathbf{q}$ is a $K \times 1$ dimensional vector, where $K = |\mathcal{K}|$, and $q_i \in \{0, 1\}$. Let us define

$$f(\mathbf{q}) = \frac{B}{P_{sum}} \sum_{i=1}^{K} \log_2 \left(1 + \frac{p_u (M^* - 1) q_k \beta_k}{p_u \sum_{i=1, i \neq k}^{K} q_i \beta_i + \sigma^2} \right). \tag{2.16}$$

Then, the optimization problem in (2.15) can be written as

$$\begin{aligned} &\text{maximize} && f(\mathbf{q}) \\ &\text{subject to} && q_i \in \{0, 1\}, i = 1, 2, ..., K. \end{aligned} \tag{2.17}$$

The above optimization problem could be solved efficiently by the cross-entropy framework introduced in [118]. The detailed procedure to obtain the EE-optimal subset of users via the cross-entropy algorithm is shown in Algorithm 2.1 below. In this algorithm, $\rho_r \in (0, 1)$ is a user specified rarity parameter, N_s is the number of samples, $\alpha_s \in (0, 1)$ is a smoothing parameter, $\mathbf{1}_{\{.\}}$ denotes the indicator function, and $\lceil a \rceil$ is the smallest integer which is greater than or equal to a.

Now, we have obtained a pair of $\mathcal{M}^*$ and $\mathcal{K}^*$ with bisection, random selection and cross-entropy algorithm. Since the selected subsets $\mathcal{M}^*$ and $\mathcal{K}^*$ are not necessarily EE-optimal, we propose the following two-step iterative procedure: At each time, we obtain a subset of antennas for the given user set via bisection search and random selection, and then obtain the optimally energy efficient subset of users with the selected antenna subset using cross-entropy algorithm. Does this procedure converge to the EE-optimal $\mathcal{M}^*$ and $\mathcal{K}^*$? The

Algorithm 2.1 Cross-entropy algorithm to get EE optimum subset of users

1: Initialize parameter vector $\hat{\mathbf{v}}_0$. Let $N^e = \lceil \rho_r N_s \rceil$. Set $n_s = K$, $t = 1$ (iteration counter).

2: Generate i.i.d. multivariate Bernoulli random vectors $\mathbf{q}_1$, $\mathbf{q}_2$, $\cdots$, $\mathbf{q}_N$. The probability of each entry in $\mathbf{q}_i$ equals to 1 corresponds to the respective element in $\hat{\mathbf{v}}_{t-1}$. Determine $f(\mathbf{q}_i)$ for every i, and sort them in increasing order: $f(\mathbf{q}_1) \leq f(\mathbf{q}_2) \leq ... \leq f(\mathbf{q}_N)$. Let $\hat{\gamma}_t$ be the sample $(1-\rho_r)$-quantile of $f(\mathbf{q}_i)$, i.e., $\hat{\gamma}_t = f(\mathbf{q}_{N_s - N^e + 1})$.

3: Set $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_{N_s}]$, obtain $\hat{\mathbf{v}}_t$ with $\mathbf{Q}$, where $\hat{v}_{t,i} = \sum_{k=1}^{N_s} (\mathbf{1}_{\{f(\mathbf{q}_i) \geq \hat{\gamma}_t\}} \mathbf{Q}_{k,i}) / \sum_{k=1}^{N_s} \mathbf{1}_{\{f(\mathbf{q}_i) \geq \hat{\gamma}_t\}}$.

4: $\hat{\mathbf{v}}_t = \alpha_s \hat{\mathbf{v}}_t + (1 - \alpha_s) \hat{\mathbf{v}}_{t-1}$.

5: Set $d_t = \max_{1 \leq j \leq n_s} \{\min\{\hat{v}_{t,j}, 1 - \hat{v}_{t,j}\}\}$. If $d_t \leq 10^{-3}$, stop; otherwise, set $t = t + 1$, and return to step 2.

following theorem provides us with the characterization that by using bisection search, random selection and cross-entropy algorithm step by step, we can improve the EE at each iteration and approach the EE-optimal $\mathcal{M}^*$ and $\mathcal{K}^*$.

Theorem 2.1 *Consider a single-cell uplink massive MIMO communication system which consists of a BS with M antennas and K single antenna users. Assume that we determine the number of antennas M_1 and a subset of users $\mathcal{K}_1$ ($K_1 = |\mathcal{K}_1|$) via bisection and cross-entropy algorithms, and the achieved EE is EE_1. Then, we set $\mathcal{K}_1$ as the total user set, and we obtain the number of antennas M_2 and a subset of users K_2 ($K_2 = |\mathcal{K}_2|$) leading to the new EE of EE_2. The EE of the second iteration is no smaller than that obtained in the first iteration, i.e., $EE_1 \leq EE_2$.*

Proof 2.1 *At the beginning of the second iteration, we set $\mathcal{K}_1$ as the total user set and obtain the antenna subset $\mathcal{M}_2$ (possibly different from $\mathcal{M}_1$) with bisection search and random selection. Thus, $EE(\mathcal{M}_2, \mathcal{K}_1) \geqslant EE(\mathcal{M}_1, \mathcal{K}_1)$. With antenna subset $\mathcal{M}_2$ and the total user set $\mathcal{K}$, we select the user subset $\mathcal{K}_2$ with cross-entropy algorithm. Since $\mathcal{K}_2$ is the best user subset that maximizes the energy efficiency with antenna set $\mathcal{M}_2$, we have $EE(\mathcal{M}_2, \mathcal{K}_2) \geqslant EE(\mathcal{M}_2, \mathcal{K}_1)$. Therefore, $EE(\mathcal{M}_2, \mathcal{K}_2) \geqslant EE(\mathcal{M}_1, \mathcal{K}_1)$, i.e., $EE_1 \leq EE_2$.* □

From Theorem 1, we know that the EE obtained in the second iteration is no smaller than that obtained at the first time. With the two-step iterative procedure, we obtain a

sequence EE_1, EE_2, EE_3, $\cdots$, where EE_k is the optimal EE achieved at the kth iteration. By Theorem 1, we have $EE_1 \leqslant EE_2 \leqslant EE_3 \leqslant \cdots$. Note that for a given antenna set $\mathcal{M}$ and user set $\mathcal{K}$, the maximum EE that we can obtain is upper bounded. Therefore, this iterative procedure will converge to the EE-optimal $\mathcal{M}^*$ and $\mathcal{K}^*$.

Remark 2.1 *The M^* we determined is in general a real number, while the number of antennas should be integer valued. If M^* is real valued, we can determine the corresponding subset of users $\mathcal{K}_1$, $\mathcal{K}_2$ with $\lceil M^* \rceil - 1$, $\lceil M^* \rceil$, respectively, using the cross-entropy algorithm. Then we select either $\lceil M^* \rceil - 1$ or $\lceil M^* \rceil$, as the optimal number of antennas, depending on which leads to a larger EE.*

2.3 Numerical Results

The cell radius used in the simulations is $r_c = 250m$, and the cell hole radius is $r_h = 35m$ (there is no user in the cell hole). The path loss decay exponent of the wireless channel is $\alpha = 2.5$, and the shadow fading standard deviation is $\sigma_{shadow} = 8dB$. The bandwidth is $20MHz$. The noise power spectral density is $\frac{N_0}{2} = -133dBm/Hz$. SNR is $30dB$. VLSI processing efficiency is $\varrho_e = 5Gflop/W$. Power amplifier efficiency at the user terminals are 22%. RF front-end power consumption P_{RF} is $97.5mW$. Some other parameters of the power consumption model are given in Table 2.1. We assume that the user distribution within the cell does not change, i.e., $\beta_k, k \in \{1, ..., K\}$ stays the same during simulation.

Fig. 2.1 provides the simulation results of EE versus M and K for the MRC receiver. The grid plots the EE with no antenna and user selection. The red curve is the EE with only antenna selection at the BS. For a given K, this curve shows the points which have the largest EE. We can see that the red curve is a bit lower than the grid. That is because we use a lower bound on the channel capacity in the algorithm. But this does not prevent us from getting the optimum M. Simulation results show that the optimum M which we obtain with the bisection algorithm matches well with the Monte Carlo ergodic results. The

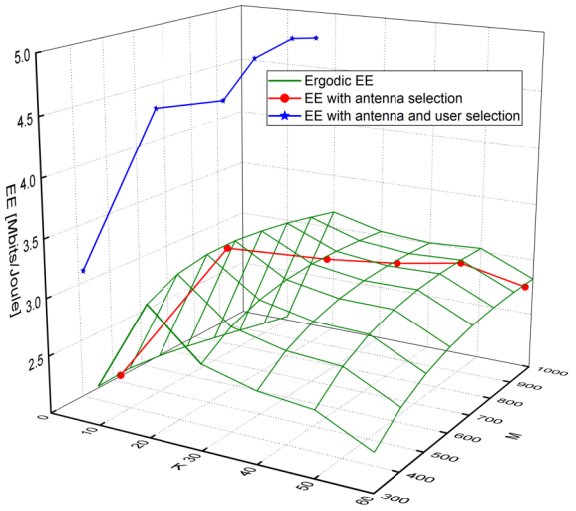

Figure 2.1: EE versus M and K for MRC receiver with different strategies.

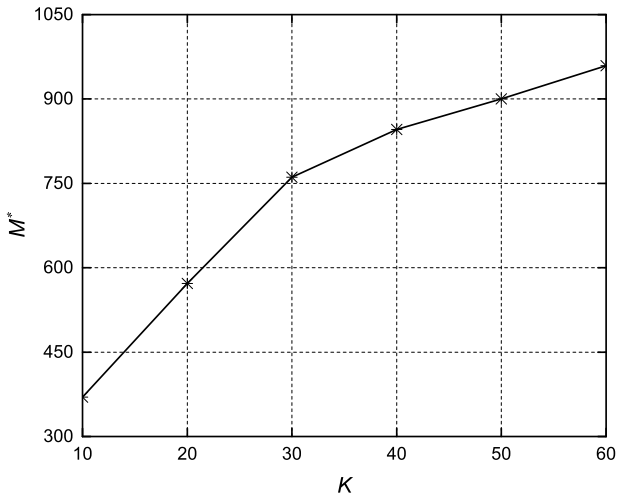

Figure 2.2: The optimum number of M^* versus K for MRC receiver.

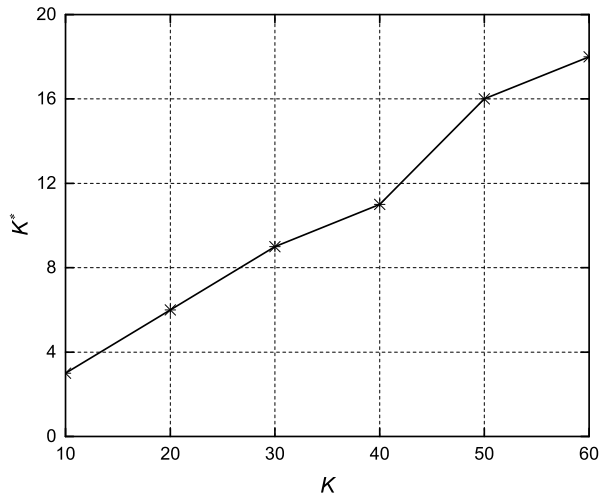

Figure 2.3: The optimum number of K^* versus K for MRC receiver.

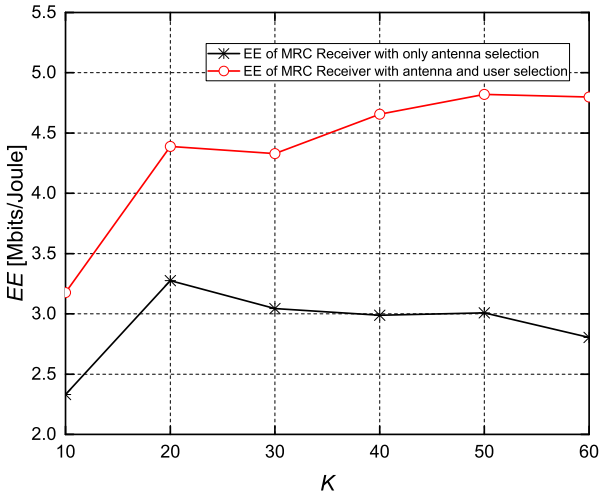

Figure 2.4: Energy Efficiency of antenna selection with and without user selection for MRC receiver.

blue curve demonstrates the EE with both antenna and user selection. The EE has been improved by 71.16% for the MRC receiver when $K = 60$. We can also conclude that the improvement in EE becomes increasingly larger as M and K grow.

In order to show the results in Fig. 1 in more detail, we present Figs. 2, 3 and 4. Fig. 2.2 depicts the optimum number of antennas M^* versus the total number of users K. Fig. 2.3 gives the overall optimum number of users K^* versus the total number of users K. Fig. 2.4 plots the EE with only antenna selection or both antenna and user selection at the BS versus K.

Chapter 3

Statistical Learning Based Joint Antenna Selection and User Scheduling for Single-Cell Massive MIMO Systems

In this chapter, we propose a learning-based stochastic gradient descent algorithm to obtain the optimal joint uplink and downlink EE of single-cell massive MIMO systems with joint antenna selection and user scheduling, under a limitation on the number of available RF chains. With Jensen's inequality and the power consumption model, the original joint antenna selection and user scheduling problem is converted into a combinatorial optimization problem. The learning-based stochastic gradient descent algorithm proposed in this chapter to solve the corresponding combinatorial optimization problem is based on the learning framework proposed in [119]. However, the original learning framework for combinatorial optimization problem does not converge to the optimum value of the objective function, and it is difficult to generate random samples based on the given distribution parameters directly when the constraints for the problem are strict. By generating a population of N_s samples instead of

only one sample and selecting the fittest one at each iteration, and using an efficient subset selection method for rare event simulation proposed in [120], we overcome the disadvantages of the original learning algorithm, and devise an efficient learning-based stochastic gradient descent algorithm for the joint antenna selection and user scheduling problem considered in this chapter.

3.1 System Model

Consider a single-cell massive MIMO system consisting of one BS with the antenna set $\mathcal{M}$, and K single-antenna users. The user set is denoted by $\mathcal{K}$. And we have $|\mathcal{M}| = M$ and $|\mathcal{K}| = K$, where $|\cdot|$ denotes the cardinality of a given set. Without loss of generality, we assume that the system operates over a flat-fading channel [1], the transmission bandwidth is B Hz, and the channel coherence bandwidth is B_c Hz. U symbols are transmitted during a time-frequency coherence block. The uplink and downlink transmissions are considered together with fixed ratios of ς^{ul} and ς^{dl}, respectively, with $\varsigma^{ul} + \varsigma^{dl} = 1$. During each channel coherence interval, $U\varsigma^{ul}$ uplink symbols are transmitted first, then the $U\varsigma^{dl}$ downlink symbols. We assume that BS and all users are perfectly synchronized and operating according to the time-division duplex (TDD) protocol. The uplink and downlink channels are considered to be reciprocal and the uplink channel estimation at the BS could be used for both uplink reception and downlink transmission. As shown in Fig. 3.1, $\tau^{ul}K$ pilot symbols are used during uplink transmission for channel estimation, while another $\tau^{dl}K$ pilot symbols are used during downlink transmission to estimate each user's effective channel and interference variance under the current precoding [1]. We assume τ^{ul}, $\tau^{dl} \geqslant 1$ to enable orthogonal pilot sequences among users.

The pilot sequences used by the K users during channel estimation are $\sqrt{p_{pilot}}\mathbf{\Phi}^T$, where $\mathbf{\Phi}^H\mathbf{\Phi} = \tau^{ul}K\mathbf{I}_K$ and p_{pilot} is the transmit power of pilot signals. Then, the received pilot

[1] For frequency selective channels, orthogonal frequency division multiplexing (OFDM) can, for instance, be used to create flat-fading subchannels.

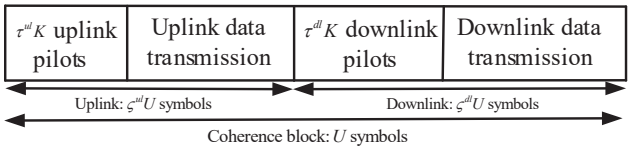

Figure 3.1: Illustration of TDD protocol, where ς^{ul} and ς^{dl} are the fractions of uplink and downlink transmission, respectively.

signal at the BS during each channel coherence interval is

$$\mathbf{Y}_{pilot} = \sqrt{p_{pilot}}\mathbf{G}\mathbf{\Phi}^T + \mathbf{N} \tag{3.1}$$

where $\mathbf{N}$ is the AWGN noise matrix with i.i.d. complex Gaussian $\mathcal{CN}(0, \sigma^2)$ components with zero mean and variance σ^2, and $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_K] \in \mathcal{C}^{M\times K}$ is the matrix describing the channel from all the users to the BS. More specifically, $\mathbf{g}_k$ is the uplink channel vector from the kth user to all the antennas at the BS. For flat fading channels, $\mathbf{g}_k$ can be written as [2],

$$\mathbf{g}_k = \beta_k \mathbf{h}_k$$

where the vector $\mathbf{h}_k$ of fading coefficients is composed of i.i.d. $\mathcal{CN}(0, 1)$ elements. β_k describes the effects of path loss and shadow fading, and can be expressed as

$$\beta_k = \frac{z_k}{(r_k/d_0)^\alpha}$$

where r_k is the distance between the kth user and the BS, d_0 is a reference distance, α is the path loss decay exponent, and the shadow fading coefficient z_k is a log-normal random variable, i.e., $10\log_{10}(z_k)$ is zero-mean Gaussian distributed with standard deviation σ_{shad}.

The received uplink data signal at the BS is

$$\mathbf{y}^{ul} = \sqrt{p^{ul}}\mathbf{G}\mathbf{x} + \mathbf{n} \tag{3.2}$$

where $\sqrt{p^{ul}}\mathbf{x}$ denotes the signal vector transmitted from the users to the BS (with $E\{\mathbf{x}\mathbf{x}^H\} =$

$\mathbf{I}_K$) and $\mathbf{n} \sim CN(0, \sigma^2\mathbf{I}_M)$ is the AWGN noise vector at the BS.

During downlink data transmission, the received signal vector of all users can be expressed as

$$\mathbf{y}^{dl} = \sqrt{p^{dl}}\mathbf{G}^H\mathbf{W}\mathbf{x} + \mathbf{n} \tag{3.3}$$

where $\sqrt{p^{dl}}\mathbf{x}$ is the signal transmitted from the BS to users, and $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_K] \in \mathcal{C}^{M\times K}$ is the precoding matrix at the BS.

3.2 Linear Processing in Single-Cell Massive MIMO Systems

As the number of antennas at the BS is very large in massive MIMO systems, linear precoding schemes can obtain near-optimal performance [121]. Therefore, we consider linear processing in this chapter. More specifically, we assume that linear processing with MRC or ZF receiver is employed at the BS during uplink data transmission, and maximum ratio transmission (MRT) or ZF precoding is used during downlink transmission, under both perfect and imperfect CSI assumptions. Finally, we obtain lower bounds for the achievable data rates of different users with different linear processing methods.

3.2.1 Linear Processing with Perfect CSI

In this subsection, we assume the availability of perfect CSI at the BS. Both MRC and ZF receiving strategies are considered at the BS during uplink data transmission, while MRT and ZF precoding schemes are considered for downlink data transmission.

3.2.1.1 Uplink

The received uplink signal at the BS after linear processing can be expressed as

$$\hat{\mathbf{y}}^{ul} = \sqrt{p^{ul}}\mathbf{AGx} + \mathbf{An} \tag{3.4}$$

where $\mathbf{A}$ is the linear combination matrix at the BS and is given by

$$\mathbf{A} = \begin{cases} \mathbf{G}^H & \text{for MRC,} \\ (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H & \text{for ZF.} \end{cases} \tag{3.5}$$

The kth component of $\hat{\mathbf{y}}^{ul}$ can be expressed as

$$\hat{\mathbf{y}}_k^{ul} = \sqrt{p^{ul}}\mathbf{a}_k\mathbf{g}_k x_k + \sqrt{p^{ul}} \sum_{i=1, i\neq k}^{K} \mathbf{a}_k\mathbf{g}_i x_i + \mathbf{a}_k\mathbf{n} \tag{3.6}$$

where $\mathbf{a}_k$ is the kth row of $\mathbf{A}$. Therefore, the SINR corresponding to kth user's signal received at the BS is

$$\gamma_k^{ul} = \frac{p^{ul}\left|\mathbf{a}_k\mathbf{g}_k\right|^2}{p^{ul}\sum\limits_{i=1, i\neq k}^{K}\left|\mathbf{a}_k\mathbf{g}_i\right|^2 + \left\|\mathbf{a}_k\right\|^2\sigma^2}. \tag{3.7}$$

Then the achievable uplink data rate (bps/Hz) of kth user is

$$R_k^{ul} = E\{\log_2\left(1+\gamma_k^{ul}\right)\}. \tag{3.8}$$

Using Jensen's inequality and the characteristics of wireless channels, we obtain the following lower bounds on R_k^{ul} when MRC and ZF schemes are employed, respectively, at the BS [117]:

$$\widetilde{R}_k^{MRC,ul} = \log_2\left(1 + \frac{p^{ul}(M-1)\beta_k}{p^{ul}\sum\limits_{i=1, i\neq k}^{K}\beta_i + \sigma^2}\right), \tag{3.9}$$

$$\widetilde{R}_k^{ZF,ul} = \log_2\left(1 + \frac{p^{ul}(M-K)\beta_k}{\sigma^2}\right). \tag{3.10}$$

Since the number of antennas at the BS should be no less than the number of users to achieve space multiplicity and diversity gains in MIMO systems, the condition $M \geq K$ should be satisfied in (3.9). Because $E\{[(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}]_{kk}\} = \frac{1}{(M-K)\beta_k}$, $M \geq K+1$ should be satisfied in (3.10).

3.2.1.2 Downlink

We assume MRT or ZF precoding is employed at the BS for downlink transmission and the precoding vector can be expressed as

$$\mathbf{w}_k = \frac{\mathbf{a}_k^H}{\|\mathbf{a}_k\|}. \tag{3.11}$$

The received signal at the kth user is

$$\mathbf{y}_k^{dl} = \sqrt{p^{dl}}\mathbf{g}_k^H\mathbf{w}_k x_k + \sqrt{p^{dl}}\sum_{i=1,i\neq k}^{K}\mathbf{g}_k^H\mathbf{w}_i x_i + n_k. \tag{3.12}$$

Thus, the SINR at the kth user during downlink transmission is

$$\gamma_k^{dl} = \frac{p^{dl}\left|\mathbf{g}_k^H\mathbf{w}_k\right|^2}{p^{dl}\sum\limits_{i=1,i\neq k}^{K}\left|\mathbf{g}_k^H\mathbf{w}_i\right|^2 + \sigma^2}. \tag{3.13}$$

Then the achievable downlink data rate (bits/s/Hz) of the kth user is

$$R_k^{dl} = E\{\log_2(1+\gamma_k^{dl})\}. \tag{3.14}$$

Following the same procedure as in [117], we obtain the following lower bounds on R_k^{dl} with MRT and ZF precoding, respectively:

$$\widetilde{R}_k^{MRT,dl} = \log_2\left(1 + \frac{p^{dl}(M-1)\beta_k}{p^{dl}(K-1)\beta_k + \sigma^2}\right), \tag{3.15}$$

$$\widetilde{R}_k^{ZF,dl} = \log_2\left(1 + \frac{p^{dl}(M-K)\beta_k}{\sigma^2}\right). \tag{3.16}$$

Similar to (3.9) and (3.10), (3.15) and (3.16) require that the conditions of $M \geq K$ and $M \geq K+1$, respectively, are satisfied. Then, the lower bound on the total achievable data rate in the system is

$$\begin{aligned}\widetilde{R} &= \widetilde{R}^{ul} + \widetilde{R}^{dl} \\ &= a\sum_{k=1}^{K}\widetilde{R}_k^{ul} + b\sum_{k=1}^{K}\widetilde{R}_k^{dl}\end{aligned} \tag{3.17}$$

where

$$a = \varsigma^{ul}\left(1 - \frac{\tau^{ul}K}{U\varsigma^{ul}}\right)B \tag{3.18}$$

$$b = \varsigma^{dl}\left(1 - \frac{\tau^{dl}K}{U\varsigma^{dl}}\right)B \tag{3.19}$$

in which the factor $1-\frac{\tau^{ul}K}{U\varsigma^{ul}}$ and $1-\frac{\tau^{dl}K}{U\varsigma^{dl}}$ account for the uplink and downlink pilot overhead, respectively [1].

3.2.2 Linear Processing with Imperfect CSI

In practice, CSI obtained via channel estimation is imperfect. Let us denote the channel estimate obtained from (3.1) as $\hat{\mathbf{G}}$. Then, the channel estimation error is

$$\boldsymbol{\varepsilon} = \hat{\mathbf{G}} - \mathbf{G}. \tag{3.20}$$

The MMSE estimate of $\mathbf{G}$ given $\mathbf{Y}_{pilot}$ is [122]

$$\widehat{\mathbf{G}} = \frac{1}{\sqrt{p_{pilot}}}\mathbf{Y}_{pilot}\mathbf{\Phi}^{*}\left(\frac{\sigma^2}{p_{pilot}}\mathbf{D}^{-1} + \tau^{ul}K\mathbf{I}_K\right)^{-1} \tag{3.21}$$

where $\mathbf{\Phi}^{*}$ is the complex conjugate of the pilot sequence matrix, and $\mathbf{D} = \text{diag}(\beta_1, \beta_2, ..., \beta_K)$. The error covariance

$$\begin{aligned}\mathbf{C}_{MMSE} &= E\left\{vec(\boldsymbol{\varepsilon}) * vec(\boldsymbol{\varepsilon})^H\right\} \\ &= \left(\mathbf{D}^{-1} + \frac{\tau^{ul}Kp_{pilot}}{\sigma^2}\mathbf{I}_K\right)^{-1} \otimes \mathbf{I}_M\end{aligned} \tag{3.22}$$

where $vec(\cdot)$ and $\otimes$ are matrix vectorization and kronecker product operation, respectively. Therefore, each element of $\boldsymbol{\varepsilon}$ is independent of each other, and the kth column of $\boldsymbol{\varepsilon}$, denoted as $\boldsymbol{\varepsilon}_k$, is a vector of random variables with zero mean and variance

$$\gamma_k^v = \frac{\beta_k\sigma^2}{\tau^{ul}Kp_{pilot}\beta_k + \sigma^2}, k = 1, 2, ..., K. \tag{3.23}$$

Substituting (3.1) into (3.21), we can obtain

$$\widehat{\mathbf{G}} = \left(\tau^{ul}K\mathbf{G} + \frac{\mathbf{N}\mathbf{\Phi}^{*}}{\sqrt{p_{pilot}}}\right)\left(\frac{\sigma^2}{p_{pilot}}\mathbf{D}^{-1} + \tau^{ul}K\mathbf{I}_K\right)^{-1}. \tag{3.24}$$

Therefore, elements of $\widehat{\mathbf{G}}$ are independent of each other, and the kth column of $\hat{\mathbf{G}}$, denoted as $\hat{\mathbf{g}}_k$, is a vector of random variables with zero mean and variance

$$\hat{\beta}_k = \frac{\tau^{ul}Kp_{pilot}\beta_k^2}{\tau^{ul}Kp_{pilot}\beta_k + \sigma^2}, k = 1, 2, ..., K. \tag{3.25}$$

Moreover, $\boldsymbol{\varepsilon}$ is independent of $\mathbf{G}$ and $\hat{\mathbf{G}}$. Now, we can obtain the following lower bounds on the achievable uplink data rate of the kth user under imperfect CSI [117]:

$$\widetilde{R}_k^{MRC,ul} = \log_2\left(1 + \frac{(M-1)\hat{\beta}_k}{\sum_{i=1,i\neq k}^{K}\hat{\beta}_i + \sum_{i=1}^{K}\gamma_i^v + \frac{\sigma^2}{p^{ul}}}\right) \tag{3.26}$$

$$\widetilde{R}_k^{ZF,ul} = \log_2\left(1 + \frac{(M-K)\hat{\beta}_k}{\sum_{i=1}^{K}\gamma_i^v + \frac{\sigma^2}{p^{ul}}}\right) \tag{3.27}$$

where the constraints $M \geq K$ and $M \geq K+1$ are required for (3.26) and (3.27), respectively. For the downlink case, the vector of received signals can be expressed as

$$\begin{aligned}\mathbf{y}^{dl} &= \sqrt{p^{dl}}\mathbf{G}^H\hat{\mathbf{W}}\mathbf{x} + \mathbf{n} \\ &= \sqrt{p^{dl}}\hat{\mathbf{G}}^H\hat{\mathbf{W}}\mathbf{x} - \sqrt{p^{dl}}\boldsymbol{\varepsilon}^H\hat{\mathbf{W}}\mathbf{x} + \mathbf{n}.\end{aligned} \tag{3.28}$$

The received signal at the kth user is

$$\hat{y}_k^{dl} = \sqrt{p^{dl}}\hat{\mathbf{g}}_k^H\hat{\mathbf{w}}_k x_k + \sqrt{p^{dl}}\sum_{i=1,i\neq k}^{K}\hat{\mathbf{g}}_k^H\hat{\mathbf{w}}_i x_i - \sqrt{p^{dl}}\sum_{i=1}^{K}\varepsilon_k^H\hat{\mathbf{w}}_i x_i + \mathbf{n}. \tag{3.29}$$

Then, we can obtain the following lower bounds on the achievable downlink data rate to the kth user under imperfect CSI [117]:

$$\widetilde{R}_k^{MRT,dl} = \log_2\left(1 + \frac{(M-1)\hat{\beta}_k}{(K-1)\hat{\beta}_k + K\gamma_k^v + \frac{\sigma^2}{p^{dl}}}\right) \tag{3.30}$$

$$\widetilde{R}_k^{ZF,dl} = \log_2\left(1 + \frac{(M-K)\hat{\beta}_k}{K\gamma_k^v + \frac{\sigma^2}{p^{dl}}}\right) \tag{3.31}$$

Similarly, $M \geq K$ and $M \geq K+1$, respectively, are required to be satisfied for (3.30) and (3.31).

3.3 Energy Efficiency under a Constraint on the Number of RF Chains

We have obtained lower bounds on the achievable data rate with different linear processing methods under perfect or imperfect CSI in the previous section. In this section, we formulate the EE of massive MIMO systems with a limitation on the number of RF chains. First, we introduce the power consumption model for massive MIMO systems. Subsequently, we express the energy-efficiency maximization problem in massive MIMO systems subject to a constraint on the number of RF chains.

3.3.1 Power Consumption Model

We use the power consumption model proposed in [1]. For the sake of completeness, we describe the main characterizations here. The total power consumption consists of the power consumed by the power amplifiers, transceiver chains, channel estimation process, channel coding and decoding units, load-dependent backhaul, linear processing at the BS, and the fixed power consumption (such as power required for site-cooling, control signaling, load-independent power of backhaul infrastructure and baseband processors). Now, the total power can be expressed as follows:

$$P_{sum} = c + dM + fR \tag{3.32}$$

where

$$c = \begin{cases} c_1 & \text{MRC/MRT,} \\ c_1 + \frac{BK^3}{3UL_{BS}} & \text{ZF/MRT, MRC/ZF and ZF/ZF,} \end{cases} \tag{3.33}$$

$$d = \begin{cases} d_1 + \frac{3BK}{UL_{BS}} & \text{MRC/MRT}, \\ d_1 + \frac{B(3K^2+4K)}{UL_{BS}} & \text{ZF/MRT and MRC/ZF}, \\ d_1 + \frac{B(3K^2+K)}{UL_{BS}} & \text{ZF/ZF}, \end{cases} \tag{3.34}$$

$$f = P_{COD} + P_{DEC} + P_{BT}, \tag{3.35}$$

$$c_1 = K\left(\frac{p^{ul}}{\eta^{ul}} + \frac{p^{dl}}{\eta^{dl}}\right) + \frac{2BK^2}{U}\left(\frac{\tau^{ul}M}{L_{BS}} + \frac{2\tau^{dl}}{L_U}\right) + P_{FIX} + P_{SYN} + KP_U, \tag{3.36}$$

$$d_1 = P_{BS} + \frac{2BK}{L_{BS}}\left(1 - \frac{(\tau^{ul} + \tau^{dl})K}{U}\right). \tag{3.37}$$

Above, different expressions in (3.33) and (3.34) are for different combinations of uplink and downlink linear processing schemes at the BS. For instance, MRC/MRT notation describes that MRC is used at the BS for reception in uplink and MRT is employed at the BS for downlink transmission. The differences in (3.33) and (3.34) are the results of differences in the power consumption of different linear processing methods. The description of the parameters in these equations along with their typical values are given in Table 3.1. Readers can also refer to [1] for more details.

Table 3.1: Baseband power consumption parameters in (3.32) to (3.37)

Parameter description	**Value**
Power amplifier efficiency at the users: η^{ul}	0.3
Power amplifier efficiency at the BS: η^{dl}	0.39
Fixed power consumption: P_{FIX}	18 W
Power consumed by local oscillator at BS: P_{SYN}	2 W
Power for the circuit components at each user: P_U	0.1 W
Power for the circuit components at BS: P_{BS}	1 W
Computational efficiency at BS: L_{BS}	12.8 Gflops/W
Power required for coding of data signals: P_{COD}	0.1 W/(Gbit/s)
Power required for decoding of data signals: P_{DEC}	0.8 W/(Gbit/s)
Power required for backhaul traffic: P_{BT}	0.25 W/(Gbit/s)

3.3.2 Energy Efficiency for ZF/ZF under RF Chain Constraints

As the formula of the lower bounds on the achievable data rate with different linear processing methods are similar, we consider only ZF receiver in uplink and ZF precoding in downlink in this subsection. Note that since the formulas for the lower bounds on the achievable data rate under imperfect CSI are almost the same as those under perfect CSI, we only formulate the energy-efficiency maximization problem under perfect CSI. The results for the combination of other linear processing methods under perfect or imperfect CSI can be obtained easily following the same procedure introduced in this chapter. For the sake of brevity in notations, we will eliminate the subscript "ZF/ZF" in the following equations for the results of ZF receiver in uplink and ZF precoding in downlink. The lower bound on the EE with ZF reception in uplink and ZF precoding in downlink under perfect CSI can be expressed as

$$EE(\mathcal{M},\mathcal{K}) = \frac{\widetilde{R}(\mathcal{M},\mathcal{K})}{c + dM + f\widetilde{R}(\mathcal{M},\mathcal{K})} \tag{3.38}$$

where

$$\widetilde{R}(\mathcal{M},\mathcal{K}) = a\sum_{k=1}^{K}\widetilde{R}_k^{ZF,ul} + b\sum_{k=1}^{K}\widetilde{R}_k^{ZF,dl}. \tag{3.39}$$

Our goal is to find a subset of antennas $\mathcal{M}_1$ at the BS and a subset of users $\mathcal{K}_1$ which maximize $EE(\mathcal{M},\mathcal{K})$ under a limitation on the number of RF chains. Therefore, the original energy-efficiency maximization problem could be written as follows:

$$\begin{aligned} &\underset{\mathcal{M}_1\subseteq\mathcal{M},\mathcal{K}_1\subseteq\mathcal{K}}{\text{maximize}} \; EE(\mathcal{M}_1,\mathcal{K}_1) \\ &\text{subject to} \;\; |\mathcal{M}_1| \le N_{RF}, |\mathcal{K}_1| \le |\mathcal{M}_1| - 1 \end{aligned} \tag{3.40}$$

where N_{RF} is the number of available RF chains at the BS.

The above problem is an NP-hard problem, and it cannot be solved analytically. We

assume $\mathbf{q}$ is an $N \times 1$ vector, where $N = K + M$, and $q_i \in \{0, 1\}$ for all $i \in \{1, \ldots, N\}$. The first K elements of $\mathbf{q}$ correspond to the user selection results, and the latter M elements correspond to antennas. Define

$$\phi(\mathbf{q}) = \frac{a \sum_{k=1}^{K} \log_2 \left(1 + \frac{p^{ul}(M_1 - K_1) q_k \beta_k}{\sigma^2}\right) + b \sum_{k=1}^{K} \log_2 \left(1 + \frac{p^{dl}(M_1 - K_1) q_k \beta_k}{\sigma^2}\right)}{c + dM + f\left(a \sum_{k=1}^{K} \log_2 \left(1 + \frac{p^{ul}(M_1 - K_1) q_k \beta_k}{\sigma^2}\right) + b \sum_{k=1}^{K} \log_2 \left(1 + \frac{p^{dl}(M_1 - K_1) q_k \beta_k}{\sigma^2}\right)\right)} \tag{3.41}$$

where $K_1 = \sum_{i=1}^{K} q_i$ is the number of selected users, and $M_1 = \sum_{i=K+1}^{N} q_i$ is the number of selected antennas. Then, the optimization problem in (3.40) can be rewritten as

$$\begin{aligned} &\text{maximize } \phi(\mathbf{q}) \\ &\text{subject to } \sum_{i=K+1}^{N} q_i \leqslant N_{RF}, \\ &\qquad \sum_{i=1}^{K} q_i \leqslant \sum_{i=K+1}^{N} q_i - 1, \\ &\qquad q_i \in \{0, 1\}, i = 1, 2, ..., K, \end{aligned} \tag{3.42}$$

which is a typical combinatorial optimization problem, and it can be solved efficiently via the Gibbs-sampling based method. In the following sections, we study how to obtain the energy-efficiency maximizing subset of users and antennas via a learning-based stochastic gradient descent method.

3.4 Learning-Based Stochastic Gradient Descent Combinatorial Optimization Algorithm

In this section, we analyze how to solve the combinatorial optimization problem using the learning-based stochastic gradient descent method. For the completeness of this chapter, we initially review the original learning-based stochastic gradient descent method proposed

in [119], and then address its advantages and disadvantages in solving the problem in (3.42). Following this, we provide several extensions to the original learning-based stochastic gradient descent method to make it an efficient algorithm for solving the problem in (3.42).

3.4.1 Learning-Based Stochastic Gradient Descent Method

Suppose we have a combinatorial optimization problem with N features $\mathbf{q} = [q_1, q_2, ..., q_N]^T$, $q_i \in \{0, 1\}$, and we want to minimize the objective function $\varphi(\mathbf{q})$. In [119], a learning-based stochastic gradient descent algorithm is proposed for solving this problem based on the characteristics of Gibbs distribution and dynamical systems.

The Gibbs distribution maps each value of the objective function onto a probability defined by

$$p^*_{T_t}(\mathbf{q}) = \frac{\exp(-\varphi(\mathbf{q})/T_t)}{\sum_{\mathbf{y}\in\mathcal{S}} \exp(-\varphi(\mathbf{q})/T_t)} \tag{3.43}$$

where $T_t > 0$ is the analogue of a temperature and $\mathcal{S} = \{0, 1\}^N$ is the set of all possible $\mathbf{q}$. Let us define $\mathcal{U} = \{\mathbf{q} \in \mathcal{S} | \forall \mathbf{q}_1 \in \mathcal{S}, \varphi(\mathbf{q}) \leq \varphi(\mathbf{q}_1)\}$. The Gibbs distribution $p^*_{T_t}$ converges to a uniform distribution on $\mathcal{U}$ when T_t tends to zero. In other words, we can obtain the optimal solutions by finding the limit distribution of $p^*_{T_t}$ as $T_t \to 0$. However, this is difficult in practice. So instead of finding the limit distribution directly, we search for a distribution which has the smallest Kullback-Leibler (KL) divergence to an implicit Gibbs distribution. The KL divergence between p and $p^*_{T_t}$ is

$$D(p, p^*_{T_t}) = -\sum_{\mathbf{q}\in\mathcal{S}} p(\mathbf{q}) \ln \frac{p^*_{T_t}(\mathbf{q})}{p(\mathbf{q})} \tag{3.44}$$

where $\ln(\cdot)$ is the natural logarithm function. Then, the problem is converted into the minimization of the free energy of the system:

$$F_e = \sum\nolimits_{\mathbf{q}\in\mathcal{S}} p(\mathbf{q})\varphi(\mathbf{q}) + T_t \sum\nolimits_{\mathbf{q}\in\mathcal{S}} p(\mathbf{q}) \ln(p(\mathbf{q})). \tag{3.45}$$

This is still a discrete problem, and it is not easy to solve in practice. By introducing $\boldsymbol{\theta}_p = [\theta_{1,p}, \theta_{2,p}, ..., \theta_{N,p}]^T$ as the probability distribution parameter for the N-dimensional random vector $\mathbf{q} = [q_1, q_2, ..., q_N]^T$, we can convert the discrete optimization into a continuous optimization problem. Then, we introduce the following dynamical system

$$\frac{\partial \boldsymbol{\theta}_p}{\partial t} + \frac{\partial F_e}{\partial \boldsymbol{\theta}_p} = 0. \tag{3.46}$$

At last, we obtain the following statistical update rule:

$$\theta_{i,p}(t+1) = \theta_{i,p}(t) - \alpha_l \left(\varphi(\mathbf{q}) + T_t(1 + \ln(p(\mathbf{q})))\right) \frac{\partial \ln(p(\mathbf{q}))}{\partial \theta_{i,p}} \tag{3.47}$$

where α_l is the learning rate, and $\frac{\partial \ln(p(\mathbf{q}))}{\partial \theta_{i,p}}$ is the gradient.

Suppose we choose the random variables q_i to be independent binomially distributed. Then, the joint distribution of the random vector $\mathbf{q}$ is as follows:

$$p(\mathbf{q}) = \prod_{i=1}^{N} \left(q_i p_i + (1 - q_i)(1 - p_i)\right) \tag{3.48}$$

where p_i is the probability that q_i equals to 1. The relationship between p_i and $\theta_{i,p}$ is

$$p_i = \frac{1}{2}(1 + \tanh(\beta_l \theta_{i,p})), \tag{3.49}$$

and the gradient of $\ln(p(\mathbf{q}))$ is

$$\frac{\partial \ln(p(\mathbf{q}))}{\partial \theta_{i,p}} = 2\beta_l (q_i - p_i). \tag{3.50}$$

With this, the update rule in (3.47) becomes

$$\theta_{i,p}(t+1) = \theta_{i,p}(t) - 2\alpha_l \beta_l \left(\varphi(\mathbf{q}) + T_t(1 + \ln(p(\mathbf{q})))\right)(q_i - p_i). \tag{3.51}$$

As noted before, the Gibbs distribution converges to a uniform distribution which achieves the optimal solutions, and Algorithm 3.1 below will obtain the optimal solutions with sufficiently many iterations. If we use the Metropolis algorithm to update the Gibbs distribution to get the optimal solutions, then we have the well-known simulated annealing algorithm. As the objective function decreases fastest in the gradient direction, Algorithm 3.1 will arrive at the optimal solution quicker than the simulated annealing algorithm.

If we are interested in finding the maximum value of an objective function, we just need to add a negative sign before the objective function, and then substitute it into our algorithm.

Below, we provide the learning-based combinatorial optimization algorithm using the stochastic gradient descent as Algorithm 3.1.

Algorithm 3.1 Learning-based combinatorial optimization algorithm.

1: Initialize the multivariate Bernoulli distribution parameter $\boldsymbol{\theta}_p$, learning rate α_l, and β_l;
2: Generate an N-dimensional Bernoulli random vector $\mathbf{q}$ which satisfies all the constraints in the combinatorial optimization problem, with parameter vector $\boldsymbol{\theta}_p$;
3: Calculate the objective function value $\varphi(\mathbf{q})$;
4: Update $\boldsymbol{\theta}_p$ with (3.51);
5: If the stop criteria is met, stop; otherwise, go to step 2.

3.4.2 Drawbacks of Algorithm 3.1 and the Corresponding Solutions

Although Algorithm 3.1 could arrive at the maximum value of the objective function, fluctuations occur, slowing the convergence.

Another drawback lies in step 2 of Algorithm 3.1, which should generate an N-dimensional multivariate Bernoulli random vector $\mathbf{q}$ which satisfies all the constraints with parameter vector $\boldsymbol{\theta}_p$. It may become difficult to produce this N-dimensional multivariate Bernoulli random vector directly as the constraints for the combinatorial optimization problem are strict, e.g., especially when the probability of the N-dimensional multivariate Bernoulli random vector is very small, such as less than 10^{-10}. Such low probabilities may be experienced in practice

and such cases require a large number of samples to get a realization of the event.

These observations motivate us to provide the following modifications to Algorithm 3.1 and use an efficient rare event simulation method to overcome these drawbacks:

3.4.2.1 Convergence

The fluctuation problem of Algorithm 3.1 could be solved by some revisions in steps 2 and 3. Instead of generating only one sample, we generate a population of N_s individuals from the N-dimensional multivariate Bernoulli distribution, which satisfy the constraints with parameter vector $\boldsymbol{\theta}_p$ in step 2. Then we determine the objective function value for each individual and select the one which leads to the smallest value in step 3. Numerical results demonstrate that Algorithm 3.1 converges to the optimal solutions quickly with these changes.

3.4.2.2 Rare event simulation

The second drawback of Algorithm 3.1 is overcome with the efficient subset simulation method for rare event estimation, proposed by Au and Beck in [120]. The basic idea of subset simulation method is to decompose the rare event F_n into a sequence of progressively "less-rare" nested events $F_n = F_{m,n} \subset F_{m-1,n} \subset \cdots \subset F_{1,n}$, where $F_{1,n}$ is a relatively frequent event [123]. Then the small probability $P(F_n)$ of the rare event F_n can be represented as

$$P(F_n) = P(F_{1,n}) \cdot P(F_{2,n}|F_{1,n}) \cdot ... \cdot P(F_{m,n}|F_{m-1,n}) \tag{3.52}$$

where $P(F_{k,n}|F_{k-1,n}) = P(F_{k,n})/P(F_{k-1,n})$ is the conditional probability of $F_{k,n}$ given the occurrence of $F_{k-1,n}$, for $k = 2, ..., m$. With this, the estimation of the rare event problem is transferred to the product of relatively frequent events. In practice, it is always not obvious how to decompose the rare event into a sequence of relatively frequent events. This could be done adaptively via the Markov chain Monte Carlo technique [123].

With the former two revisions, we obtain the learning-based stochastic gradient descent algorithm for EE maximization as described in Algorithm 3.2 below.

Algorithm 3.2 Learning-based stochastic gradient descent algorithm for EE maximization.

1: Initialize the multivariate Bernoulli distribution parameter $\boldsymbol{\theta}_p$, learning rate α_l, β_l, and the number of samples at each iteration N_s;
2: Generate N_s N-dimensional multivariate Bernoulli random vectors $\mathbf{q}$ with parameter vector $\boldsymbol{\theta}_p$;
3: If none of the N_s multivariate Bernoulli random vectors $\mathbf{q}$ generated in step 2 satisfies the constraints in (3.42), we use the subset simulation method introduced in this section to produce another N_s multivariate Bernoulli random vectors $\mathbf{q}$, instead of the vectors produced in step 2.
4: Calculate $\phi(\mathbf{q})$ for each of the multivariate Bernoulli random vectors $\mathbf{q}$ generated in step 2 or 3, which satisfy the constraints in (3.42). Then, select the one that maximize $\phi(\mathbf{q})$;
5: Update $\boldsymbol{\theta}_p$ with (3.51). As the goal is to maximize the EE, we substitute $\varphi(\mathbf{q}) = -\phi(\mathbf{q})$ into (3.51);
6: If the stop criteria is met, stop; otherwise, go to step 2.

We usually stop learning when EE converges to some value. The subset of antennas at the BS and subset of users are selected jointly in Algorithm 3.2. Numerical simulation results in the following section will show that Algorithm 3.2 is very efficient to solve the EE maximization problem subject to a limitation on the number of RF chains.

In order to see the effect of RF chains constraint on the maximum EE that the system could achieve, We also consider the case where there is no RF chains constraint for the selected number of antennas. In this situation, it is equivalent to the case that the number of RF chains constraint equals the total number of antennas at the BS, i.e., $N_{RF} = M$ in equation (3.42).

3.5 Numerical Results

In this section, we provide numerical results to analyze the performance. More specifically, we primarily focus the EE achieved when a ZF receiver is used at the BS for uplink data reception, and ZF precoding is employed at the BS for downlink data transmission (abbreviated as the ZF/ZF strategy). We analyze the EE as a function of the SNR and the number

of available RF chains. Performance levels achieved with other uplink receivers and downlink precoders, addressed in Section 3.2, can be determined similarly. In Section 3.5.4, we compare our learning-based stochastic gradient descent method with the algorithm proposed in [1], and demonstrate that our algorithm can further improve the EE of the system.

In Section 3.5.1, 3.5.2 and 3.5.3, we consider a single-cell massive MIMO system with a radius of 1000 m. Users are uniformly distributed in the cell, and we assume that there is no user within the radius of 100 m. The path loss decay exponent is $\alpha = 3.8$, and the shadow fading has a standard deviation of $\sigma_{shad} = 8$ dB. The transmission bandwidth is $B = 20$ MHz, while the channel coherence bandwidth is $B_c = 180$ kHz. The number of symbols transmitted during a time-frequency coherence block is $U = 1800$. The total noise power at the BS is -96 dBm. The relative pilot length during uplink and downlink channel estimation is $\tau^{ul} = \tau^{dl} = 1$, i.e., the number of pilots equal to the number of users. The uplink and downlink transmission ratios are $\varsigma^{ul} = 0.4$ and $\varsigma^{dl} = 0.6$. The parameters related to the baseband power consumption model are given in Table 3.1. The energy efficiency is averaged with 5000 independent realizations.

3.5.1 Convergence of Algorithm 3.2

Algorithm 3.2 can be regarded as a variant of stochastic gradient decent algorithms, which have been widely used to solve the optimization problems with nonconvex objective functions. Indeed, while being different, Algorithm 3.2 has certain similarities to the stochastic gradient descent algorithms addressed in [124–126], where the convergence properties of such algorithms are discussed in detail. Additionally, while establishing theoretical guarantees for the convergence is challenging, the convergence of Algorithm 3.2 can be addressed and demonstrated by numerical analysis in the settings considered in the simulations. Fig. 3.2 plots the achieved EE with Algorithm 3.2 versus the number of iterations with imperfect CSI under RF chain constraints for a single realization of the channel coefficients. We observe that Algorithm 3.2 converges after about 70 iterations. Numerical results show that

the convergence tendency of Algorithm 3.2 under different settings are similar to Fig. 3.2.

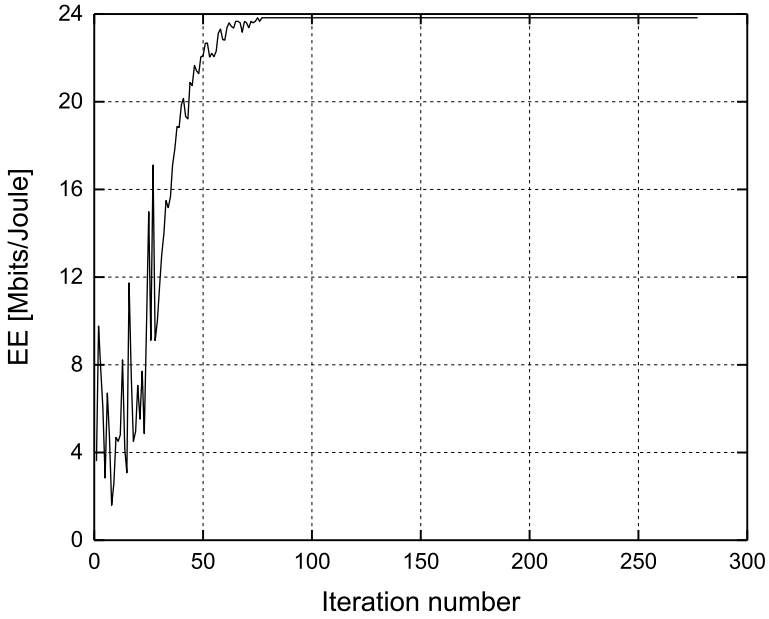

Figure 3.2: EE versus number of iterations under imperfect CSI, SNR=20 dB, RF=15, 60 antennas at the BS and 50 users uniformly distributed in the cell.

Moreover, since the optimal solution of combinatorial optimization problems can be obtained by exhaustive search, it is obvious that if it achieves the same performance as exhaustive search, Algorithm 3.2 converges to the optimal solution. In order to perform exhaustive search, we consider a relatively low-dimensional setting and assume that there are 8 antennas at the BS, and 6 users uniformly distributed in the cell.

In Fig. 3.3, we compare the maximum EE levels achieved with Algorithm 3.2 and also with exhaustive search in single-cell massive MIMO systems, assuming both perfect and imperfect CSI, with and without a limitation on the number of available RF chains. We observe that the performances of Algorithm 3.2 and exhaustive search are indistinguishable, and hence Algorithm 3.2 can attain approximately optimal solutions for problem of joint antenna selection and user scheduling to maximize the achievable joint uplink and downlink EE, i.e., Algorithm 3.2 converges to the approximately optimal solution points of the corresponding combinatorial optimization problem.

Fig. 3.4 plots the average number of iterations (averaged over channel fading) needed for

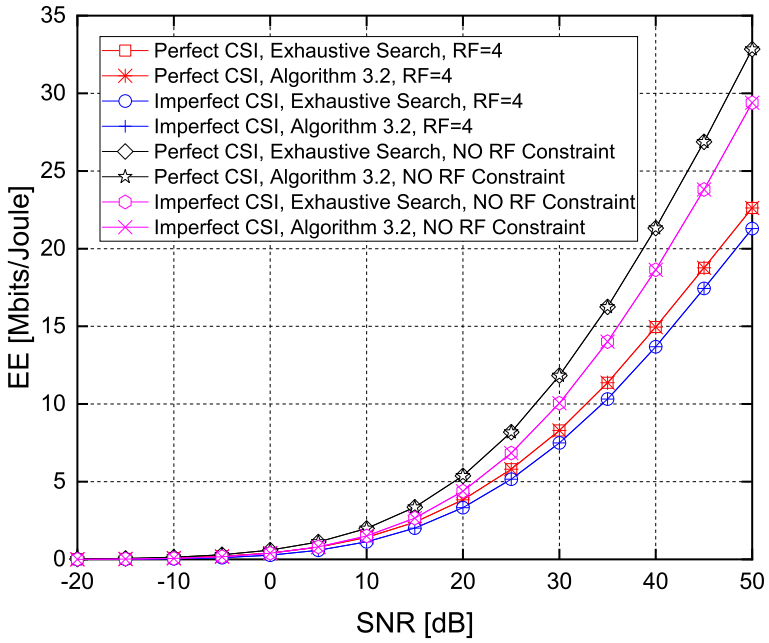

Figure 3.3: Comparison of the maximum EE levels versus SNR achieved with Algorithm 3.2 and exhaustive search method in single-cell massive MIMO systems.

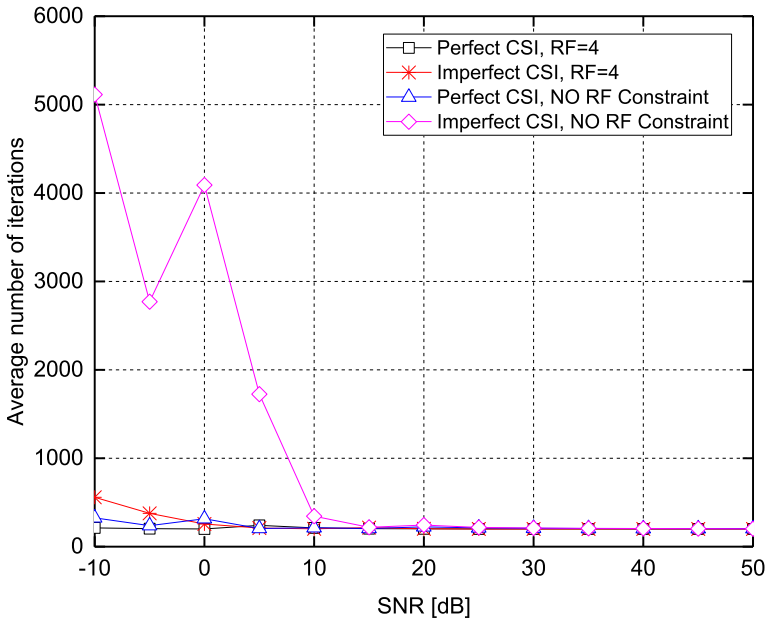

Figure 3.4: Average number of iterations needed for the converge of Algorithm 3.2.

the converge of Algorithm 3.2. In the numerical simulations, when the successive number of objective function values, whose difference between the former and latter is smaller than a given tolerance, is larger than 200, we deem Algorithm 3.2 as converged and stop the iteration process. Therefore, the lower bound of the average iteration number is 200. With this stopping criterion, the total number of iterations before Algorithm 3.2 is stopped in the simulation provided in Fig. 3.2 is 277 (although convergence is attained after about 70 iterations as seen in the figure).

3.5.2 Computational Complexity of Algorithm 3.2

Since Algorithm 3.2 is a Gibbs distribution based stochastic gradient descent method, it is difficult to analyze the computational complexity theoretically. Thus, the computational complexity is analyzed via numerical results. From the convergence analysis before, we know that Algorithm 3.2 can achieve approximately optimal solutions for the corresponding combinatorial optimization problem considered in this chapter, which could also be obtained by exhaustive search. Therefore, we here compare the computational complexities of Algorithm 3.2 and exhaustive search. The former is comprised of generating multivariate Bernoulli random vectors, checking the RF chain constraint, and computing $\phi(\mathbf{q})$, while the latter consists of only checking the RF chain constraint and computing $\phi(\mathbf{q})$. Numerical results demonstrate that the time used for the computation of $\phi(\mathbf{q})$ dominates the computational complexity of both Algorithm 3.2 and exhaustive search. Therefore, the number of times $\phi(\mathbf{q})$ is computed during the simulations is used as a criterion for the computational complexity comparison. In this section, we consider 60 antennas at the BS, and 50 users uniformly distributed in the cell.

Fig. 3.5 plots the curves for the computational complexities of Algorithm 3.2 and the exhaustive search method versus SNR, the number of antennas at BS and the number of users, respectively. The y-axis corresponds to the average number of times $\phi(\mathbf{q})$ is computed during the numerical simulations. All the curves in Fig. 3.5 show that the computational

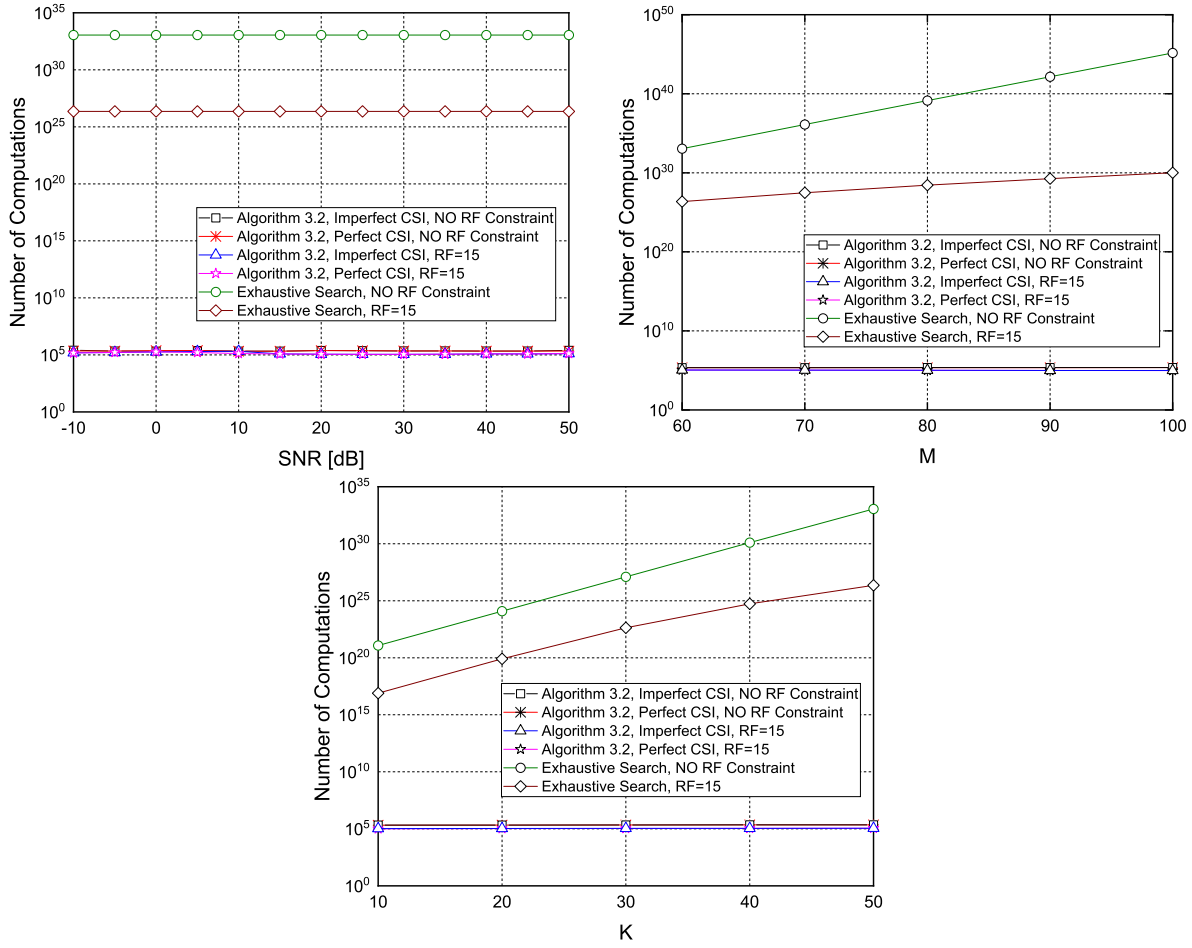

Figure 3.5: Average number of computations of $\phi(\mathbf{q})$ vs. SNR, number of BS antennas, M, and number of users, K, when employing Algorithm 3.2 and exhaustive search method.

complexity of Algorithm 3.2 is significantly (e.g., many orders of magnitude) less than that of the exhaustive search method.

3.5.3 Maximum EE achieved by the ZF/ZF strategy

In this subsection, we assume that there are 60 antennas at the BS, and 50 users uniformly distributed in the cell. In Figs. 3.6, 3.7 and 3.8, we plot the maximum EE (that the single-cell massive MIMO system could achieve) as a function of SNR with perfect and imperfect CSI and with and without RF chain constraints. We observe in Figs. 3.6 and 3.7 that there is almost no difference in EE with and without RF chains constraints at low SNRs, regardless of whether there is perfect (Fig. 3.6) or imperfect CSI (Fig. 3.7). On the other

hand, at medium and large SNR levels, the presence of RF chain constraints leads to a noticeable decrease in the maximum EE. We note that the curves approach straight lines in the high SNR regime, but the slopes decrease when SNR is larger than 20 dB as the RF chain constraints become more stringent. Finally, we observe in Fig. 3.8 that compared to perfect CSI case, the maximum EE is smaller under imperfect CSI with the same SNR. However, the curves in this figure demonstrate that the performance gap is less than 10% under imperfect CSI, compared with that under perfect CSI. Besides, since the magnitude of P_{FIX}, P_{SYN} and P_{BS} are very large compared with the summation of uplink and downlink transmitting power, the dominate component of the power consumption model in (3.32) increases slower than the increasing rate of system's total achievable data rate. Therefore, other than bell-shaped curves, the maximum EE achieved by the system keeps increasing as the SNR increases.

Additionally, we have the following intriguing observations. Under perfect CSI, we notice in Fig. 3.6 that the EE curve with RF = 45 overlaps with the one without a constraint on the number of RF chains, while the EE curves with RF = 35 and RF = 45 both overlap in Fig. 3.7 with the one without RF chain constraints under imperfect CSI. This indicates that under imperfect CSI, a smaller number of RF chains is needed to attain the same performance level achieved in the absence of RF chain constraints. At the same time, it is important to note that the best performance under imperfect CSI is less than the best performance under perfect CSI.

The maximum EE that can be achieved in single-cell massive MIMO systems versus the number of allowed RF chains for different SNRs under perfect and imperfect CSI are plotted in Figs. 3.9, 3.10 and 3.11. Under both perfect and imperfect CSI, EE initially increases as the number of allowed RF chains grows, but the rate of increase slows and EE starts staying fixed after a certain threshold on the number of RF chains. We observe that this RF chain threshold is larger at higher SNR levels. However, at the same SNR, the RF thresholds are interestingly equal under perfect and imperfect CSI, although, as noted before, the maximum

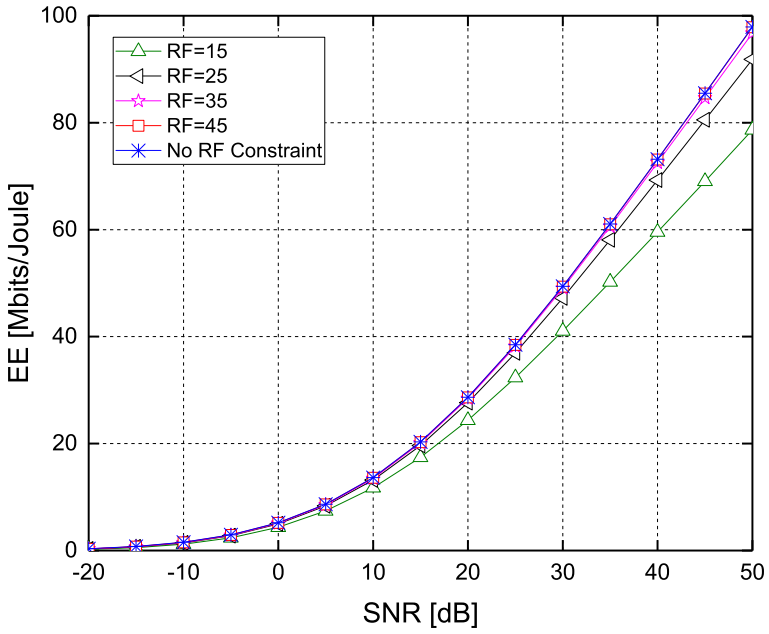

Figure 3.6: Maximum EE achieved in a single-cell massive MIMO system versus SNR with different RF constraints under perfect CSI.

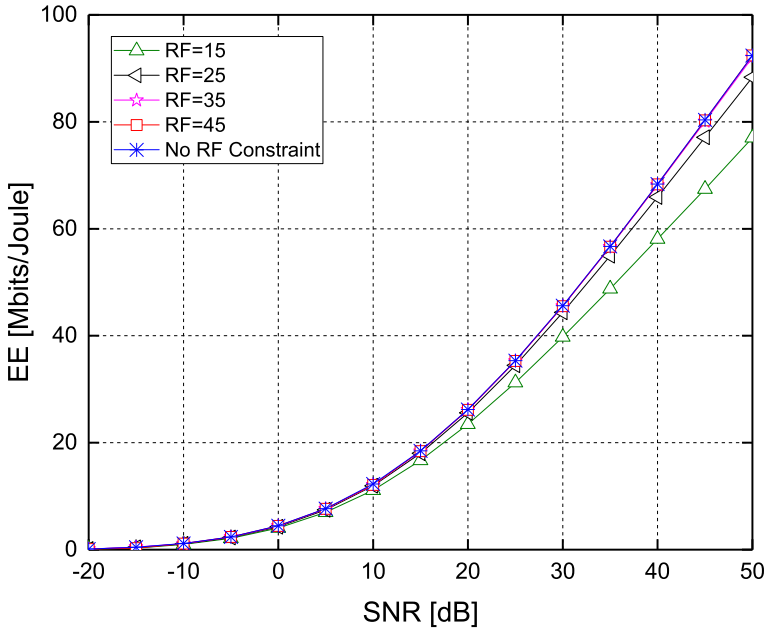

Figure 3.7: Maximum EE achieved in a single-cell massive MIMO system versus SNR with different RF constraints under imperfect CSI.

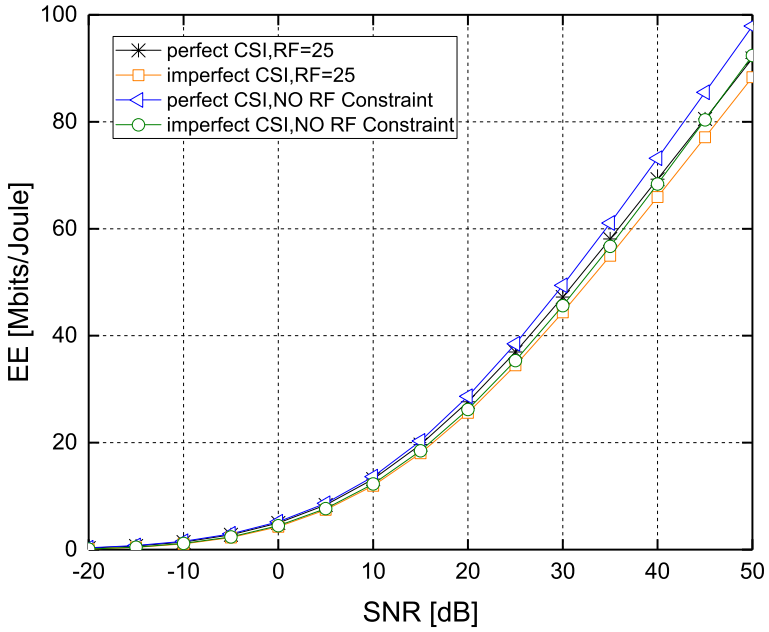

Figure 3.8: Comparison of maximum EE achieved in a single-cell massive MIMO system versus SNR under perfect and imperfect CSI.

EE that can be achieved under imperfect CSI is less than that under perfect CSI.

We remark that the results in Fig. 3.10 could guide the design of single-cell massive MIMO systems in practice. As there are large numbers of antennas at the BS, how many RF chains should be employed is a critical consideration in the design of massive MIMO systems. For instance, in the design of a real system, given the EE requirements, one can determine approximately how many RF chains should be set up at the BS from the results in Fig. 3.10.

3.5.4 Comparison of Algorithm 3.2 and the Method in [1]

Since [1] also addresses joint uplink and downlink EE maximization in single-cell massive MIMO systems, we in this section compare the performance of Algorithm 3.2 with that of the method proposed in [1]. Specifically, we assume that the users are distributed in a circular cell with maximum radius $d_{max} = 250m$ and minimum radius $d_{min} = 35m$. Path-loss is the dominant component in the large-scale fading of users' channels. The large-scale fading is expressed as $\beta_k = \beta_0/(d_k)^\alpha$, where d_k is the distance from the kth user to the BS, the

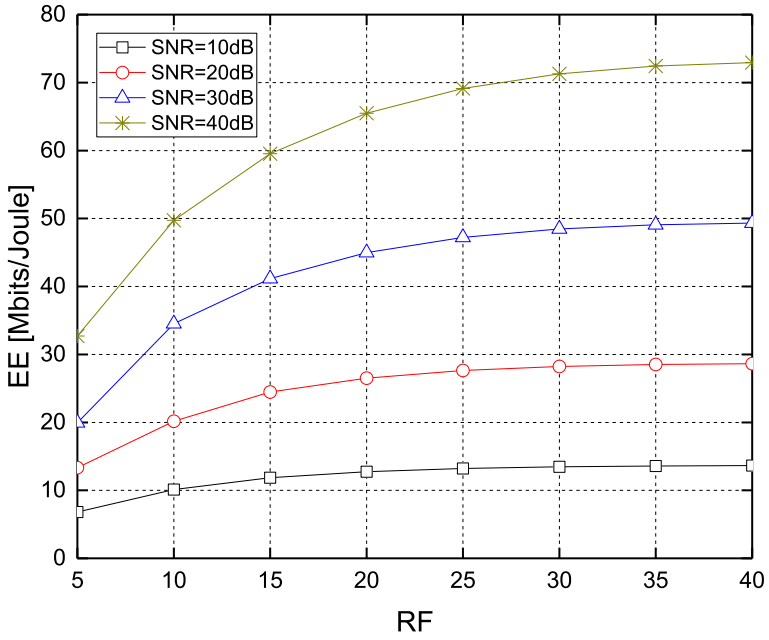

Figure 3.9: Maximum EE achieved in a single-cell massive MIMO system versus RF chains constraint with different SNR under perfect CSI.

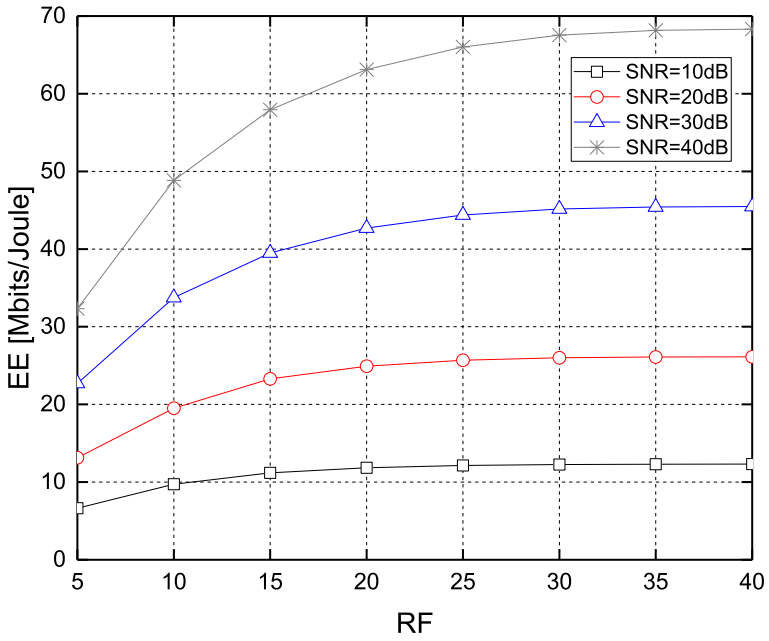

Figure 3.10: Maximum EE achieved in a single-cell massive MIMO system versus RF chains constraint with different SNR under imperfect CSI.

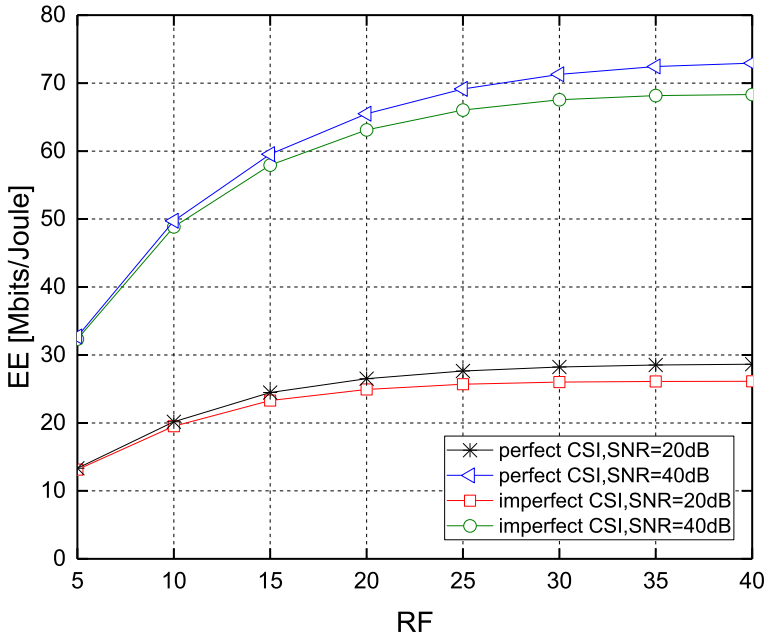

Figure 3.11: Comparison of maximum EE achieved in a single-cell massive MIMO system versus RF chains constraint under perfect and imperfect CSI.

path-loss decay exponent $\alpha = 3.76$, and the large-scale fading at d_{min} is $\beta_0 = 35.3$ dB. We assume that there are 220 antennas at the BS, and 150 users uniformly distributed in the cell. The other parameters are the same as described at the beginning of this section.

Algorithm 3.2 and [1] achieve optimal joint uplink and downlink EE with two entirely different methods. For instance, EE maximization is achieved with joint antenna selection and user scheduling in Algorithm 3.2, while the algorithm proposed in [1] maximizes EE with random antenna and user selection (Specifically, in [1], with a specified power allocation algorithm, each user is assumed to attain the same rate. Then, the EE maximization problem reduces to finding the optimal number of antennas and users, which maximize the EE of the system. Given the optimal numbers of antennas and users, the set of users and antennas are selected randomly). In other words, as its most important aspect, Algorithm 3.2 performs joint antenna selection and user scheduling, while the algorithm proposed in [1] is used to obtain the optimal number of antennas and users. Additionally, Algorithm 3.2 can be employed under any power allocation scheme, while the algorithm proposed in [1] only works under the power allocation scheme which leads to the same rate for all the users.

In order to have fairness in the comparison, we assume equal total transmit power for both Algorithm 3.2 and the method in [1], i.e., the total transmit power for the entire set of users is $B\varsigma^{ul}\sigma^2\rho S_\beta K/\eta^{ul}$ or $B\varsigma^{dl}\sigma^2\rho S_\beta K/\eta^{dl}$, for uplink or downlink communication, respectively, where the design parameter $\rho = p^{ul}\beta_k/\sigma^2 = p^{dl}\beta_k/\sigma^2$, and $S_\beta = E\{\beta^{-1}\}$. For ZF/ZF strategy, each user's transmit power for the algorithm proposed in [1] is proportional to $1/\|\mathbf{g}_k\|^2$. We also employ Algorithm 3.2 under this power allocation scheme. Besides, since Algorithm 3.2 can operate under any power allocation scheme, we also provide numerical results with Algorithm 3.2 under other power allocation schemes, such as equal power allocation, transmit power proportional to β_k or $1/\beta_k$. We would like to point out here that when the transmit power is proportional to $1/\beta_k$, each user would achieve the same rate lower bound.

We assume that the pilot and data signals are transmitted with the same power, and ZF/ZF strategy is used during uplink and downlink data transmissions. Since RF chain constraint is not considered in [1], we will not include this constraint in this section.

Fig. 3.12 and Fig. 3.13 display the maximum EE achieved in a single-cell massive MIMO system versus ρ (proportional to the transmit power) considering perfect and imperfect CSI, respectively. Comparing with the algorithm proposed in [1], Algorithm 3.2 can further improve the EE under both assumptions of perfect and imperfect CSI when equal power allocation scheme is used, or the transmit power is proportional to β_k. The maximum EE achieved with Algorithm 3.2 under transmit power proportional to $1/\beta_k$ overlaps with that achieved under transmit power proportional to $1/\|\mathbf{g}_k\|^2$, for both perfect and imperfect CSI. With equal power allocation and perfect CSI, while the curves of our algorithm and the alternating optimization algorithm in [1] have similar shapes, Algorithm 3.2 improves the EE by more than 40%, compared with the approach in [1]. Under imperfect CSI, the authors in [1] used an exhaustive search method. The curve with equal power allocation in Fig. 3.13 shows that our algorithm can also achieve a substantial improvement, compared with the exhaustive search method in [1].

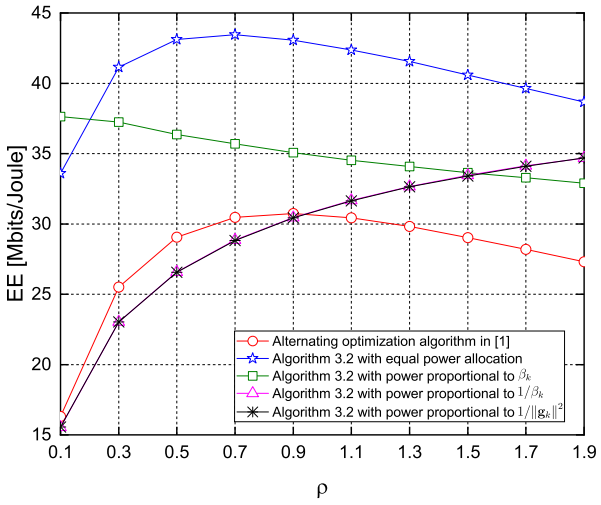

Figure 3.12: Maximum EE achieved in a single-cell massive MIMO system versus ρ under perfect CSI.

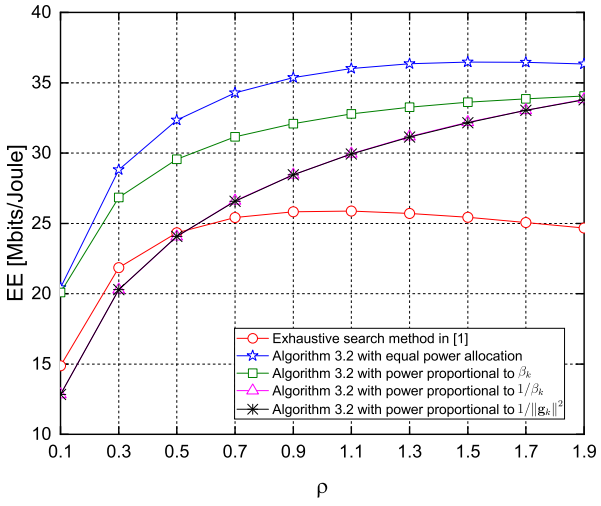

Figure 3.13: Maximum EE achieved in a single-cell massive MIMO system versus ρ under imperfect CSI.

Chapter 4

Joint Activity Detection and Channel Estimation in Cell-Free Massive MIMO Networks with Massive Connectivity

In this chapter, we focus on the asymptotic analysis of SMV based MMSE estimation in cell-free massive MIMO systems with massive connectivity. We address the JADECE problem using the frameworks of MMSE estimation and likelihood ratio test. Prior studies on the MMSE estimation of sparse signals have developed methodologies, including a closed-form method in [78], an approximation method in [79], and AMP and CB-AMP algorithms. Since we consider cell-free massive MIMO in this paper, the non-zero components of the sparse signal vectors to be estimated are i.n.i.d.. However, to the best of our knowledge, all the asymptotic analyses of the MMSE estimation with replica method in the literature focus on the sparse signal vectors with i.i.d. non-zero components. With i.n.i.d. non-zero components, analysis becomes more challenging. Although the authors in [78] have obtained closed-form expressions for the MMSE estimates of the sparse signals with heteroscedastic

non-zero entries, there is a requirement for the measurement matrix to be unitary, while we in our setting have an underdetermined system as a result of the facts that limited number of pilots can be used, and condition of having a unitary measurement matrix cannot be satisfied. An approximation method for the MMSE estimation of sparse signals is provided in [79]. However, when the number of variables becomes large, the computational complexity will grow quickly and the accumulated error will be high.

Inspired by the decoupling property of replica symmetric postulated MMSE estimation for sparse signal vectors with i.i.d. non-zero components [61, 127], we establish a decoupling principle of SMV based MMSE estimation for sparse signal vectors with i.n.i.d. non-zero components, which plays a key role in our theoretical analysis of SMV based MMSE estimation in cell-free massive MIMO networks with massive connectivity. Subsequently, Using the decoupling principle, likelihood ratio test and the optimal fusion rule, we obtain detection rules for the activity of users based on the received pilot signals at only one AP, and also based on the cooperation of the received pilot signals from the entire set of APs for centralized and distributed detection. Then, we determine the false alarm and miss detection probabilities of activity detection in cell-free massive MIMO networks with massive connectivity. We show that the error probabilities of both centralized and distributed activity detection schemes tend to zero when the number of APs tends to infinity while the asymptotic ratio between the number of users and pilots is kept constant. Moreover, we analyze oracle estimation, which provides a lower bound on the mean square error of MMSE estimation for sparse signals, as a benchmark scheme in the case of known user activities. In particular, we investigate the asymptotic behavior of oracle estimation in cell-free massive MIMO systems with massive connectivity via random matrix theory, and provide comparisons. Furthermore, we conduct a numerical analysis on MMV based MMSE estimation for cell-free massive MIMO networks with massive connectivity and demonstrated the performance differences with that of SMV based MMSE estimation via numerical results.

4.1 System Model

We consider a cell-free massive MIMO network that consists of M APs and N single antenna users as depicted in Fig. 4.1. There is only one antenna at each AP. All the APs and users are uniformly distributed in a circular area with radius R_r, and all the APs are connected to the CPU through a backhaul network [128]. We consider sparse activity detection in the presence of massive connectivity, i.e., N is very large, and among all the users, only a small fraction of them are active at a given time instant. We assume the probability of each user being active is λ, and the user activities remain the same during each channel coherence interval.

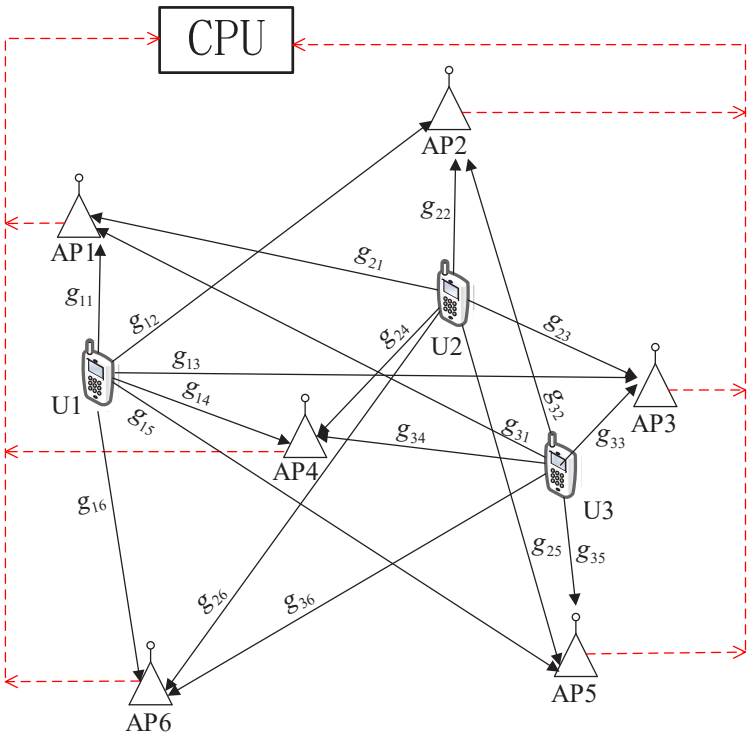

Figure 4.1: Cell-free massive MIMO network.

We assume flat fading channel between all the users and APs. The channel coefficient from the ith user to the jth AP is

$$g_{ij} = \beta_{ij}^{1/2} h_{ij} \tag{4.1}$$

where β_{ij} is the large-scale fading coefficient which can be expressed as $\beta_{ij} = \min\left(d_{ij}^{-\alpha}, d_0^{-\alpha}\right)$,

where d_{ij} is the distance from the ith user to the jth AP, d_0 is the reference distance, and α is the path loss decay exponent [65]. As all the APs and users are uniformly distributed in a circle with radius R_r, the probability density function of d_{ij} is

$$p(d_{ij}) = \frac{4d_{ij}}{\pi {R_r}^2} \left[\cos^{-1}\left(\frac{d_{ij}}{2R_r}\right) - \frac{d_{ij}}{2R_r}\left(1 - \frac{d_{ij}^2}{4R_r^2}\right)^{\frac{1}{2}} \right] \tag{4.2}$$

for $0 < d_{ij} < 2R_r$, and $p(d_{ij}) = 0$ elsewhere [129].

We denote the probability density function of β_{ij} by $p(\beta)$. Since β_{ij} changes slowly over time, we assume it is known at the central processing unit (CPU) and all APs. We denote by $\boldsymbol{\beta}_j \in \mathcal{R}^{N\times 1}$ the large-scale fading vector from all users to the jth AP. Hence, the ith element of $\boldsymbol{\beta}_j$ equals β_{ij}. We define $\boldsymbol{\Lambda}_j = diag(\boldsymbol{\beta}_j)$. Furthermore, $h_{ij} \sim \mathcal{CN}(0,1)$ is the corresponding small-sale fading coefficient from the ith user to the jth AP, and all the small-scale fading coefficients are i.i.d.. We use $\mathbf{g}_j \in \mathcal{R}^{N\times 1}$ and $\mathbf{h}_j \in \mathcal{R}^{N\times 1}$ to denote the channel coefficients and small-scale fading coefficients from all users to the jth AP, respectively. It is obvious that $\mathbf{g}_j = \boldsymbol{\Lambda}_j^{1/2}\mathbf{h}_j$ and all the elements of $\mathbf{h}_j$ are i.i.d.. Since the diagonal elements of $\boldsymbol{\Lambda}_j$ are known at the CPU and APs and they are not equal to each other, the elements of $\mathbf{g}_j$ are i.n.i.d..

We describe the activity of the ith user with binary-valued a_i, i.e., $a_i = 1$ if the ith user is active, while $a_i = 0$ if it is inactive. The probability of each user being active is λ, i.e., $P(a_i = 1) = \lambda$ and $P(a_i = 0) = 1 - \lambda$. We designate the combination of the ith user's activity and the channel coefficient from the ith user to the jth AP as the effective channel coefficient θ_{ij}, i.e., $\theta_{ij} = a_i g_{ij}$. Additionally, we denote by $b_{ij} = a_i h_{ij}$ the effective small-scale fading coefficient from the ith user to the jth AP. Similarly, we use the vectors $\boldsymbol{\theta}_j$ and $\mathbf{b}_j$ to denote the effective channel coefficients and effective small-scale fading coefficients from all users to the jth AP, respectively. It is obvious that $\boldsymbol{\theta}_j = \boldsymbol{\Lambda}_j\mathbf{b}_j$, and the non-zero elements of θ_j are i.n.i.d. while the non-zero elements of $\mathbf{b}_j$ are i.i.d..

In each channel coherence interval, active users will send pilots to APs first, and then

transmit messages. The pilots are used by APs to recognize the activity of the users. We assume that L symbols are used for pilot transmission during each channel coherence interval. The pilot matrix is denoted by an $L \times N$ complex matrix $\mathbf{\Phi}$, and each element of $\mathbf{\Phi}$ is independent and circularly symmetrically distributed with zero mean and variance $\frac{1}{L}$. Then, the received pilot signal at the jth AP is

$$\mathbf{y}_j = \mathbf{\Phi}\boldsymbol{\theta}_j + \mathbf{n}_j \tag{4.3a}$$

$$= \mathbf{\Phi}\mathbf{\Lambda}_j^{1/2}\mathbf{b}_j + \mathbf{n}_j \tag{4.3b}$$

where $\mathbf{n}_j \in \mathcal{CN}(0, \sigma_0^2\mathbf{I}_L)$ is the i.i.d. additive white Gaussian noise vector at the jth AP.

Finally, the aggregated received pilot signal of the entire set of APs can be expressed as

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{\Theta} + \mathbf{N} \tag{4.4}$$

where $\mathbf{\Theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_M]$ is the effective channel matrix from all users to the entire set of APs, and $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \cdots, \mathbf{n}_M]$ is the aggregated noise matrix at all APs.

Remark 4.1 *For facilitating the theoretical analysis, we consider SMV based MMSE estimation of $\boldsymbol{\theta}_j$ in (4.3a), although the estimation of $\mathbf{\Theta}$ in (4.4) is an MMV problem. The performance gap introduced by dividing the MMV problem into multiple SMV problems is shown via numerical results in Section 4.7.*

Remark 4.2 *Through Section 4.2 to Section 4.4, we consider SMV based MMSE estimation and activity detection with the received pilot signals at only one AP. For instance, if we consider the jth AP, then the large-scale fading coefficients, effective channel coefficients, effective small-scale fading coefficients, received pilot signals and noise are $\mathbf{\Lambda}_j$, $\boldsymbol{\theta}_j$, $\mathbf{b}_j$, $\mathbf{y}_j$ and $\mathbf{n}_j$, respectively. For the sake of notational brevity, we omit the subscript "j" in these symbols through Section 4.2 to Section 4.4 when there is no confusion.*

For the completeness of this chapter, in the next section, we briefly address preliminary

characterizations on SMV based postulated MMSE estimation and the CB-AMP algorithm, which will be used in our subsequent analysis.

4.2 Preliminaries

Several preliminary characterizations from prior work are provided in this section. First, we introduce the decoupling principle of SMV based postulated MMSE estimation of sparse signal vectors with i.i.d. non-zero components. Then, we describe the SMV based CB-AMP algorithm for MMSE estimation of sparse signals.

4.2.1 SMV Based Postulated MMSE Estimation

From Section II, we know that the elements of $\mathbf{b}$ are i.i.d., and the variance of the AWGN noise at AP is σ_0^2. We denote the probability density function of each element of $\mathbf{b}$ by $p_0(b)$. Now we can express

$$p_0(b) = (1-\lambda)\,\delta(b) + \lambda \mathcal{CN}(b; 0, 1) \tag{4.5}$$

where $\delta(\cdot)$ is the Dirac delta function, and $\mathcal{CN}(x_0; c_0, d_0)$ stands for the probability density function (at point x_0) of a circularly symmetric complex Gaussian random variable x with mean c_0 and variance d_0.

Suppose we are given the postulated prior distribution of each element of $\mathbf{b}$ as $p_{post}(b)$, and the postulated AWGN noise variance is σ_{post}^2. Then, the postulated MMSE (PMMSE) estimate of $\mathbf{b}$ is defined as

$$\widehat{\mathbf{b}}^{\,pmmse}(\mathbf{y}) = E\left\{\mathbf{b}|\mathbf{y}; p_{post}(b), \sigma_{post}^2\right\}. \tag{4.6}$$

Note that in the case in which $p_{post}(b) = p_0(b)$ and $\sigma_{post}^2 = \sigma_0^2$, $\widehat{\mathbf{b}}^{\,pmmse}(\mathbf{y})$ is the MMSE estimate of $\mathbf{b}$ given the received signal $\mathbf{y}$. Thus, postulated MMSE estimation is used in this chapter to aid the asymptotic analysis of SMV based MMSE estimation for sparse signal

vectors with both i.i.d. and i.n.i.d. non-zero components.

Now, let us consider a scalar Gaussian noise corrupted received signal

$$z = b + \sqrt{\eta} n \tag{4.7}$$

where $n \sim \mathcal{CN}(0,1)$. We define the conditional MMSE estimate of b as

$$\widehat{b}_{scalar}^{\,pmmse}\left(z; p_{post}(b), \eta\right) = E\left\{b | z; p_{post}(b), \eta\right\}. \tag{4.8}$$

Given two distributions, $p_{post}^1(b)$ and $p_{post}^2(b)$, and two noise levels, η_1 and η_2, we define

$$MSE\left(p_{post}^1(b), p_{post}^2(b), \eta_1, \eta_2\right) = \int \left| b - \widehat{b}_{scalar}^{\,pmmse}\left(z; p_{post}^1(b), \eta_1\right)\right|^2 p\left(b|z; p_{post}^2(b), \eta_2\right) db. \tag{4.9}$$

Equation (4.9) shows the mean square error of postulated MMSE estimation for signal distribution $p_{post}^2(b)$ and noise variance η_2 under postulated signal distribution $p_{post}^1(b)$ and postulated noise variance η_1. When $p_{post}^1(b) = p_{post}^2(b) = p_0(b)$ and $\eta_1 = \eta_2 = \sigma_0^2$, equation (4.9) gives the mean square error of MMSE estimation for $\mathbf{b}$. Then, we can obtain the decoupling principle of replica symmetric PMMSE estimation for sparse signal vectors with i.i.d. non-zero components. We present this characterization as a property below.

Property 4.1 *[Decoupling Principle of Replica Symmetric SMV based MMSE Estimation for Sparse Signal Vectors with i.i.d. non-zero components] [61, 127]: Consider the linear model in (3b). Assume that the number of pilots used for estimation, L, changes with the number of users, N (using $L(N)$ to denote the relationship between L and N), and $\lim_{N\to\infty} \frac{N}{L(N)} = \gamma$. Let $\widehat{\mathbf{b}}^{\,pmmse}(\mathbf{y})$ be the MMSE estimate of $\mathbf{b}$ based on the prior distribution $p_0(b)$ and noise variance σ_0^2. Then, under replica symmetry, there exist effective noise levels σ_{eff}^2 and $\sigma_{p\text{-}eff}^2$ such that:*
(1) As $N \to \infty$, the random vector $(b_k, \beta_k, \widehat{b}_k^{\,pmmse})$ converges in distribution to the random vector $(b, \beta, \widehat{b})$. Here, b_k and $\widehat{b}_k^{\,pmmse}$ are the kth elements of $\mathbf{b}$ and $\widehat{\mathbf{b}}^{\,pmmse}(\mathbf{y})$, respectively.

β_k is the kth diagonal element of $\mathbf{\Lambda}$. b, β and n are independent with distributions $b \sim p_0(b)$, $\beta \sim p(\beta)$, $n \sim \mathcal{CN}(0,1)$, and

$$z^{mmse} = b + \sqrt{\eta} n, \tag{4.10a}$$

$$\widehat{b} = \widehat{b}_{scalar}^{\,pmmse}\left(z^{mmse}; p_0(b), \eta_p\right) \tag{4.10b}$$

where $\eta = \frac{\sigma_{eff}^2}{\beta}$ and $\eta_p = \frac{\sigma_{p\text{-}eff}^2}{\beta}$.

(2) The effective noise levels satisfy the equations

$$\sigma_{eff}^2 = \sigma_0^2 + \gamma E\left\{\beta MSE(p_0(b), p_0(b), \eta_p, \eta, z)\right\}, \tag{4.11a}$$

$$\sigma_{p\text{-}eff}^2 = \sigma_0^2 + \gamma E\left\{\beta MSE(p_0(b), p_0(b), \eta_p, \eta_p, z)\right\} \tag{4.11b}$$

where the expectations are taken over $\beta \sim p(\beta)$ and z is generated by (4.10a).

Property 4.1 shows that the MMSE estimation of the SMV problem in (4.3b) can be decoupled into scalar Gaussian noise corrupted linear problems as described in (4.10a), and the corresponding noise variance in (4.10a) can be obtained by solving the fixed point equations in (4.11). Then, we can readily find the MMSE estimates in (4.10b) and the corresponding mean square error.

Remark 4.3 *Since we consider SMV based MMSE estimation in this chapter, the postulated prior of $\mathbf{b}$ and the postulated noise level used in Property 4.1 are the prior distribution $p_0(b)$ and noise variance σ_0^2, respectively.*

Remark 4.4 *In general, there may exist multiple solutions for the fixed point equations in (4.11). In this case, the true solution is the minimizer of a certain free energy function as described in [61].*

4.2.2 SMV Based CB-AMP Algorithm

The CB-AMP algorithm is used during numerical simulations to obtain the SMV based MMSE estimates of the sparse effective channel coefficients in cell-free massive MIMO networks. For the sake of completeness in this chapter, we briefly introduce the CB-AMP algorithm proposed in [57]. We summarize the recursions of the CB-AMP algorithm in Algorithm 4.1 below.

Algorithm 4.1 SMV based CB-AMP algorithm [57].

For the linear model in (4.3a) with a given pilot matrix $\mathbf{\Phi}$, received signal $\mathbf{y}$, and the prior probability function $p(\boldsymbol{\theta})$, the CB-AMP algorithm generates a sequence of estimates $\widehat{\boldsymbol{\theta}}^t$, $\widehat{\mathbf{r}}^t$, for $t = 1, 2, \ldots$ through the following recursions:

(1) Initialization: Set $t = 1$, $\widehat{\boldsymbol{\theta}}^1 = \overline{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\kappa}}^1 = \mathrm{var}(\boldsymbol{\theta})$, $\mathbf{z}^0 = \mathbf{1}_L$, $\mathbf{p} = \mathbf{y}$, where $\overline{\boldsymbol{\theta}}$ and $\mathrm{var}(\boldsymbol{\theta})$ are the corresponding mean and variance for each element of $\boldsymbol{\theta}$, and $\mathbf{1}_M$ is an $1 \times M$ vector whose elements are 1.
(2) For each $j \in [L]$ (where $[L]$ stands for the set which includes the integers from 1 to L), calculate

$$z_j^t = \sum_i |\phi_{ji}|^2 \widehat{\kappa}_i^t, \tag{4.12a}$$

$$p_j^t = \sum_i \phi_{ji} \widehat{\theta}_i^t - \frac{z_j^t}{\sigma_0^2 + z_j^{t-1}} \left(y_a - p_a^{t-1}\right) \tag{4.12b}$$

where ϕ_{ji} is the corresponding element on the jth row and ith column of $\mathbf{\Phi}$.
(3) For each $i \in [N]$, calculate

$$\tau_i^t = \left[\sum_j \frac{|\phi_{ji}|^2}{\sigma_0^2 + z_j^t}\right]^{-1}, \tag{4.13a}$$

$$\widehat{r}_i^t = \widehat{\theta}_i^t + \tau_i^t \sum_j \frac{\phi_{ji}^* \left(y_j - p_j^t\right)}{\sigma_0^2 + z_j^t}, \tag{4.13b}$$

$$\widehat{\theta}_i^{t+1} = g_{in}\left(\widehat{r}_i^t, \tau_i^t\right), \tag{4.13c}$$

$$\widehat{\kappa}_i^{t+1} = \varpi_{in}\left(\widehat{r}_i^t, \tau_i^t\right). \tag{4.13d}$$

Then let $t \leftarrow t+1$, and return to step (2) and repeat until a sufficient number of iterations have been performed or a given termination condition is satisfied.

Considering the AWGN output channel and the recursions of the CB-AMP algorithm,

the estimated signal is decoupled into scalar AWGN Gaussian noise corrupted signals, i.e.,

$$\widehat{r}_i^t = \theta_i + \sqrt{\xi^t} n \tag{4.14}$$

where ξ^t is the variance of the corrupting noise, and it satisfies the following state evolution equation [57]:

$$\xi^{t+1} = \sigma_0^2 + \gamma E\left\{\left|\theta_i - \widehat{\theta}_i^t\right|^2\right\}. \tag{4.15}$$

The MMSE estimates of $\widehat{\theta}_i^t$ given $\widehat{r}_i^t$ for the linear model in (4.14) is [64]

$$\begin{aligned}\widehat{\theta}_i^{t+1}\left(\widehat{r}_i^t; \xi^t, \lambda, \beta_i\right) &= E\left\{\theta | \widehat{r}_i^t; \xi^t, \lambda, \beta_i\right\} \\ &= G\left(\left|\widehat{r}_i^t\right|^2; \xi^t, \lambda, \beta_i\right) \frac{\beta_i}{\beta_i + \xi^t} \widehat{r}_i^t\end{aligned} \tag{4.16}$$

where

$$G\left(\left|\widehat{r}_i^t\right|^2; \xi^t, \lambda, \beta_i\right) = \frac{1}{1 + \frac{(1-\lambda)(\beta_i + \xi^t)}{\lambda \xi^t} \exp\left(-\frac{\beta_i |\widehat{r}_i^t|^2}{\xi^t(\beta_i + \xi^t)}\right)}.$$

Then we can obtain the MMSE estimate of $\widehat{\theta}_i^{t+1}$ as

$$g_{in}(\widehat{r}_i^t, \tau_i^t) = \widehat{\theta}_i^{t+1}\left(\widehat{r}_i^t; \tau_i^t, \lambda, \beta_i\right) \tag{4.17}$$

where τ_i^t provided in (4.13a) is the estimated value of ξ^t. Finally, the estimated variance of $\widehat{\theta}_i^{t+1}$ can be expressed as

$$\begin{aligned}\varpi_{in}(\widehat{r}_i^t, \tau_i^t) &= \tau_i^t \frac{\partial}{\partial \widehat{r}_i^t} g_{in}\left(\widehat{r}_i^t, \tau_i^t\right) \\ &= \beta_i G\left[\frac{\beta_i(\tau_i^t + |\widehat{r}_i^t|^2) + |\tau_i^t|^2}{(\beta_i + \tau_i^t)^2} - G\frac{\beta_i |\widehat{r}_i^t|^2}{(\beta_i + \tau_i^t)^2}\right]\end{aligned} \tag{4.18}$$

where G stands for the function $G\left(\left|\widehat{r}_i^t\right|^2; \tau_i^t, \lambda, \beta_i\right)$.

4.3 SMV Based MMSE Estimation in Cell-Free Massive MIMO Networks with Massive Connectivity

In this chapter, we focus on the asymptotic analysis of SMV based MMSE estimation for sparse signal vectors with i.n.i.d. non-zero components, which is the case for the linear model in (4.3a). Since it is challenging to find closed-form results, CB-AMP is an efficient algorithm to acquire the MMSE estimate of $\boldsymbol{\theta}_j$ numerically. However, to the best of our knowledge, there is no previous work which has focused on the asymptotic analysis of this setting. Several previous studies in the literature addressed the asymptotic analysis of SMV based MMSE estimation for sparse signal vectors with i.i.d. non-zero components. In order to use these results, we need to convert the problem in (4.3a) into (4.3b), in which case the unknown sparse vector $\mathbf{b_j}$ has i.i.d. non-zero components. On the other hand, in the same situation with (4.3a), it is challenging for us to find a closed-form MMSE estimate of $\mathbf{b_j}$ in (4.3b), and CB-AMP is an efficient algorithm to be considered. However, CB-AMP has strict requirements on the structure of the measurement matrix to ensure that the algorithm works [57, 130]. For a given pilot matrix $\boldsymbol{\Phi}$, the equivalent measurement matrix in (4.3b) is $\boldsymbol{\Phi}\boldsymbol{\Lambda}_j$ as we consider only $\mathbf{y}_j$ for SMV based MMSE estimation when seeking to estimate $\mathbf{b}_j$. Since the elements of $\boldsymbol{\Lambda}_j$ are random variables, it is difficult to control the structure of $\boldsymbol{\Phi}\boldsymbol{\Lambda}_j$ to meet the measurement matrix requirements of the CB-AMP algorithm. Thus, it is challenging to find the MMSE estimate of $\mathbf{b_j}$ in (4.3b) numerically with the CB-AMP algorithm. In summary, we can not perform asymptotic and numerical analysis with the same linear model, i.e., linear model in (4.3a) is suitable for numerical analysis but challenging for asymptotic analysis, while (4.3b) is challenging for numerical analyze but suitable for asymptotic analysis.

In this section, by considering the relationship between the MMSE estimate of unknown vectors in (4.3a) and (4.3b), we arrive at a decoupling principle as represented in Property 4.2, with which we can perform asymptotic analysis on the MMSE estimate of $\boldsymbol{\theta}_j$ in (4.3a).

Then, the asymptotic and numerical analyses can be performed within the same linear model (4.3a), facilitating the performance analysis on SMV based MMSE estimation for sparse signal vectors with i.n.i.d. non-zero components (the case of cell-free massive MIMO being included), and also the likelihood ratio test in the following sections.

Initially, to provide a benchmark, we analyze the "oracle estimation" with which the MMSE estimation is performed using the linear model (4.3a) under the idealistic assumption that the support of the users (i.e., their being active or inactive) is known at the APs. With the prior information on user support, it is obvious that the oracle estimation leads to a lower bound on the mean square error of MMSE estimation using the linear model (4.3a) with unknown user support. Additionally, the asymptotic analyze of oracle estimation is performed with random matrix theory.

In order for the simplicity of descriptions, we define several notations first. Let us denote the user activity support of $\boldsymbol{\theta}$ as $\mathbf{s}^* = [a_1, a_2, \cdots, a_N]^T$. For a given user activity support $\mathbf{s} \in \{0,1\}^N$ and its corresponding index set of non-zero elements, $I_{\mathbf{s}}$ ($i \in I_{\mathbf{s}}$ if $i \leqslant N$ and $s_i = 1$), we define $\boldsymbol{\Phi}_{\mathbf{s}}$ and $\boldsymbol{\Lambda}_{\mathbf{s}}$ as matrices consisting of the columns of $\boldsymbol{\Phi}$ and $\boldsymbol{\Lambda}$ on the index set $I_{\mathbf{s}}$, respectively. Similarly, $\boldsymbol{\theta}_{\mathbf{s}}$ is comprised of the entries on the index set $I_{\mathbf{s}}$.

4.3.1 Oracle Estimation

Assume $\mathbf{s}^*$ is known at all APs. Then,

$$p(\mathbf{y}|\boldsymbol{\Phi}_{\mathbf{s}^*}, \boldsymbol{\theta}_{\mathbf{s}^*}) = \frac{1}{\pi^L \sigma_0^{2L}} \exp\left(-\frac{\|\mathbf{y} - \boldsymbol{\Phi}_{\mathbf{s}^*}\boldsymbol{\theta}_{\mathbf{s}^*}\|^2}{\sigma_0^2}\right) \tag{4.19}$$

and

$$p(\boldsymbol{\theta}_{\mathbf{s}^*}) = \frac{1}{\pi^{|\mathbf{s}^*|}\,|\boldsymbol{\Lambda}_{\mathbf{s}^*}|} \exp\left(-\boldsymbol{\theta}_{\mathbf{s}^*}^H \boldsymbol{\Lambda}_{\mathbf{s}^*}^{-1} \boldsymbol{\theta}_{\mathbf{s}^*}\right) \tag{4.20}$$

where $|\mathbf{s}^*|$ stands for the cardinality of $\mathbf{s}^*$, and $|\mathbf{\Lambda}_{\mathbf{s}^*}|$ denotes the determinant of $\mathbf{\Lambda}_{\mathbf{s}^*}$. Therefore, the oracle estimate is

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{oracle} &= \int \boldsymbol{\theta}_{\mathbf{s}^*} p(\boldsymbol{\theta}_{\mathbf{s}^*}|\mathbf{\Phi}_{\mathbf{s}^*}, \mathbf{y}) d\boldsymbol{\theta}_{\mathbf{s}^*} \\ &= \left(\mathbf{\Phi}_{\mathbf{s}^*}^H \mathbf{\Phi}_{\mathbf{s}^*} + \sigma_0^2 \mathbf{\Lambda}_{\mathbf{s}^*}^{-1}\right)^{-1} \mathbf{\Phi}_{\mathbf{s}^*}^H \mathbf{y} \end{aligned} \tag{4.21}$$

and the corresponding mean square error is

$$MSE_{oracle} = \frac{1}{N} tr\left(\left(\frac{1}{\sigma_0^2}\mathbf{\Phi}_{\mathbf{s}^*}^H \mathbf{\Phi}_{\mathbf{s}^*} + \mathbf{\Lambda}_{\mathbf{s}^*}^{-1}\right)^{-1}\right) \tag{4.22}$$

where $tr(\cdot)$ is the trace operator. Now, let us consider the asymptotic behavior of the mean square error of the oracle estimate, i.e., determine $MSE_{m^*} = \lim_{N\to\infty} MSE_{oracle}$, where $m^* = \lim_{N\to\infty} \frac{|\mathbf{s}^*|}{N}$. Recall that the elements of $\mathbf{\Phi}_{\mathbf{s}^*}$ are i.i.d. with zero mean and variance $1/L$. Then we can obtain

$$\mathcal{R}_{\frac{1}{\sigma_0^2}\mathbf{\Phi}_{\mathbf{s}^*}^H \mathbf{\Phi}_{\mathbf{s}^*}}(z) = \frac{1}{\sigma_0^2 - m^*\gamma z} \tag{4.23}$$

where $\mathcal{R}_{\mathbf{X}}(\cdot)$ is the R-transform of random matrix $\mathbf{X}$ [131].

Via random matrix theory, the η-transform of $\mathbf{\Lambda}_{\mathbf{s}^*}^{-1}$ is

$$\eta_{\mathbf{\Lambda}_{\mathbf{s}^*}^{-1}}(\varsigma) = E\left\{\frac{\beta}{\beta + \varsigma}\right\} \tag{4.24}$$

where β is a random variable that has the same distribution as the diagonal elements of $\mathbf{\Lambda}_{\mathbf{s}^*}$, and $\varsigma > 0$ [131]. Using the relationship between R-transform and η-transform, we can obtain

$$\mathcal{R}_{\mathbf{\Lambda}_{\mathbf{s}^*}^{-1}}(z) = -\frac{1}{\varsigma} - \frac{1}{z} \tag{4.25}$$

where

$$z = -\varsigma \eta_{\mathbf{\Lambda}_{\mathbf{s}^*}^{-1}}(\varsigma). \tag{4.26}$$

Let us define $\mathbf{F}^{\mathbf{s}^*} = \frac{1}{\sigma_0^2}\mathbf{\Phi}_{\mathbf{s}^*}^H\mathbf{\Phi}_{\mathbf{s}^*} + \mathbf{\Lambda}_{\mathbf{s}^*}^{-1}$. Since $\frac{1}{\sigma_0^2}\mathbf{\Phi}_{\mathbf{s}^*}^H\mathbf{\Phi}_{\mathbf{s}^*}$ is unitarily invariant and $\mathbf{\Lambda}_{\mathbf{s}^*}^{-1}$ is a deterministic matrix with bounded eigenvalues, $\frac{1}{\sigma_0^2}\mathbf{\Phi}_{\mathbf{s}^*}^H\mathbf{\Phi}_{\mathbf{s}^*}$ and $\mathbf{\Lambda}_{\mathbf{s}^*}^{-1}$ are asymptotically free. Thus,

$$\begin{aligned}\mathcal{R}_{\mathbf{F}^{\mathbf{s}^*}}(z) &= \mathcal{R}_{\frac{1}{\sigma_0^2}\mathbf{\Phi}_{\mathbf{s}^*}^H\mathbf{\Phi}_{\mathbf{s}^*}}(z) + \mathcal{R}_{\mathbf{\Lambda}_{\mathbf{s}^*}^{-1}}(z) \\ &= \frac{1}{\sigma_0^2 - m^*\gamma z} - \frac{1}{\varsigma} - \frac{1}{z}. \end{aligned} \tag{4.27}$$

Therefore,

$$\begin{aligned}\mathcal{S}_{\mathbf{F}^{\mathbf{s}^*}}^{-1}(-z) &= \mathcal{R}_{\mathbf{F}^{\mathbf{s}^*}}(z) + \frac{1}{z} \\ &= \frac{1}{\sigma_0^2 - m^*\gamma z} - \frac{1}{\varsigma} \end{aligned} \tag{4.28}$$

where $\mathcal{S}_{\mathbf{X}}^{-1}(\cdot)$ stands for the inverse of Stieltjes transform of the random matrix $\mathbf{X}$. Setting $\mathcal{S}_{\mathbf{F}^{\mathbf{s}^*}}^{-1}(-z_0) = 0$, we can obtain

$$E\left\{\frac{\beta}{\beta + \varsigma_0}\right\} = \frac{\varsigma_0 - \sigma_0^2}{m^*\gamma\varsigma_0}. \tag{4.29}$$

Then, with the definition of Stieltjes transform, we have

$$\begin{aligned}\frac{1}{m^*}MSE_{m^*} &= \lim_{|\mathbf{s}^*|\to\infty}\frac{1}{|\mathbf{s}^*|}tr\left(\left(\frac{1}{\sigma_0^2}\mathbf{\Phi}_{\mathbf{s}^*}^H\mathbf{\Phi}_{\mathbf{s}^*} + \mathbf{\Lambda}_{\mathbf{s}^*}^{-1}\right)^{-1}\right) \\ &= \mathcal{S}_{\mathbf{F}^{\mathbf{s}^*}}(0) \\ &= \mathcal{S}_{\mathbf{F}^{\mathbf{s}^*}}\left(\mathcal{S}_{\mathbf{F}^{\mathbf{s}^*}}^{-1}(-z_0)\right) \\ &= -z_0 \\ &= \varsigma_0\eta_{\mathbf{\Lambda}_{\mathbf{s}^*}^{-1}}(\varsigma_0) \\ &= \frac{\varsigma_0 - \sigma_0^2}{m^*\gamma} \end{aligned} \tag{4.30}$$

where $\mathcal{S}_{\mathbf{X}}(\cdot)$ is the Stieltjes transform of random matrix $\mathbf{X}$, and ς_0 can be obtained via the

fixed point equation (4.29).

In this chapter, we consider the scenario in which the users are active independently with the same probability λ. Therefore, $\lambda = \lim_{N\to\infty} \frac{|\mathbf{s}^*|}{N} = m^*$. Thus, the asymptotic mean square error of the oracle estimation is

$$\begin{aligned} MSE_{oracle,A} &= MSE_{\lambda} \\ &= \frac{\varsigma^* - \sigma_0^2}{\gamma} \end{aligned} \tag{4.31}$$

where ς^* is the solution of the fixed point equation

$$E\left\{\frac{\beta}{\beta+\varsigma^*}\right\} = \frac{\varsigma^* - \sigma_0^2}{\lambda\gamma\varsigma^*}.$$

4.3.2 SMV Based MMSE Estimation

In this subsection, we focus on the asymptotic analysis of SMV based MMSE estimation for sparse signal vectors with i.n.i.d. non-zero components, which is the statistical description of linear model (4.3a). By comparing the formulas for the MMSE estimate of unknown vectors in (4.3a) and (4.3b), we find a closed-form relationship between the MMSE estimates of $\boldsymbol{\theta}$ and $\mathbf{b}$. Inspired by the decoupling principle in Property 4.1 for (4.3b), we propose a similar decoupling principle for the linear model (4.3a) as represented in Property 4.2. Finally, we arrive at two different methods to compute the effective noise level σ_{eff} for the MMSE estimates of linear model (4.3a).

The SMV based MMSE estimator for the unknown vector in linear model (4.3a) is [79]

$$\widehat{\boldsymbol{\theta}}_{mmse} = \sum\nolimits_{\mathbf{s}\in\{0,1\}^N} p(\mathbf{s}|\mathbf{y})\widehat{\boldsymbol{\theta}}_{\mathbf{s}} \tag{4.32}$$

where $p(\mathbf{s}|\mathbf{y})$ is the conditional probability that the user's activity support equals $\mathbf{s}$, and

$$\widehat{\boldsymbol{\theta}}_{\mathbf{s}} = \left(\boldsymbol{\Phi}_{\mathbf{s}}^H\boldsymbol{\Phi}_{\mathbf{s}} + \sigma_0^2\boldsymbol{\Lambda}_{\mathbf{s}}^{-1}\right)^{-1}\boldsymbol{\Phi}_{\mathbf{s}}^H\mathbf{y}. \tag{4.33}$$

Similarly, the MMSE estimator for the unknown vector in linear model (4.3b) is

$$\widehat{\mathbf{b}}_{mmse} = \sum\nolimits_{\mathbf{s}\in\{0,1\}^N} p(\mathbf{s}|\mathbf{y})\widehat{\mathbf{b}}_{\mathbf{s}} \tag{4.34}$$

and

$$\begin{aligned}\widehat{\mathbf{b}}_{\mathbf{s}} &= \left(\boldsymbol{\Lambda}_{\mathbf{s}}^{1/2}\boldsymbol{\Phi}_{\mathbf{s}}^H\boldsymbol{\Phi}_{\mathbf{s}}\boldsymbol{\Lambda}_{\mathbf{s}}^{1/2} + \sigma_0^2\mathbf{I}_{|\mathbf{s}|}\right)^{-1}\boldsymbol{\Lambda}_{\mathbf{s}}^{1/2}\boldsymbol{\Phi}_{\mathbf{s}}^H\mathbf{y} \\ &= \boldsymbol{\Lambda}_{\mathbf{s}}^{-1/2}\left(\boldsymbol{\Phi}_{\mathbf{s}}^H\boldsymbol{\Phi}_{\mathbf{s}} + \sigma_0^2\boldsymbol{\Lambda}_{\mathbf{s}}^{-1}\right)^{-1}\boldsymbol{\Phi}_{\mathbf{s}}^H\mathbf{y}. \end{aligned} \tag{4.35}$$

For a given received signal $\mathbf{y}$, we have the same $p(\mathbf{s}|\mathbf{y})$ in (4.32) and (4.34), and $\widehat{\boldsymbol{\theta}}_{\mathbf{s}} = \boldsymbol{\Lambda}_{\mathbf{s}}^{1/2}\widehat{\mathbf{b}}_{\mathbf{s}}$. Therefore,

$$\widehat{\boldsymbol{\theta}}_{mmse} = \boldsymbol{\Lambda}^{1/2}\widehat{\mathbf{b}}_{mmse}. \tag{4.36}$$

Decoupling principle in Property 4.1 indicates that the MMSE estimate for linear model (4.3b) can be decomposed into Gaussian noise corrupted scalar models described in (4.10a). As a result of the relationship in (4.36), it is natural for us to consider using $z_1^{mmse} = \sqrt{\beta}z^{mmse}$ to describe the decoupling property of the MMSE estimate for linear model (4.3a), i.e.,

$$z_1^{mmse} = \theta + \sigma_{\mathit{eff}}n. \tag{4.37}$$

Now, let us try to verify the assumption of (4.37) via the MMSE estimates of the scalar linear models in (4.10a) and (4.37). With the result in (4.16), it is easy for us to obtain that

$$\widehat{b}_{mmse} = G\left(|z^{mmse}|^2;\eta,\lambda,1\right)\frac{z^{mmse}}{1+\eta}, \tag{4.38}$$

$$\widehat{\theta}_{mmse} = G\left(|z_1^{mmse}|^2;\sigma_{\mathit{eff}}^2,\lambda,\beta\right)\frac{\beta z_1^{mmse}}{\beta+\sigma_{\mathit{eff}}^2}, \tag{4.39}$$

and

$$G\left(|z_1^{mmse}|^2;\sigma_{\mathit{eff}}^2,\lambda,\beta\right) = G\left(|z|^2;\eta,\lambda,1\right) \tag{4.40}$$

when $z_1^{mmse} = \sqrt{\beta} z^{mmse}$. Therefore,

$$\widehat{\theta}_{mmse} = \sqrt{\beta}\widehat{b}_{mmse}, \tag{4.41}$$

which is consistent with the relationship in (4.36) and verifies our assumption of using (4.37) to describe the decoupling property of the MMSE estimate for linear model (4.3a).

Combining (4.36), (4.37) and Property 1, we conclude that the MMSE estimator of linear model (4.3a) breaks down the received signal into Gaussian noise corrupted versions as described in (4.37), and the parameter σ_{eff} satisfies the fixed point equations in (4.11). We call this property as the decoupling principle of the MMSE estimation of linear model (4.3a) and summarize it in Property 4.2 as follows:

Property 4.2 *[Decoupling Principle of Replica Symmetric SMV based MMSE Estimation for Sparse Signal Vectors with i.n.i.d. non-zero Components]: Consider the linear model in (4.3a), where* $\boldsymbol{\theta} = \boldsymbol{\Lambda}\mathbf{b}$, $\boldsymbol{\Lambda}$ *is a diagonal matrix whose diagonal elements (denoted by the random variable* β*) are i.i.d. distributed with distribution* $\beta \sim p(\beta)$*. The diagonal elements of* $\boldsymbol{\Lambda}$ *are realizations from the random distribution* $p(\beta)$ *and we assume that they are known for further signal processing. The elements of* $\mathbf{b}$ *are i.i.d. distributed with distribution* $b \sim p_0(b)$*. As a consequence of knowing the diagonal elements of* $\boldsymbol{\Lambda}$*, the elements of* $\boldsymbol{\theta}$ *are i.n.i.d.. Assume that the number of pilots used for estimation,* L*, and the number of users,* N *satisfy the conditions as described in Property 4.1. Let* $\widehat{\boldsymbol{\theta}}^{mmse}(\mathbf{y})$ *be the MMSE estimate of* $\boldsymbol{\theta}$*. Then, under replica symmetry, as* $N \to \infty$*, the random vector* $(\theta_k, \beta_k, \widehat{\theta}_k^{mmse})$ *converges in distribution to the random vector* $(\theta, \beta, \widehat{\theta})$*. Here,* θ_k *and* $\widehat{\theta}_k^{mmse}$ *are the* k*th elements of* $\boldsymbol{\theta}$ *and* $\widehat{\boldsymbol{\theta}}^{mmse}(\mathbf{y})$*, respectively.* β_k *is the* k*th diagonal element of* $\boldsymbol{\Lambda}$*.* b*,* β *and* n *are independent with distributions* $b \sim p_0(b)$*,* $\beta \sim p(\beta)$*,* $n \sim \mathcal{CN}(0,1)$*. There exist effective noise levels* σ_{eff}^2

and $\sigma^2_{p\text{-}eff}$ *such that:*

$$z_1^{mmse} = \theta + \sigma_{eff} n, \tag{4.42a}$$

$$\widehat{\theta} = \sqrt{\beta}\widehat{b}_{scalar}^{\,pmmse}\left(z_1^{mmse}; p_0(b), \eta_p\right), \tag{4.42b}$$

where $\eta_p = \frac{\sigma^2_{p\text{-}eff}}{\beta}$. *The effective noise levels* σ^2_{eff} *and* $\sigma^2_{p\text{-}eff}$ *satisfy the fixed point equations in (4.11).*

Proof 4.1 *From the results in Property 4.1, we can acquire that as* $N \to \infty$, *the random vector* $(\frac{\theta_k}{\sqrt{\beta_k}}, \beta_k, \widehat{b}_k^{\,pmmse})$ *converges in distribution to the random vector* $(\frac{\theta}{\sqrt{\beta}}, \beta, \widehat{b})$, *and the scalar Gaussian noise corrupted* θ *can be expressed as* $z^{mmse} = \frac{\theta}{\sqrt{\beta}} + \sqrt{\eta} n$. *As noted above, the MMSE estimates* $\widehat{\theta}_{mmse}$ *and* $\widehat{b}_{mmse}$ *have the relationship* $\widehat{\theta}_{mmse} = \sqrt{\beta}\widehat{b}_{mmse}$ *when the scalar versions of Gaussian noise corrupted* θ *and* b, *which are denoted as* z_1^{mmse} *and* z^{mmse}, *have the relationship* $z_1^{mmse} = \sqrt{\beta} z^{mmse}$. *Therefore, when the decoupled scalar Gaussian noise corrupted* θ *is expressed as* $z_1^{mmse} = \sqrt{\beta} z^{mmse} = \theta + \sigma_{eff} n$, *the random vector* $(\theta_k, \beta_k, \widehat{\theta}_k^{mmse})$ *converges in distribution to the random vector* $(\theta, \beta, \sqrt{\beta}\widehat{b})$ *as* $N \to \infty$, *which has the same distribution as that of the random vector* $(\theta, \beta, \widehat{\theta})$. *Hence, we have convergence in distribution to* $(\theta, \beta, \widehat{\theta})$, *and Property 4.2 follows from Property 4.1. For details on the proof of Property 4.1, we refer to [61].*

Similar to Property 4.1, Property 4.2 shows that the MMSE estimation of the SMV problem in (4.3a) can be decoupled into scalar Gaussian noise corrupted linear problems as represented in (4.42a), and the corresponding noise variance in (4.42a) can be obtained by solving the fixed point equations in (4.11). Then it becomes obvious to find the MMSE estimates in (4.42b) and the corresponding mean square error.

Now, we seek to find the value of σ_{eff}. Note that we can determine σ_{eff} by solving the fixed point equations in (4.11). We refer to this method as *Property1-*σ_{eff}. The PMMSE

estimator in (4.10b) is

$$\widehat{b}_{scalar}^{\,pmmse}(z;p_0(b),\eta_p) = G\left(|z|^2;\eta_p,\lambda,1\right)\frac{z}{1+\eta_p}. \tag{4.43}$$

Then, with (4.9), the mean square error in (4.11b) is [64]

$$\begin{aligned} MSE\left(p_0(b),p_0(b),\eta_p,\eta_p,z\right) =& \int \left|b-\widehat{b}_{scalar}^{\,pmmse}(z;p_0(b),\eta_p)\right|^2 p\left(b|z;p_0(b),\eta_p\right)db \\ =& \lambda\left[1-\frac{\eta_p^2}{1+\eta_p}\omega\left(\frac{(1+\eta_p)(1-\lambda)}{\lambda\eta_p},\eta_p\right)\right] \end{aligned} \tag{4.44}$$

where

$$\omega(a,b) = \int_0^\infty \frac{te^{-bt}}{1+ae^{-t}}dt.$$

Following the same procedure as in [64], we can obtain the mean square error in (4.11a) as

$$\begin{aligned} MSE\left(p_0(b),p_0(b),\eta_p,\eta,z\right) =& \int \left|b-\widehat{b}_{scalar}^{\,pmmse}(z;p_0(b),\eta_p)\right|^2 p\left(b|z;p_0(b),\eta\right)db \\ =& \lambda\left[1-\frac{2\eta_p^2(1+\eta_p)}{(1+\eta)^2}\omega\left(\overline{a},\overline{b}\right)+\frac{\eta_p^2}{1+\eta}\omega_2(\overline{a},\overline{b},\overline{c},\overline{d})\right] \end{aligned} \tag{4.45}$$

where

$$\omega_2(a,b,c,d) = \int_0^\infty \frac{te^{-bt}\left(1-ce^{-dt}\right)}{(1+ae^{-t})^2}dt,$$

$$\overline{a} = \frac{(1+\eta_p)(1-\lambda)}{\lambda\eta_p},$$

$$\overline{b} = \frac{\eta_p(1+\eta_p)}{1+\eta},$$

$$\overline{c} = \frac{(1+\eta)(1-\lambda)}{\lambda\eta},$$

$$\overline{d} = \overline{b}/\eta.$$

Then, we determine σ_{eff} by substituting (4.44), (4.45) and the distribution of β into the fixed point equations in (4.11).

With the CB-AMP algorithm and also the theoretical result in (4.42a), we have another approach to obtain σ_{eff}, specifically by using the state evolution equation (4.15). When the CB-AMP algorithm converges, the noise variance in (4.14) should satisfy $\xi^{t+1} = \xi^t$. Therefore, σ_{eff}^2 satisfies the fixed point equation

$$\sigma_{\mathit{eff}}^2 = \sigma_0^2 + \gamma E\left\{\left|\theta - \widehat{\theta}_{mmse}\right|^2\right\} \tag{4.46}$$

where $\widehat{\theta}_{mmse}$ is given in (4.39). Therefore, we can obtain σ_{eff} by solving the fixed point equation (4.46). We call this method as *State-*σ_{eff}.

The mean square error of the MMSE estimate for the linear model (4.37) is

$$E\left\{\left|\theta - \widehat{\theta}_{mmse}\right|^2\right\} = \lambda\left(E\{\beta\} - \int_{\beta_{\min}}^{\beta_{\max}}\int_0^{\infty} f(\sigma_{\mathit{eff}}, \beta, t)dtd\beta\right) \tag{4.47}$$

where

$$f(\sigma_{\mathit{eff}}, \beta, t) = \frac{\lambda t p(\beta) e^{-t\sigma_{\mathit{eff}}^2/\beta}}{\left(\frac{\beta}{\sigma_{\mathit{eff}}^2}+1\right)^2\left(\frac{\lambda}{\beta+\sigma_{\mathit{eff}}^2}+\frac{(1-\lambda)e^{-t}}{\sigma_{\mathit{eff}}^2}\right)},$$

$\beta_{\min} = (2R)^{-\alpha}$, and $\beta_{\max} = d_0^{-\alpha}$.

By now, we have introduced the *Property1-*σ_{eff} and *State-*σ_{eff} method to compute the effective noise level σ_{eff} in (4.42a). Although derived with the state evolution equations of the CB-AMP algorithm, the set up of *State-*σ_{eff} method is conditioned on the theoretical result in (4.42a) which is introduced by the decoupling principle. In other words, decoupling principle in Property 4.2 provides the theoretical basis for the asymptotic analysis of SMV based MMSE estimation for sparse signal vectors with i.n.i.d. non-zero components in this chapter, including both *Property1-*σ_{eff} and *State-*σ_{eff} method, and the likelihood ratio test introduced in the following sections.

Since the mean square error $E\{|\theta - \widehat{\theta}_{mmse}|^2\}$ monotonically increases with σ_{eff} [132], there exists at most one solution for the fixed point equation (4.46). However, the fixed point equations in (4.11) may have multiple solutions. Therefore, unlike in the *State-*σ_{eff} method, we

may need to search for the true solution among several values to find σ_{eff} when *Property1-*σ_{eff} method is used. Thus, *State-*σ_{eff} method leads to a more efficient numerical evaluation in computing σ_{eff}, compared with the *Property1-*σ_{eff} method, while the set up of the former is conditioned on the theoretical result which is introduced by the decoupling principle (facilitating the theoretical analysis that lead to the latter method).

4.4 Likelihood Ratio Test With SMV Based MMSE Estimation

In this section, we consider the activity detection problem based on the received signal at only one AP. From the previous section, we know that the MMSE estimation breaks down the received signal of the linear model (4.3a) into Gaussian noise corrupted scalar versions described in (4.42a). Then, we can obtain the corresponding likelihood functions as follows:

$$p(z_1^{mmse}|a=1) = \frac{1}{\pi\left(\beta+\sigma_{\mathit{eff}}^2\right)} \exp\left(-\frac{|z_1^{mmse}|^2}{\beta+\sigma_{\mathit{eff}}^2}\right), \tag{4.48}$$

$$p(z_1^{mmse}|a=0) = \frac{1}{\pi\sigma_{\mathit{eff}}^2} \exp\left(-\frac{|z_1^{mmse}|^2}{\sigma_{\mathit{eff}}^2}\right). \tag{4.49}$$

Now, the likelihood ratio is

$$\begin{aligned} R\left(z_1^{mmse}\right) &= \frac{p(z_1^{mmse}|a=1)}{p(z_1^{mmse}|a=0)} \\ &= \frac{\sigma_{\mathit{eff}}^2}{\beta+\sigma_{\mathit{eff}}^2} \exp\left(\frac{\beta|z_1^{mmse}|^2}{\sigma_{\mathit{eff}}^2(\beta+\sigma_{\mathit{eff}}^2)}\right). \end{aligned} \tag{4.50}$$

Finally, we can obtain the likelihood ratio test rule

$$R\left(z_1^{mmse}\right) \underset{a=0}{\overset{a=1}{\gtrless}} \frac{1-\lambda}{\lambda}. \tag{4.51}$$

After some algebraic operations, we have the following threshold detection rule:

$$|z_1^{mmse}|^2 \underset{a=0}{\overset{a=1}{\gtrless}} l'\left(\sigma_{\mathit{eff}}^2, \lambda, \beta\right) \tag{4.52}$$

where

$$l'\left(\sigma_{\mathit{eff}}^2, \lambda, \beta\right) = \frac{\sigma_{\mathit{eff}}^2(\beta + \sigma_{\mathit{eff}}^2)}{\beta} \log\left(\frac{(1-\lambda)(\beta + \sigma_{\mathit{eff}}^2)}{\lambda \sigma_{\mathit{eff}}^2}\right).$$

Then, the false alarm probability is

$$P_F = E\left\{\exp\left(-\frac{l'\left(\sigma_{\mathit{eff}}^2, \lambda, \beta\right)}{\sigma_{\mathit{eff}}^2}\right)\right\} \tag{4.53}$$

and the miss detection probability is

$$P_M = 1 - E\left\{\exp\left(-\frac{l'\left(\sigma_{\mathit{eff}}^2, \lambda, \beta\right)}{\beta + \sigma_{\mathit{eff}}^2}\right)\right\} \tag{4.54}$$

where the expectation is over the large-scale fading coefficient β. Finally, the error probability is

$$P_{err} = (1-\lambda)P_F + \lambda P_M. \tag{4.55}$$

4.5 Activity Detection With SMV Based MMSE Estimation in Cell-Free Massive MIMO Networks

In this section, we consider activity detection in cell-free massive MIMO systems with massive connectivity based on the cooperation of received pilot signals at the entire set of APs. As noted before, with the CB-AMP algorithm, we can obtain the SMV based MMSE estimates of the effective channel coefficients from all users to every AP, and the received pilot signals at each AP are decomposed into scalar Gaussian noise corrupted versions which can be

expressed as

$$z_2^{mmse} = \theta_{ij} + \sigma_{eff} n. \tag{4.56}$$

Within this cooperative framework, there are two different approaches, namely, centralized and distributed activity detection. In the centralized detection method, received pilot signals at all APs are transmitted to the CPU, and the decisions on the activity of all users are made based on all received pilot signals. In the distributed detection method, each AP makes its own decisions on the activity of all users based on its own received pilot signals, and subsequently the decisions and the corresponding detection reliabilities are transmitted to the CPU. The final decisions on the activity of all users are made at CPU based on the decisions from all APs while taking into account corresponding detection reliabilities.

4.5.1 Centralized Activity Detection

We denote the decoupled Gaussian noise corrupted signal of the MMSE estimate of the effective channel coefficients from the ith user to all the APs as $\mathbf{z}_i^{mmse} \in \mathcal{C}^{1\times M}$, which can be expressed as

$$\mathbf{z}_i^{mmse} = \boldsymbol{\theta}_i + \sigma_{eff}\mathbf{n}. \tag{4.57}$$

As also done in [9] and [10], we assume that each component of the noise vector $\mathbf{n}$ is independent of other components. Note that the probability density function of $\boldsymbol{\theta}_i$ is

$$p(\boldsymbol{\theta}_i) = (1-\lambda)\delta(\boldsymbol{\theta}_i) + \lambda\mathcal{CN}(\mathbf{0}, \boldsymbol{\Lambda}_i), \tag{4.58}$$

and the conditional probability density functions of $\mathbf{z}_i^{mmse}$ are

$$p\left(\mathbf{z}_i^{mmse}|a_i = 0\right) = \frac{1}{\pi^M \sigma_{eff}^{2M}} \exp\left(-\frac{|\mathbf{z}_i^{mmse}|^2}{\sigma_{eff}^2}\right) \tag{4.59}$$

and

$$p\left(\mathbf{z}_i^{mmse}|a_i=1\right)=\frac{1}{\pi^M\left|\boldsymbol{\Sigma}_i\right|}\exp\left(-\mathbf{z}_i^{mmse}\boldsymbol{\Sigma}_i^{-1}(\mathbf{z}_i^{mmse})^H\right), \tag{4.60}$$

where $\boldsymbol{\Sigma}_i=\boldsymbol{\Lambda}_i+\sigma_{\mathit{eff}}^2\mathbf{I}_M$. Then we can obtain the probability density function

$$p\left(\mathbf{z}_i^{mmse}\right)=(1-\lambda)\,p\left(\mathbf{z}_i^{mmse}|a_i=0\right)+\lambda p\left(\mathbf{z}_i^{mmse}|a_i=1\right). \tag{4.61}$$

Note that the conditional probability density function is given by

$$p(\mathbf{z}_i^{mmse}|\boldsymbol{\theta}_i)=\frac{1}{\pi^M\sigma_{\mathit{eff}}^{2M}}\exp\left(-\frac{\left|\mathbf{z}_i^{mmse}-\boldsymbol{\theta}_i\right|^2}{\sigma_{\mathit{eff}}^2}\right). \tag{4.62}$$

Then we can obtain the posterior conditional probability density function as

$$p(\boldsymbol{\theta}_i|\mathbf{z}_i^{mmse})=\frac{p(\mathbf{z}_i^{mmse}|\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i)}{p\left(\mathbf{z}_i^{mmse}\right)}. \tag{4.63}$$

Finally, we can express the MMSE estimate of $\boldsymbol{\theta}_j$ as

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_i^{mmse}&=E\left\{\boldsymbol{\theta}_i|\mathbf{z}_i^{mmse}\right\}\\&=\frac{\mathbf{z}_i^{mmse}\left(\sigma_{\mathit{eff}}^2\boldsymbol{\Lambda}_i^{-1}+\mathbf{I}_M\right)^{-1}}{1+\frac{(1-\lambda)|\boldsymbol{\Sigma}_i|}{\lambda\sigma_{\mathit{eff}}^{2M}}\exp\left(-\mathbf{z}_i^{mmse}\left(\frac{1}{\sigma_{\mathit{eff}}^2}\mathbf{I}_M-\boldsymbol{\Sigma}_i^{-1}\right)(\mathbf{z}_i^{mmse})^H\right)}.\end{aligned} \tag{4.64}$$

It is obvious that the elements of $\widehat{\boldsymbol{\theta}}_i^{mmse}$ are same with $\widehat{\theta}_{mmse}$ obtained in (4.39) with the scalar Gaussian noise corrupted linear model. Since the channel coefficients are independent of each other, $\boldsymbol{\Sigma}_j$ is a diagonal matrix. Thus,

$$\widehat{\boldsymbol{\theta}}_i^{mmse}=\frac{\mathbf{z}_i^{mmse}\left(\sigma_{\mathit{eff}}^2\boldsymbol{\Lambda}_i^{-1}+\mathbf{I}_M\right)^{-1}}{1+\frac{(1-\lambda)}{\lambda}\exp\left(-M\left(\varsigma_i-\kappa_i\right)\right)} \tag{4.65}$$

where

$$\varsigma_i=\frac{\mathbf{z}_i^{mmse}\left(\mathbf{I}_M-\sigma_{\mathit{eff}}^2\boldsymbol{\Sigma}_i^{-1}\right)(\mathbf{z}_i^{mmse})^H}{M\sigma_{\mathit{eff}}^2} \tag{4.66}$$

and

$$\kappa_i = \frac{1}{M}\sum\nolimits_{j=1}^{M} \log\left(\beta_{ij} + \sigma_{\mathit{eff}}^2\right) - 2\log\left(\sigma_{\mathit{eff}}\right). \tag{4.67}$$

As M grows without bound, we have $\widehat{\boldsymbol{\theta}}_i^{mmse} = \mathbf{z}_i^{mmse}\left(\sigma_{\mathit{eff}}^2\boldsymbol{\Lambda}_i^{-1} + \mathbf{I}_M\right)^{-1}$, which is MMSE estimate of $\boldsymbol{\theta}_i$ given $\mathbf{z}_i^{mmse}$, when $\varsigma_i > \kappa_i$, and we have $\widehat{\boldsymbol{\theta}}_i^{mmse} = 0$ when $\varsigma_i < \kappa_i$. Therefore, we obtain the following threshold detection rule:

$$\begin{cases} a_i = 1, \mathit{if}\ \varsigma_i > \kappa_i, \\ a_i = 0, \mathit{if}\ \varsigma_i < \kappa_i. \end{cases} \tag{4.68}$$

Let us denote $\widetilde{\mathbf{z}}_i^{mmse} = \mathbf{z}^{mmse}\left(\mathbf{I}_M - \sigma_{\mathit{eff}}^2\boldsymbol{\Sigma}_i^{-1}\right)^{1/2}/\sqrt{M}\sigma_{\mathit{eff}}$, $\varsigma_i^0 = \varsigma_i$ under the condition that $a_i = 0$, and $\varsigma_i^1 = \varsigma_i$ under the condition that $a_i = 1$. According to (4.57), $\widetilde{\mathbf{z}}_i^{mmse} \sim \mathcal{CN}\left(\mathbf{0}, \frac{\mathbf{I}_M - \sigma_{\mathit{eff}}^2\boldsymbol{\Sigma}_i^{-1}}{M}\right)$ when $a_i = 0$, and $\widetilde{\mathbf{z}}_i^{mmse} \sim \mathcal{CN}\left(\mathbf{0}, \boldsymbol{\Sigma}_i\left(\mathbf{I}_M - \sigma_{\mathit{eff}}^2\boldsymbol{\Sigma}_i^{-1}\right)/\left(M\sigma_{\mathit{eff}}^2\right)\right)$ when $a_i = 1$. Since $\varsigma_i = \widetilde{\mathbf{z}}_i^{mmse}(\widetilde{\mathbf{z}}_i^{mmse})^H$ and $\boldsymbol{\Sigma}_i$ is a diagonal matrix, ς_i is the squared summation of independent and circularly symmetric complex Gaussian random variables.

Denote the jth element of $\widetilde{\mathbf{z}}_i^{mmse}$ as z_{ij}^{mmse}. Then, $z_{ij}^{mmse} \sim \mathcal{CN}\left(0, \frac{\beta_{ij}}{M\left(\beta_{ij}+\sigma_{\mathit{eff}}^2\right)}\right)$ when $a_i = 0$, and $z_{ij}^{mmse} \sim \mathcal{CN}\left(0, \frac{\beta_{ij}}{M\sigma_{\mathit{eff}}^2}\right)$ when $a_i = 1$. Therefore, $(z_{ij}^{mmse})^2 \sim \Gamma\left(1, \frac{\beta_{ij}}{M\left(\beta_{ij}+\sigma_{\mathit{eff}}^2\right)}\right)$ when $a_i = 0$, and $(z_{ij}^{mmse})^2 \sim \Gamma\left(1, \frac{\beta_{ij}}{M\sigma_{\mathit{eff}}^2}\right)$ when $a_i = 1$, where $\Gamma(a, b)$ stands for the Gamma distribution with shape parameter a and scalar parameter b. We denote $\rho_{ij}^0 = \frac{\beta_{ij}}{M\left(\beta_{ij}+\sigma_{\mathit{eff}}^2\right)}$ and $\rho_{ij}^1 = \frac{\beta_{ij}}{M\sigma_{\mathit{eff}}^2}$.

Since $\frac{x}{1+x} < \log(1+x) < x$ when $x > 0$, we can obtain

$$\sum\nolimits_{j=1}^{M} \rho_{ij}^0 < \kappa_i < \sum\nolimits_{j=1}^{M} \rho_{ij}^1. \tag{4.69}$$

With the law of large numbers, we have $\varsigma_i^0 \to \sum_{j=1}^{M} \rho_{ij}^0$ and $\varsigma_i^1 \to \sum_{j=1}^{M} \rho_{ij}^1$ as M grows. Therefore, $\varsigma_i^0 < \kappa_i$ and $\varsigma_i^1 > \kappa_i$ when $M \to \infty$. Thus, the false alarm and miss detection probabilities of the threshold detection rule in (4.68) tend to zero as $M \to \infty$. So does the error probability.

Remark 4.5 *In order to compare the performances of different centralized activity detection methods, error probability of centralized activity detection with MMV based MMSE estimation in cell-free massive MIMO networks (which has been studied in [133]) is also provided in Section 4.7 via numerical results.*

4.5.2 Distributed Activity Detection

For distributed activity detection, each AP acquires the MMSE estimates of the effective channel coefficients from all users to this AP with its own received pilot signals. Subsequently, each AP makes its own decisions on the activity of all users based on the MMSE estimates and the threshold detection rule in (4.51), and sends these decisions and the corresponding reliabilities to the CPU. Then the final decisions on the activity of all users are made at the CPU based on the decisions from the entire set of APs, while also taking into account the corresponding reliabilities. The detection reliability of the decisions made at each AP can be determined via (4.53) and (4.54).

Let us denote the decision made by the jth AP for the ith user as d_{ij}, and consider the optimal fusion rule for the activity of the ith user. Based on d_{ij} $(j = 1, 2, \cdots M)$, we divide all the APs into two sets, $\mathcal{S}_0 = \{j|d_{ij} = 0\}$ and $\mathcal{S}_1 = \{j|d_{ij} = 1\}$. As shown in [134], the optimal fusion rule is

$$|\mathcal{S}_1| \log\left(\frac{1-P_M}{P_F}\right) + |\mathcal{S}_0| \log\left(\frac{P_M}{1-P_F}\right) \underset{a_i=0}{\overset{a_i=1}{\gtrless}} \log\left(\frac{1-\lambda}{\lambda}\right), \tag{4.70}$$

where P_F and P_M are given in (4.53) and (4.54), respectively. After several algebraic operation, it is easy to obtain the following fusion rule:

$$|\mathcal{S}_1| \log\frac{(1-P_M)(1-P_F)}{P_M P_F} \underset{a_i=0}{\overset{a_i=1}{\lessgtr}} \log\frac{1-\lambda}{\lambda} - M\log\frac{P_M}{1-P_F}. \tag{4.71}$$

Let us set $\chi = \log\frac{(1-P_M)(1-P_F)}{P_M P_F}$ and $\rho = (\log\frac{1-\lambda}{\lambda} - M\log\frac{P_M}{1-P_F})/\chi$. We assume $\lambda < 0.5$

in this chapter[1]. Then $l'\left(\sigma_{\mathit{eff}}^2, \lambda, \beta\right) > 0$ and

$$\exp\left(-\frac{l'\left(\sigma_{\mathit{eff}}^2, \lambda, \beta\right)}{\sigma_{\mathit{eff}}^2}\right) < \exp\left(-\frac{l'\left(\sigma_{\mathit{eff}}^2, \lambda, \beta\right)}{\beta + \sigma_{\mathit{eff}}^2}\right). \tag{4.72}$$

Therefore,

$$P_F + P_M < 1. \tag{4.73}$$

Thus, $\chi > 0$. Then, the optimal fusion rule is

$$|\mathcal{S}_1| \underset{a_i=0}{\overset{a_i=1}{\gtrless}} \rho. \tag{4.74}$$

When $a_i = 0$, we have $|\mathcal{S}_1| \sim B(M, P_F)$, where $B(M, P_F)$ stands for the binomial distribution. Then we can obtain the false alarm probability as

$$P_F^d = 1 - \psi\left(\lfloor \rho \rfloor\right) \tag{4.75}$$

where $\psi(\cdot)$ stands for the cumulative distribution function of the binomial distribution, and $\lfloor x \rfloor$ is the largest integer which is smaller than or equal to x. Similarly, when $a_i = 1$, $|\mathcal{S}_1| \sim B(M, 1 - P_M)$, and the corresponding miss detection probability of the system is

$$P_M^d = \psi\left(\lfloor \rho \rfloor\right). \tag{4.76}$$

Finally, the error probability of the system is

$$P_{err}^d = (1 - \lambda) P_F^d + \lambda P_M^d. \tag{4.77}$$

We define $f(x) = P_F \log(x) + (1 - P_F) \log(1 - x)$, $0 < x < 1$. It can be readily shown that $f(x)$ is monotonically decreasing within the interval $(P_F, 1]$. Since $P_F + P_M < 1$, we

[1]Massive connectivity and small active probability for each user are two important characteristics of IoT networks.

have $f(P_F) > f(1-P_M)$, i.e.,

$$P_F \log(P_F) + (1-P_F)\log(1-P_F) > P_F \log(1-P_M) + (1-P_F)\log(P_M). \tag{4.78}$$

Thus,

$$P_F \log\left(\frac{1-P_M}{P_F}\right) < (1-P_F)\log\left(\frac{1-P_F}{P_M}\right) + \frac{1}{M}\log\left(\frac{1-\lambda}{\lambda}\right). \tag{4.79}$$

Then, we can obtain

$$\lim_{M\to\infty} \frac{\rho}{MP_F} > 1. \tag{4.80}$$

Similarly, we can also obtain

$$\lim_{M\to\infty} \frac{\rho}{M(1-P_M)} < 1. \tag{4.81}$$

As M grows, we have $|\mathcal{S}_1| = MP_F$ when $a_i = 0$ and $|\mathcal{S}_1| = M(1-P_M)$ when $a_i = 1$. With (4.74), (4.80) and (4.81), we note that the false alarm and miss detection probabilities of the optimal fusion rule tend to zero as $M \to \infty$. Thus, the error probability of the optimal fusion rule also tends to zero when M grows without bound.

In cell-free massive MIMO systems, the backhaul link is from the APs to the CPU since there is no information exchange between APs. As a consequence of transmitting only activity decisions and the corresponding reliabilities by each AP, the capacity of backhaul link which is needed for distributed activity detection is much smaller than that needed for centralized activity detection, which sends the entire received pilots to the CPU by every AP.

Remark 4.6 *In this section, we have arrived at the conclusion that the error probabilities of both centralized and distributed activity detection tend to zero when the number of APs grows without bound. Since cell-free massive MIMO is a more general model that can be specialized to a single-cell massive MIMO system when all APs are located at the same node, these results confirm and generalize the conclusion in [10] that the error probabilities of activity*

detection in a single-cell massive MIMO system tend to zero when the number of antennas at base station goes to infinity.

4.6 Extensions to MMV Based MMSE Estimation and Activity Detection

Heretofore, we have considered the theoretical analysis of SMV based MMSE estimation for effective channel coefficients in cell-free massive MIMO systems with massive connectivity, and the corresponding theoretical analysis on error probabilities of activity detection rules with SMV based MMSE estimation results. In Section 4.3, the theoretical mean square error of SMV based MMSE estimation can be obtained with the *Property1-*σ_{eff} or *State-*σ_{eff} method and equation (4.46). The theoretical error probability of likelihood ratio test with SMV based MMSE estimation results depending on the received pilot signals at only one AP is shown in equations (4.52), (4.53) and (4.54) of Section 4.4. Section 4.5 addresses the theoretical analysis of likelihood ratio test with SMV based MMSE estimation results and cooperation among different APs. We do not obtain closed-form theoretical error probabilities in this section, but these probabilities are in general dependent on σ_{eff}, and they can be determined via numerical analysis, i.e., by substituting the theoretical value of σ_{eff} into (4.56) and finding the corresponding error probability numerically. Moreover, with the theoretical analysis, we find that the error probabilities of both centralized and distributed activity detection tend to zero when the number of APs tends to infinity while the asymptotic ratio between the number of users and pilots is kept constant.

In order to simplify theoretical analysis, we consider only SMV based methods in Section 4.3 through Section 4.5. However, since equation (4.4) presents an MMV problem, there can be a performance gap when the correlation between different columns of $\mathbf{Y}$ are not taken into account, and (4.4) is decomposed into multiple SMV problems. On the other hand, it is challenging to perform a theoretical analysis of MMV based MMSE estimation for effective

channel coefficients of cell-free massive MIMO systems with massive connectivity. Therefore, the performance loss between SMV and MMV based methods is shown via numerical results. In the next section, AMP algorithm for MMV proposed in [9] is employed to obtain several numerical results for the mean square error of MMV based MMSE estimation, and the error probability of centralized activity detection with MMV based MMSE estimation results are also determined numerically.

4.7 Numerical Results

In the setting for numerical analyze, we assume that the APs and users are uniformly distributed in a circular area with radius $R_r = 500$m. The path loss decay exponent is $\alpha = 2.5$, and the reference distance is set at $d_0 = 50$m. The total user number is $N = 4000$. CB-AMP and AMP for MMV algorithms are employed to obtain numerical results for the mean square error of SMV and MMV based MMSE estimates, respectively, of the effective channel coefficients in cell-free massive MIMO systems with massive connectivity. In the simulation settings of Figs. 4.2 and 4.4, we consider 10 APs in cell-free massive MIMO system, and the SNR at each AP is 30dB. In Figs. 4.3 and 4.5, we again consider 10 APs and assume that the number of pilots used for MMSE estimation is 300. In Figs. 4.6 and 4.7, the number of pilots used for MMSE estimation is 300, and the SNR at each AP is 30dB.

In the numerical results, we analyze the performance of MMSE estimation and activity detection considering different probabilities of each user being active. We plot the curves of the mean square error of oracle and MMSE estimators as a function of the number of pilots used for estimation and the SNR at each AP. The error probabilities of the likelihood ratio test based on MMSE estimation is also analyzed in this section. Additionally, we study the error probabilities of centralized and distributed detection in cell-free massive MIMO systems versus the number of APs.

For the simplification of descriptions, we first introduce several abbreviations that will

be used in the figures of the numerical results. "MSE^{T}_{SMV}" and "MSE^{N}_{SMV}" stand for the theoretical and numerical mean square errors of the MMSE estimation obtained with the replica analysis result (4.47) and CB-AMP algorithm, respectively. "MSE^{N}_{MMV}" denotes the mean square error of MMSE estimation acquired with the AMP for MMV algorithm. "$Oracle^{T}_{SMV}$" and "$Oracle^{N}_{SMV}$" indicate the theoretical and numerical mean square errors of the oracle estimation obtained with the asymptotic result (4.31) and the numerical result (4.22), respectively. "P^{T}_{err}" and "P^{N}_{err}" denote the theoretical and numerical error probabilities of the likelihood ratio test based on the decoupled scalar model (4.37) and the AWGN corrupted scalar model (4.14) involved with the CB-AMP algorithm, respectively. Similarly, "$P^{c,SMV}_{err,T}$" and "$P^{c,SMV}_{\mathrm{err,N}}$" stand for the theoretical and numerical error probabilities for the centralized likelihood ratio test with SMV based MMSE estimation results in cell-free massive MIMO systems. "$P^{d}_{err,T}$" and "$P^{d}_{\mathrm{err,N}}$" designate the theoretical and numerical results of error probabilities of the distributed likelihood ratio test under the optimal fusion rule (4.74) in cell-free massive MIMO systems. "$P^{c,SMV}_{err,T}$" and "$P^{d}_{err,T}$" are evaluated with the scalar Gaussian noise corrupted linear model (4.56), while "$P^{c,SMV}_{err,N}$" and "$P^{d}_{\mathrm{err,N}}$" are obtained with the CB-AMP algorithm and its state evolution equations (4.14). "$P^{c,MMV}_{err,N}$" denotes the error probability of centralized activity detection based on the MMSE estimates of the effective channel coefficients via AMP for MMV algorithm and the corresponding state evolution equations.

Fig. 4.2 plots the mean square error of the oracle, SMV and MMV based MMSE estimation in cell-free massive MIMO networks with massive connectivity as a function of the number of pilots L. These curves show that the numerical and theoretical results of oracle and SMV based MMSE estimation match well with each other. The mean square error decreases as the number of pilots increases, and the rate of decay in the curves increases as the probability of user being active decreases. Furthermore, compared with the MMV based MMSE estimation, the decay rate of the mean square error of SMV based estimation is smaller. The former has larger mean square error when L is small while it has smaller values

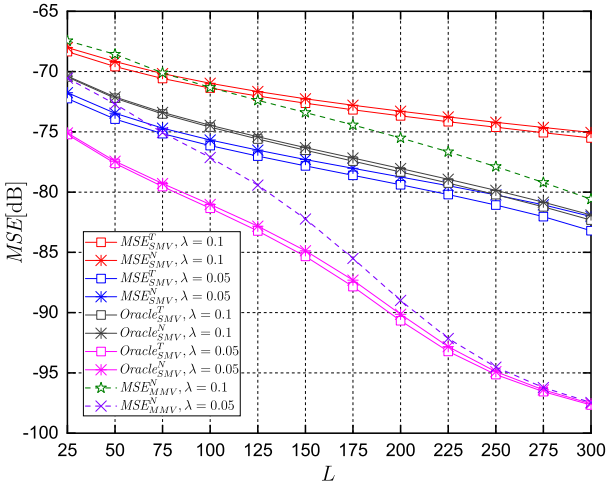

Figure 4.2: Mean square error of the oracle, SMV and MMV based MMSE estimation in cell-free Massive MIMO networks with massive connectivity versus the number of pilots L.

when L is larger than 75. Moreover, we observe that as a consequence of the uncertainty in users' activities, compared with the oracle estimation, under the same settings, the mean square error of SMV based MMSE estimation is larger, and the gap between them increases as λ decreases from 0.1 to 0.05.

Fig. 4.3 plots the mean square error of oracle, SMV and MMV based MMSE estimation in cell-free Massive MIMO networks with massive connectivity versus SNR at each AP. When SNR is small, the mean square errors diminish as SNR increases. However, when SNR is large enough, the tendency of the mean square error becomes different. Since the total number of users is 4000, the expected number of active users is 400 when $\lambda = 0.1$, and this becomes 200 when $\lambda = 0.05$. Recall that the number of pilots used for estimation is 300. Therefore, when $\lambda = 0.1$, the received signal which is comprised of the transmitted signal from all active users constitutes an underdetermined linear system. Thus, the mean square error of the oracle estimation becomes flat when SNR becomes large enough. As a consequence of the system being underdetermined and the uncertainty of users' activities, the mean square errors of both SMV and MMV based MMSE estimation also flatten out

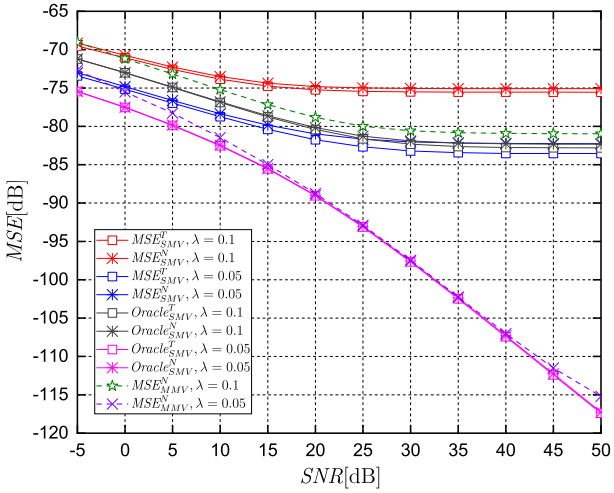

Figure 4.3: Mean square error of the oracle, SMV and MMV based MMSE estimation in cell-free Massive MIMO networks with massive connectivity versus SNR at each AP.

when SNR becomes large enough. On the contrary, the received signal from all active users constitutes an overdetermined linear system when $\lambda = 0.05$. Therefore, the mean square error of oracle estimation keeps decreasing as SNR increases, and tends to zero as SNR grows without bound. However, since users' activities are unknown to the APs, the mean square error of both SMV and MMV based MMSE estimation still flatten out after SNR reaches a certain high level when $\lambda = 0.05$. Moreover, we have also noticed that the theoretical mean square error of oracle and SMV based MMSE estimation match well with that of the numerical results. These curves also show that the mean square error of the MMV based MMSE estimation is always smaller than that of SMV based results under the same settings.

Fig. 4.4 plots the theoretical and numerical results for the error probability of the likelihood ratio test based on the received pilot signals at only one AP versus the number of pilots used in the MMSE estimation. We see that the error probabilities diminish as the number of pilots increases. We also note that when λ decreases, the error probabilities diminish more quickly.

Fig. 4.5 plots the theoretical and numerical results for the error probabilities of the

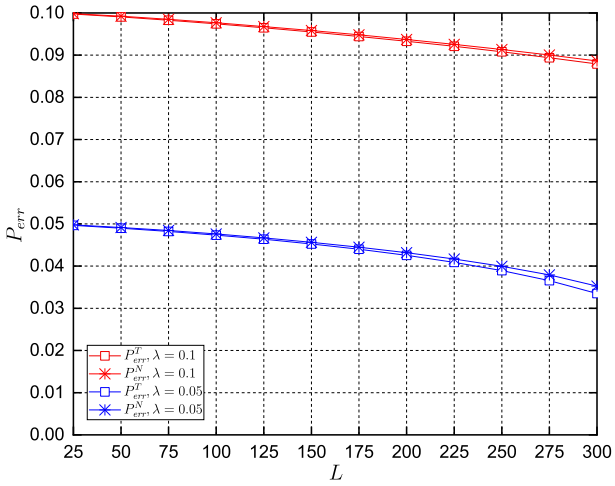

Figure 4.4: Error probability of likelihood ratio test based on the received pilot signals at only one AP versus the number of pilots used during the MMSE estimation.

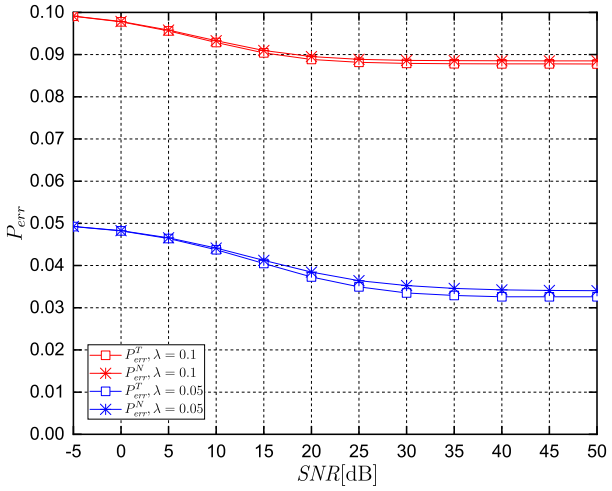

Figure 4.5: Error probability of likelihood ratio test based on the received pilot signals at only one AP versus SNR.

likelihood ratio test based on the received pilot signals at only one AP versus SNR. From the analysis of Fig. 4.3, we know that the mean square error of MMSE estimation diminishes as SNR initially grows starting from small values. After SNR becomes large enough, the mean square error flattens out. Since the likelihood ratio test is based on the MMSE estimation result, as SNR increases, error probabilities should also decrease with SNR initially, and then become flat when the SNR is beyond a certain threshold. Comparing Figs. 4.3 and 4.5, we can see that the mean square errors and error probabilities becomes flat at the same SNR value within the same numerical simulation settings, which is consistent with our theoretical analysis.

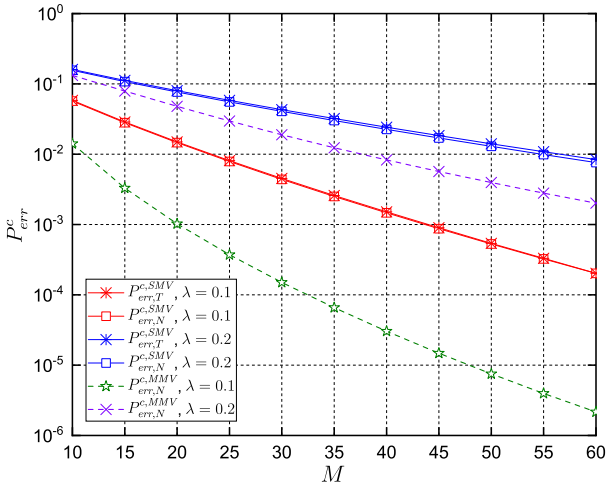

Figure 4.6: Error probability of centralized detection in cell-free Massive MIMO networks with massive connectivity versus the number of APs.

Figs. 4.6 and 4.7 plot the Error probabilities of centralized and distributed detection in cell-free massive MIMO networks with massive connectivity versus the number of APs, respectively. The error probabilities diminish as the number of APs increases, confirming our conclusion that the error probabilities of both centralized and distributed detection tend to zero when the number of AP tends to infinity while the ratio between the number of users and pilots is kept constant. When the user activity probability decreases from 0.2 to 0.1, the

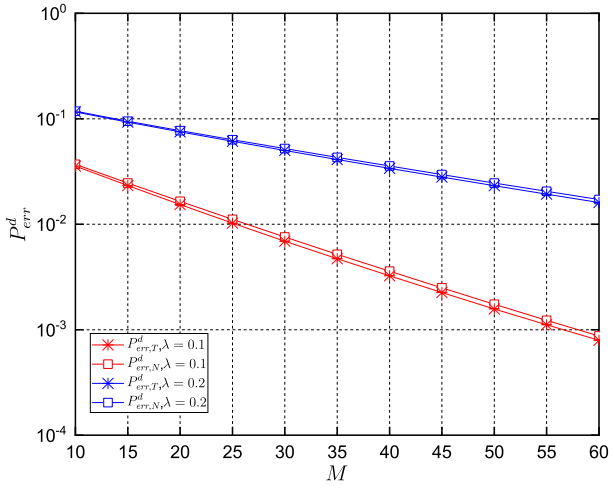

Figure 4.7: Error probability of distributed detection with the optimal fusion rule in cell-free massive MIMO networks with massive connectivity versus the number of APs.

error probabilities decrease, and the rate of decrease becomes larger. Furthermore, in Fig. 4.6, we plot the curves for centralized detection with both SMV and MMV based MMSE estimation results of the effective channel coefficients. The gaps between the SMV and MMV based results show the performance loss due to disregarding the correlation between different columns of the received pilot signals when performing channel estimation.

Chapter 5

Performance Analysis of Cell-Free Massive MIMO Systems with Massive Connectivity

In this chapter, we investigate the achievable rate of cell-free massive MIMO systems with massive connectivity. Our analysis is divided into two parts. First, we analyze the joint activity detection and channel estimation using the pilots transmitted from active users via GAMP algorithm and likelihood ratio test. MMSE estimates of the effective channel coefficients from all users to all APs can be obtained with the GAMP algorithm. With the decoupling properties of MMSE estimation for large linear systems [61], the received pilot signals are decomposed into scalar additive complex Gaussian noise corrupted versions. We determine the variance of the scalar additive complex Gaussian noise (which is also subsequently referred to as the converged noise variance) with the help of the state evolution equations of the GAMP algorithm. Then, we identify the variances of both the estimated channel and the corresponding channel estimation error. Secondly, we address the uplink data transmission from active users. With the results in [76], we analyze the achievable uplink data transmission rates in cell-free massive MIMO systems with massive connectivity

when ZF detector is employed. In this setting, we demonstrate the impact of the accuracy of activity detection on the achievable rates via numerical results.

5.1 System Model

We consider a cell-free massive MIMO system that consists of M APs and N single antenna users. All the APs and users are uniformly distributed in a circular area with radius R_r, and all the APs are connected to the CPU through a backhaul network. Let us denote by d_{ij} the distance from the ith user to the jth AP. The probability density function (PDF) of d_{ij} is

$$p(d_{ij}) = \frac{4d_{ij}}{\pi {R_r}^2}\left[\cos^{-1}\left(\frac{d_{ij}}{2R_r}\right) - \frac{d_{ij}}{2R_r}\left(1-\frac{d_{ij}^2}{4R_r^2}\right)^{\frac{1}{2}}\right] \tag{5.1}$$

for $0 < d_{ij} < 2R_r$ and $p(d_{ij}) = 0$ otherwise [129].

We consider flat fading channel between all the users and APs in this chapter. Besides, massive connectivity and sparse activity are assumed, i.e., there is a large number of users uniformly distributed in the area, while the number of active users at a given time instant is small. We call the probability of each user being active at a given time instant as the activity probability, and assume that the activity probability of all the users is the same and denoted by λ. Moreover, the activities of all the users keep unchanged during each channel coherence interval. At a given time instant, we denote the activity of the ith user by the binary-valued a_i. Specifically, $a_i = 1$ stands for the ith user being active, while $a_i = 0$ denotes inactivity. Then, the activity probability is $p(a_i = 1) = 1 - p(a_i = 0) = \lambda$.

The channel coefficient from the ith user to the jth AP is

$$g_{ij} = \beta_{ij}^{1/2} h_{ij} \tag{5.2}$$

where β_{ij} is the large-scale fading coefficient which can be expressed as $\beta_{ij} = \min\left(d_{ij}^{-\alpha}, d_0^{-\alpha}\right)$, d_0 is the reference distance, and α is the path loss decay exponent [65]. Since β_{ij} changes

slowly over time, it can be measured in advance and we assume that it is known at the CPU and all the APs. With the PDF of d_{ij} as given in (5.1), we can obtain the PDF of β_{ij} and denote it by $p(\beta)$. $h_{ij} \sim \mathcal{CN}(0,1)$ is the small scale fading coefficient.

We assume that there are T symbols in each channel coherence interval. During each channel coherence interval, L symbols are used for joint activity detection and channel estimation, and the remaining $T - L$ symbols are used for uplink data transmission from the users to APs. We call the former as the joint activity detection and channel estimation phase, and the latter as the data transmission phase. Furthermore, we assume the total bandwidth of the system is B Hz.

We denote the combination of the ith user's activity and the channel coefficient from the ith user to the jth AP as the effective channel coefficient θ_{ij}, i.e., $\theta_{ij} = a_i g_{ij}$. In order to simplify the description, we represent the effective channel coefficients from all the users to the jth AP by the $N \times 1$ dimensional vector $\boldsymbol{\theta}_j$, whose ith element is θ_{ij}. Then, in the joint activity detection and channel estimation phase, the received pilot signal at the jth AP can be expressed as

$$\mathbf{y}_{p,j} = \boldsymbol{\Phi}\boldsymbol{\theta}_j + \mathbf{n}_{p,j} \tag{5.3}$$

where $\boldsymbol{\Phi}$ is the pilot matrix, and $\mathbf{n}_{p,j} \in \mathcal{CN}(0, \sigma_0^2 \mathbf{I}_L)$ is the independent and identically distributed additive white Gaussian noise at the jth AP. We assume that the elements of the $L \times N$ complex pilot matrix $\boldsymbol{\Phi}$ are independent and circularly symmetrically distributed with zero mean and variance $1/L$. We further assume that both the number of pilots and number of users grow without bound with their ratio kept constant, i.e., $\lim_{N\to\infty} \frac{N}{L} = \gamma$.

Let us also denote the activity support of all the users, i.e., the set of the indices of active users, by $\mathcal{S}$. Moreover, we represent the channel coefficients from the ith user to all the APs by the $M \times 1$ dimensional vector $\boldsymbol{\varphi}_i$, whose jth element is g_{ij}. Now, in the uplink data transmission phase, the received signals at all the APs form an $M \times 1$ vector and can be

expressed as

$$\mathbf{y}_d = \sum_{i \in \mathcal{S}} \boldsymbol{\varphi}_i x_i + \mathbf{n}_d \tag{5.4}$$

where x_i is the data transmitted from the ith user, and $\mathbf{n}_d \in \mathcal{CN}(0, \sigma_0^2 \mathbf{I}_M)$ is the noise vector whose components represent the independent and identically distributed Gaussian noise terms at the APs. We assume that the transmitted signals from all the users satisfy $E\left\{|x_i|^2\right\} = 1$.

5.2 Performance Analysis

The GAMP algorithm proposed in [56] is used for joint activity detection and channel estimation in this chapter. Specifically, with GAMP algorithm, we obtain the MMSE estimates of the channel coefficients. We note that GAMP algorithm decomposes the received pilot signal into scalar Gaussian noise corrupted versions, and the noise variance can be obtained with the state evolution equations of the GAMP algorithm. Following this approach, we identify the statistical characterizations of the estimated channel coefficients and the corresponding channel estimation error.

In the data transmission phase, the estimated channel coefficients are used for ZF detection. With the results in [76], we analyze the achievable rate of cell-free massive MIMO systems with massive connectivity.

5.2.1 Joint Activity Detection and Channel Estimation

In this section, we address the decoupling principle of the GAMP algorithm, and analyze the joint activity detection and channel estimation phase.

Using the replica symmetric postulated MMSE decoupling properties given in [61] and [127], we can obtain the following decoupling principle of the GAMP algorithm for cell-free massive MIMO systems:

$$\widehat{y}_{p,ij} = \theta_{ij} + \sigma_{\mathit{eff}} n \tag{5.5}$$

where $\widehat{y}_{p,ij}$ is the GAMP decoupled pilot signal, and $n \sim \mathcal{CN}(0,1)$ is the additive white Gaussian noise.

Since $\theta_{ij} = a_i g_{ij}$, the PDF of θ_{ij} is

$$p(\theta_{ij}) = (1-\lambda)\delta(\theta_{ij}) + \lambda \mathcal{CN}(\theta_{ij}; 0, \beta_{ij}) \tag{5.6}$$

where $\delta(\cdot)$ is the Dirac delta function, and $\mathcal{CN}(\cdot; a, b)$ denotes the PDF of a circularly symmetric complex Gaussian random variable with mean a and variance b. Then, the MMSE estimate of θ_{ij} is

$$\widehat{\theta}_{ij} = G\left(\left|\widehat{y}_{p,ij}\right|^2; \sigma^2_{\mathit{eff}}, \lambda, \beta_{ij}\right) \frac{\beta_{ij}\widehat{y}_{p,ij}}{\beta_{ij} + \sigma^2_{\mathit{eff}}} \tag{5.7}$$

where

$$G\left(\left|\widehat{y}_{p,ij}\right|^2; \sigma^2_{\mathit{eff}}, \lambda, \beta_{ij}\right) = \frac{1}{1 + \frac{(1-\lambda)(\beta_{ij}+\sigma^2_{\mathit{eff}})}{\lambda \sigma^2_{\mathit{eff}}} \exp\left(-\frac{\beta_{ij}|\widehat{y}_{p,ij}|^2}{\sigma^2_{\mathit{eff}}(\beta_{ij}+\sigma^2_{\mathit{eff}})}\right)}. \tag{5.8}$$

The state evolution equation of the GAMP algorithm is

$$\widehat{r}^t_{ij} = \theta_{ij} + \sqrt{\xi^t} n \tag{5.9}$$

where $\widehat{r}^t_{ij}$ is the decoupled received pilot signal at the tth iteration of the GAMP algorithm, and the noise variance ξ^t satisfies the following state evolution equation [56]:

$$\xi^{t+1} = \sigma_0^2 + \gamma E\left\{\left|\theta_{ij} - \widehat{\theta}^t_{ij}\right|^2\right\} \tag{5.10}$$

where the expectation (denoted by $E\{\cdot\}$) is over both large-scale and small-scale channel fading, and $\widehat{\theta}^t_{ij}$ is the MMSE estimate of θ_{ij} based on $\widehat{r}^t_{ij}$ in (5.9), which can be obtained using the formulation in (5.7). When the GAMP algorithm converges, we have $\xi^{t+1} = \xi^t = \sigma^2_{\mathit{eff}}$.

Thus, we can obtain

$$E\left\{\left|\theta_{ij}-\widehat{\theta}_{ij}^{t}\right|^{2}\right\}=\lambda\left(E\{\beta\}-\int_{\beta_{\min}}^{\beta_{\max}}\int_{0}^{\infty}F(\sigma_{\mathit{eff}},\beta,t)dtd\beta\right) \tag{5.11}$$

where

$$F(\sigma_{\mathit{eff}},\beta,t)=\frac{\lambda tp(\beta)e^{-t\sigma_{\mathit{eff}}^{2}/\beta}}{\left(\frac{\beta}{\sigma_{\mathit{eff}}^{2}}+1\right)^{2}\left(\frac{\lambda}{\beta+\sigma_{\mathit{eff}}^{2}}+\frac{(1-\lambda)e^{-t}}{\sigma_{\mathit{eff}}^{2}}\right)}, \tag{5.12}$$

$\beta_{\min}=(2R_r)^{-\alpha}$, and $\beta_{\max}=d_0^{-\alpha}$. Then, we have the following fixed point equation when the GAMP algorithm converges:

$$\sigma_{\mathit{eff}}^{2}=\sigma_{0}^{2}+\gamma E\left\{\left|\theta_{ij}-\widehat{\theta}_{ij}^{t}\right|^{2}\right\}. \tag{5.13}$$

By solving the fixed point equation in (5.13), we can find the noise variance σ_{eff}^2 of the GAMP decoupled pilot signal. With this, we can determine the MMSE estimate of the effective channel coefficients, θ_{ij}, using (5.7). Now for active users, it is immediate that

$$\begin{aligned} var\left(\widehat{\theta}_{ij}\right) &= E_1\left\{\left(\widehat{\theta}_{ij}-E_1\left\{\widehat{\theta}_{ij}\right\}\right)\left(\widehat{\theta}_{ij}-E_1\left\{\widehat{\theta}_{ij}\right\}\right)^{*}\right\} \\ &= \frac{\beta_{ij}^{2}}{\beta_{ij}+\sigma_{\mathit{eff}}^{2}} \end{aligned} \tag{5.14}$$

where $var(z)$ is the variance of z, and $E_1\{\cdot\}$ stands for the expectation over small-scale channel fading. We define the channel estimation error as

$$\varepsilon_{ij}=\widehat{\theta}_{ij}-\theta_{ij}. \tag{5.15}$$

Then,

$$\begin{aligned} cov\left(\varepsilon_{ij},\widehat{\theta}_{ij}\right) &= E_1\left\{\left(\varepsilon_{ij}-E_1\left\{\varepsilon_{ij}\right\}\right)\left(\widehat{\theta}_{ij}-E_1\left\{\widehat{\theta}_{ij}\right\}\right)^{*}\right\} \\ &= 0 \end{aligned} \tag{5.16}$$

where $cov(z_1, z_2)$ denotes the covariance of z_1 and z_2. Therefore, ε_{ij} and $\widehat{\theta}_{ij}$ are uncorrelated. Recall that we assume the error probability of activity detection is zero in this chapter, which equals to the case that the user activities are known as priori. Since only active users are considered during the data transmission phase (λ can be regarded as 1), ε_{ij} and $\widehat{\theta}_{ij}$ are complex Gaussian random variables based on the results in (5.7) and (5.15). Thus, ε_{ij} and $\widehat{\theta}_{ij}$ are independent of each other. Then, we can obtain

$$var(\varepsilon_{ij}) = \frac{\beta_{ij}\sigma_{\mathit{eff}}^2}{\beta_{ij} + \sigma_{\mathit{eff}}^2}. \tag{5.17}$$

As noted in the introduction, the error probability in activity detection with GAMP algorithm and likelihood ratio test tends to zero as M tends to infinity and $\lim_{N\to\infty} \frac{N}{L} = \gamma$, which is assumed to be satisfied in this chapter. Thus, we will not consider the effect of activity detection errors on the performance analysis in terms of achievable rates in uplink data transmission in the next subsection, while the influence is studied via numerical simulations subsequently. Due to space limitations, we refer to [135] for more details on the activity detection with the GAMP algorithm and likelihood ratio test.

5.2.2 Achievable Rates of Uplink Data Transmission

For active users, we have $\theta_{ij} = g_{ij}$. Therefore, $g_{ij} = \theta_{ij} = \widehat{\theta}_{ij} - \varepsilon_{ij}$. Substituting g_{ij} into (5.4), we can obtain

$$\mathbf{y}_d = \sum_{i\in\mathcal{S}} \widehat{\boldsymbol{\theta}}_i x_i - \sum_{i\in\mathcal{S}} \boldsymbol{\varepsilon}_i x_i + \mathbf{n}_d \tag{5.18}$$

where $\widehat{\boldsymbol{\theta}}_i = [\widehat{\theta}_{i1}, \widehat{\theta}_{i2}, \cdots, \widehat{\theta}_{iM}]^T$ and $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iM}]^T$.

Let us denote the estimated effective channel coefficients from all active users to the entire AP set by the $|\mathcal{S}| \times M$ matrix $\mathbf{B}$, i.e., the element on the ith row and jth column of $\mathbf{B}$, b_{ij}, equals $\widehat{\theta}_{ij}$ for $i \in \mathcal{S}$. For uplink data transmission, we employ the ZF receiver at the

CPU, i.e., the detector matrix is $\mathbf{\Omega} = \mathbf{B}\left(\mathbf{B}^H\mathbf{B}\right)^{-1}$. Then, the detected signal is

$$\begin{aligned}\widehat{x}_k &= \boldsymbol{\omega}_k^H \mathbf{y}_d \\ &= \sum_{i\in\mathcal{S}} \boldsymbol{\omega}_k^H \widehat{\boldsymbol{\theta}}_i x_i - \sum_{i\in\mathcal{S}} \boldsymbol{\omega}_k^H \boldsymbol{\varepsilon}_i x_i + \boldsymbol{\omega}_k^H \mathbf{n}_d \end{aligned} \tag{5.19}$$

where $\boldsymbol{\omega}_k$ is the kth column of $\mathbf{\Omega}$. Now, using the fact that Gaussian noise is the worst-case noise [136], the achievable uplink rate can be obtained as

$$R = \frac{T-L}{T} B \widetilde{R} \tag{5.20}$$

where

$$\widetilde{R} = \sum_{i\in\mathcal{S}} E_1 \left\{ \log_2 \left(1 + \frac{1}{\sum\limits_{j\in\mathcal{S}} |\omega_{ij}|^2 \sum\limits_{k\in\mathcal{S}} var\left(\varepsilon_{jk}\right) + \|\omega_k\|_2^2} \right) \right\}. \tag{5.21}$$

Next, we can determine a lower bound $\widehat{R}$ for the achievable uplink rate as follows [76]:

$$\widehat{R} = \frac{T-L}{T} B \sum_{i\in\mathcal{S}} \log_2 \left(1 + \frac{1}{\varepsilon_{\max} + 1} \varsigma_i \left(\xi_i - 1\right) \right) \tag{5.22}$$

where

$$\varepsilon_{\max} = \max \sum_{i\in\mathcal{S}} var(\varepsilon_{ij}), \tag{5.23}$$

$$\varsigma_i = \frac{\sum\limits_{m\in\mathcal{M}_k} \left(var\left(\widehat{\theta}_{im}\right)\right)^2}{\sum\limits_{m\in\mathcal{M}_k} var\left(\widehat{\theta}_{im}\right)}, \tag{5.24}$$

$$\xi_i = \frac{\left(\sum\limits_{m\in\mathcal{M}_k} var\left(\widehat{\theta}_{im}\right)\right)^2}{\sum\limits_{m\in\mathcal{M}_k} \left(var\left(\widehat{\theta}_{im}\right)\right)^2}, \tag{5.25}$$

$\mathcal{M}_k = \mathcal{M}/\mathcal{A}_k$, $\mathcal{M} = \{m|\forall m = 1, 2, \cdots, M\}$, $\mathcal{A}_k = Unique\left(\left\{ m_i^* = \arg\max\limits_m var(\widehat{\theta}_{ij}) \,|\forall m \neq k \right\}\right)$, and $Unique(\mathcal{T})$ returns the same values as in set $\mathcal{T}$ but with no repetitions.

5.3 Numerical Analysis

As discussed in the introduction, the achievable rates in massive MIMO system with massive connectivity have been addressed in [137]. With the state evolution equations of AMP algorithm and large SNR approximation, the converged noise variance of the decoupled received pilot signal is determined as

$$\tau_{\infty}^2 = \frac{\sigma_0^2}{1 - K/L} \tag{5.26}$$

where K is the number of active users. Then, based on τ_{∞}^2 and the MMSE channel estimate, uplink achievable rates with different detectors at the base station are identified. τ_{∞}^2 obtained in [137] is in the same position as σ_{eff}^2 in (5.5). However, the use of (5.26) requires the satisfaction of two conditions, namely large SNR and $K < L$, while our method to obtain σ_{eff}^2 with the fixed point equation in (5.13) does not need these conditions and hence is applicable more generally.

In this section, we compare the achievable uplink rates with both τ_{∞}^2 and σ_{eff}^2. Moreover, we can also obtain the noise variance of the decoupled received pilot signal numerically, and we refer to the achievable uplink rate with this method as the GAMP simulation result. Additionally, if the user activity is known a priori, we can assign orthogonal pilots to all the active users, and activity detection phase is not needed. It is obvious that the achievable uplink rate with prior knowledge on user activity provides an upper bound for the achievable rate without such information on user activity. As we discussed before, when M tends to infinity, the error rates in activity detection tend to zero, which is equivalent to the case in which the prior information for user activity is available. Therefore, as M grows without bound, the achievable uplink rate with σ_{eff}^2 should overlap with the achievable rate with known user activities. Indeed, numerical results in this section verify this conclusion.

We assume that there are 1000 users uniformly distributed in a circular region with radius 500m. The path-loss decay exponent is $\alpha = 2.5$. There are 1000 symbols in a channel

coherence interval, and the bandwidth of the system is 1 MHz. The sum rate is averaged over 10^5 realizations of this setting.

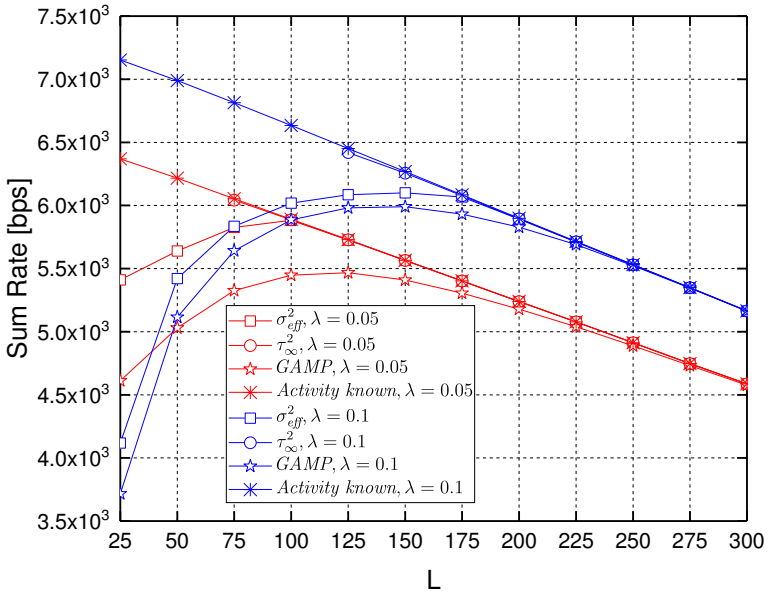

Figure 5.1: Achievable uplink rate versus the number of pilots

Fig. 5.1 plots the curves for the achievable uplink rate versus the number of pilots used in the activity detection and channel estimation phase, for different activity probabilities with 100 APs uniformly distributed in the circular region and SNR = 30 dB. Since we can obtain τ^2_∞ only when $L > K$, the sum rates based on τ^2_∞ are plotted after $L = 75$ when $\lambda = 0.05$ and $L = 125$ when $\lambda = 0.1$. Since the number of symbols in a channel coherence interval is fixed, fewer symbols will be used in the uplink data transmission phase if more symbols are used for activity detection and channel estimation. Therefore, the sum rate of the system decreases with increasing L when user activities are known at the CPU. On the other hand, a tradeoff exists when user activity needs to be detected. Specifically, when L is small, error probability in activity detection is relatively high, leading to low achievable sum rates. Hence, a reduction in the sum rate is experienced due to uncertainty in user activity detection. Since the error probability in activity detection diminishes as L increases, sum rates initially increase. However, as L exceeds a threshold, smaller duration of time being available for uplink data transmission starts being the dominant factor and sum rates start

decreasing. Therefore, there exists a tradeoff between sum rate and the number of pilots in the design of practical cell-free massive MIMO systems with massive connectivity and uncertainty in user activity. Comparing the curves for $\lambda = 0.05$ and $\lambda = 0.1$, we observe that the performance gap due to user activity uncertainty is smaller in the former case (due to smaller error rates in the presence of lower activity probabilities), while larger sum rates are achieved in the latter case (with $\lambda = 0.1$) due to generally higher number of active users. We also notice that when L is sufficiently large, the achievable sum rates based on σ^2_{eff} and τ^2_∞ and also the sum rate obtained via GAMP simulations all start overlapping with that achieved when user activities are perfectly known a priori. Hence, in all cases, performance upper bound is approached as L grows.

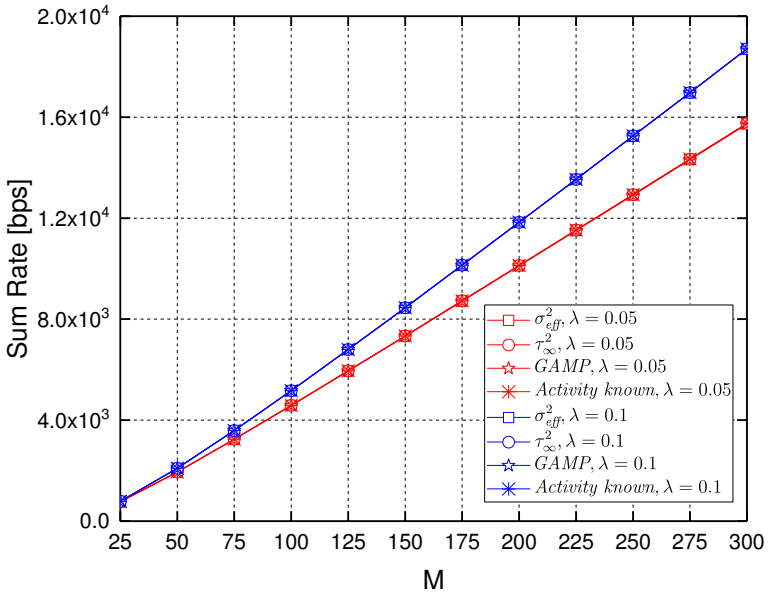

Figure 5.2: Achievable uplink rate versus the number of APs.

Fig. 5.2 plots the curves for achievable uplink rates versus the number of APs, for different activity probabilities with $L = 300$ and SNR $= 30$ dB. These curves demonstrate that the sum rates based on σ^2_{eff} and τ^2_∞ overlap with that achieved with prior information on user activities (labeled as "activity known") and also with GAMP simulation results. This shows that the influence of user activity uncertainty on the achievable sum rate can be safely disregarded under assumption of relatively large M, L and N (e.g., $L = 300$ in this

figure), i.e., having no errors in activity detection is a valid assumption. We also see that the achievable sum rates linearly increase as the number of APs grows, and the slope becomes higher as the activity probability increases. Furthermore, the achievable sum rates, similarly as before, diminish when λ is decreased from 0.1 to 0.05.

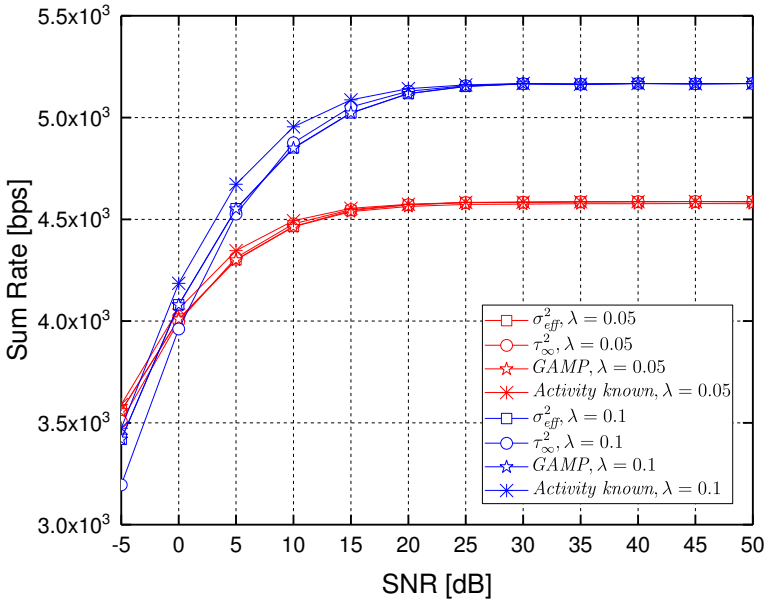

Figure 5.3: Achievable uplink rate versus SNR at APs.

Fig. 5.3 plots the curves for the achievable uplink rates versus SNR at APs for different activity probabilities with $L = 300$ and $M = 100$. When SNR is small, the error probability in activity detection is relatively large, which leads to the gaps between achievable sum rates with prior information on user activities and those achieved with σ^2_{eff}, τ^2_{∞} and GAMP algorithm. These performance gaps can be regarded as the cost of uncertainty in the detection of user activity. We also notice that the achievable sum rates obtained with σ^2_{eff} and GAMP algorithm overlap, while a smaller sum rate is achieved with τ^2_{∞} when SNR is low. This verifies that σ^2_{eff} predicts the converged noise variance of GAMP algorithm better than τ^2_{∞} at low SNR levels, which is expected due to the fact that τ^2_{∞} is derived under the large SNR assumption. We further observe in the figure that the achievable sum rates increase and the cost of uncertainty in the detection of user activity diminishes as SNR grows. Indeed, when SNR is sufficiently large, the sum rates based on σ^2_{eff}, τ^2_{∞} and GAMP algorithm overlap

with that achieved with prior information on user activities. Similarly as before, when λ is decreased from 0.1 to 0.05, both the achievable sum rates and the performance gap due to user activity uncertainty diminish.

Chapter 6

Channel Estimation for Intelligent Reflecting Surface Assisted Wireless Communications

In the prior work on channel estimation in IRS assisted wireless communication systems, proposed methods generally have high costs in terms of the computational complexity and number of pilot symbols. In this chapter, we focus on finding an efficient channel estimation method for IRS assisted wireless communication systems. By considering the equivalent channel from the BS to the users, we can address the channel estimation problem employing conventional methods for wireless networks. With this approach, we determine that each row vector of the equivalent channel has a Bessel K distribution, and all the rows are independent of each other. By introducing a GSM model, we obtain the MMSE estimate of the equivalent channel, and identify analytical upper and lower bounds on the mean square error. Applying the central limit theorem, we perform an asymptotic analysis of the channel estimation, through which we show that the upper bound on the mean square error of the MMSE estimation equals the asymptotic mean square error of the MMSE estimation when the number of reflecting elements at the IRS tends to infinity. Compared with prior

work on channel estimation in IRS assisted wireless communication systems, our proposed channel estimation method is completed in one stage via transmitting orthogonal pilots from the users, and enables us to obtain analytical expressions for the MMSE estimate of the equivalent channel coefficients through a more efficient scheme with low computational complexity.

6.1 System Model

We consider an IRS assisted wireless communication system as depicted in Fig. 6.1, which is comprised of one BS equipped with M antennas, an intelligent reflecting surface with M_1 reflecting elements, and N users. All the users are uniformly distributed in a circle with radius R_r and the IRS is located at the center of the circle. $\mathbf{G}_1 \in \mathcal{C}^{N \times M_1}$, $\mathbf{G}_2 \in \mathcal{C}^{M_1 \times M}$ and $\mathbf{V} \in \mathcal{C}^{M_1 \times M_1}$ denote the channel coefficients in the link from the IRS to the users, channel coefficients from the BS to the IRS, and the scattering matrix at the IRS, respectively. We assume that the direct link between the BS and users is not operational as a result of unfavorable propagation conditions (e.g., due to blockages) [87, 115].

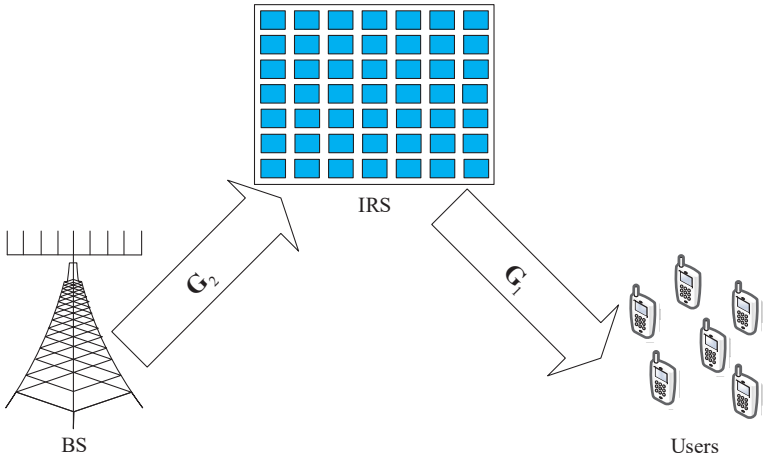

Figure 6.1: IRS assisted wireless communication systems.

Furthermore, we assume flat fading channels between the BS, IRS and users, i.e., $\mathbf{G}_1 = \text{diag}\left(\sqrt{\beta_1}\right)\mathbf{H}_1$ and $\mathbf{G}_2 = \sqrt{\beta_2}\mathbf{H}_2$, where $\text{diag}\left(\cdot\right)$ is the diagonalization operation for a given

vector [47, 138]. $\boldsymbol{\beta}_1 = [\beta_{1,1}, \beta_{1,2}, \cdots, \beta_{1,N}]$ is the path-loss vector from the IRS to the entire set of users, where $\beta_{1,i} = \beta_{10} + 10\alpha_1 \log_{10}\left(\frac{d_{1,i}}{d_{10}}\right)$ is the path-loss from the IRS to the ith user in dB, where $d_{1,i}$ is the distance from the IRS to the ith user, d_{10} is the reference distance for the path-loss between the IRS and users, β_{10} is the corresponding path-loss at d_{10}, and α_1 is the path-loss decay exponent. $\mathbf{H}_1 \in \mathcal{C}^{N \times M_1}$ is the small-scale fading channel coefficients from the IRS to all the users, and all the elements of $\mathbf{H}_1$ are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and unit variance. Similarly, the path-loss between BS and IRS is $\beta_2 = \beta_{20} + 10\alpha_2 \log_{10}\left(\frac{d_2}{d_{20}}\right)$ in dB, where d_2 is the distance between the BS and IRS, d_{20} is the reference distance for the path-loss between the BS and IRS, β_{20} is the corresponding path-loss at d_{20}, and α_2 is the path-loss decay exponent. All the elements of the small-scale fading matrix $\mathbf{H}_2$ are i.i.d. complex Gaussian random variables with zero mean and unit variance. Since path-loss changes slowly over time, it can be measured in advance and we assume $\boldsymbol{\beta}_1$ and β_2 are known at the BS.

As noted in the literature, we assume that the scattering matrix $\mathbf{V}$ is reconfigurable and known at the BS [93, 139]. We have $\mathbf{V} = \text{diag}\left(v_1 \exp(j\theta_1), v_2 \exp(j\theta_2), \cdots, v_{M_1} \exp(j\theta_{M_1})\right)$, where $v_i \in [0, 1]$ and $\theta_i \in [0, 2\pi]$ represent the amplitude and the phase coefficient for the ith element of the IRS, respectively. For the sake of simplicity in the analysis, we assume $v_1 = v_2 = \cdots = v_{M_1} = v$ in this chapter.

6.2 Statistics of the equivalent channel from the BS to the users

In order to perform channel estimation at the BS, we need to know the distribution of the equivalent channel matrix from the BS to the users, which is denoted as $\mathbf{G} = \mathbf{G}_1 \mathbf{V} \mathbf{G}_2$. We denote g_{ik} as the element on the ith row and kth column of $\mathbf{G}$, and $\mathbf{g}_i \in \mathcal{C}^{1 \times M}$ as the ith row of $\mathbf{G}$. Additionally, $h_{1,ik}$, $h_{2,ik}$, $\mathbf{h}_{1,i}$ and $\mathbf{h}_{2,i}$ represent the elements on the ith row and kth column of $\mathbf{H}_1$ and $\mathbf{H}_2$, and the ith row of $\mathbf{H}_1$ and $\mathbf{H}_2$, respectively.

6.2.1 Distribution of g_{ik}

With the definitions given above, g_{ik} can be expressed as

$$\begin{aligned} g_{ik} &= v\sqrt{\beta_{1,i}\beta_2}\sum_{m=1}^{M_1} h_{1,im}h_{2,mk}\exp(j\theta_m) \\ &= v\sqrt{\beta_{1,i}\beta_2}\sum_{m=1}^{M_1} h_{1,im}h_{3,mk} \end{aligned} \tag{6.1}$$

where $h_{3,mk} = h_{2,mk}\exp(j\theta_m)$, and $h_{3,mk} \sim \mathcal{CN}(0,1)$ as a result of the distribution of $h_{2,mk}$. Since $h_{1,im}$ and $h_{2,mk}$ are independent of each other, $h_{1,im}$ is also independent of $h_{3,mk}$.

Let us denote $g_{ik} = g_1 + jg_2$ and $t = t_1 + jt_2$. Then, the characteristic function of g_{ik} is [140]

$$\Psi_{g_{ik}}(t_1,t_2) = \frac{1}{\left(1+\frac{\beta_{1,i}\beta_2 v^2}{4}\left(t_1^2+t_2^2\right)\right)^{M_1}}. \tag{6.2}$$

Thus, the probability density function of g_{ik} can be obtained as

$$\begin{aligned} & p_{g_{ik}}(g_1,g_2) \\ =&\frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Psi_{g_{ik}}(t_1,t_2)\exp\left(-j\left(g_1t_1+g_2t_2\right)\right)dt_1dt_2 \\ =&\frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp\left(-j\left(g_1t_1+g_2t_2\right)\right)}{\left(1+\frac{\beta_{1,i}\beta_2 v^2}{4}\left(t_1^2+t_2^2\right)\right)^{M_1}}dt_1dt_2 \\ =&\frac{1}{4\pi^2}\int_{0}^{\infty}\int_{0}^{2\pi}\frac{r\exp\left(-jr\left(g_1\cos(\theta)+g_2\sin(\theta)\right)\right)}{\left(1+\frac{\beta_{1,i}\beta_2 v^2 r^2}{4}\right)^{M_1}}d\theta dr \\ =&\frac{1}{2\pi}\int_{0}^{\infty}\frac{r}{\left(1+\frac{\beta_{1,i}\beta_2 v^2 r^2}{4}\right)^{M_1}}J_0\left(r\sqrt{g_1^2+g_2^2}\right)dr \\ =&\frac{2(g_1^2+g_2^2)^{\frac{M_1-1}{2}}}{\pi\Gamma(M_1)(\beta_{1,i}\beta_2 v^2)^{\frac{M_1+1}{2}}}K_{M_1-1}\left(\frac{2}{v\sqrt{\beta_{1,i}\beta_2}}\sqrt{g_1^2+g_2^2}\right) \end{aligned} \tag{6.3}$$

where $J_0(\cdot)$ represents zeroth-order Bessel function of the first kind, and $K_n(\cdot)$ denotes nth-order modified Bessel function of the second kind.

6.2.2 Distribution of $\mathbf{g}_i$

From (6.1), we can obtain that

$$g_{ij_1}g_{ij_2} = \beta_{1,i}\beta_2 v^2 \sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_1} h_{1,im_1}h_{1,im_2}h_{3,m_1k_1}h_{3,m_2k_2}. \tag{6.4}$$

Therefore, $E\{g_{ij_1}g_{ij_2}\} = 0$, where $E\{\cdot\}$ stands for the expectation of the random variable. As $E\{g_{ij_1}\} = E\{g_{ij_2}\} = 0$, we have $cov(g_{ij_1}, g_{ij_2}) = E\{g_{ij_1}g_{ij_2}\} = 0$, where $cov(a,b)$ represents the covariance between random variables a and b. Therefore, all the elements of $\mathbf{g}_i$ are uncorrelated with each other. Moreover, $p(g_{ik}|\mathbf{h}_{1,i}) \sim \mathcal{CN}\left(0, \beta_{1,i}\beta_2 v^2 \sum\limits_{m=1}^{M_1}|h_{1,im}|^2\right)$. Thus, $p(\mathbf{g}_i|\mathbf{h}_{1,i}) \sim \mathcal{CN}\left(0, \left(\beta_{1,i}\beta_2 v^2 \sum\limits_{m=1}^{M_1}|h_{1,im}|^2\right)\mathbf{I}_M\right)$, where $\mathbf{I}_M$ denotes the identity matrix with dimension M. Then, we can obtain the joint probability density function of $\mathbf{g}_i$ and $\mathbf{h}_{1,i}$ as

$$\begin{aligned}
&p(\mathbf{g}_i, \mathbf{h}_{1,i}) \\
=&p(\mathbf{g}_i|\mathbf{h}_{1,i})p(\mathbf{h}_{1,i}) \\
=&\frac{1}{\pi^{M+M_1}\left(\beta_{1,i}\beta_2 v^2 \sum\limits_{m=1}^{M_1}|h_{1,im}|^2\right)^M}\exp\left(\varphi(\beta_{1,i},\beta_2,v,\mathbf{g}_i,\mathbf{h}_{1,i})\right)
\end{aligned} \tag{6.5}$$

where

$$\varphi(\beta_{1,i},\beta_2,v,\mathbf{g}_i,\mathbf{h}_{1,i}) = -\frac{\|\mathbf{g}_i\|^2}{\beta_{1,i}\beta_2 v^2 \sum\limits_{m=1}^{M_1}|h_{1,im}|^2} - \|\mathbf{h}_{1,i}\|^2. \tag{6.6}$$

Now, the probability density function of $\mathbf{g}_i$ can be obtained as

$$\begin{aligned}
&p_{\mathbf{g}}(\mathbf{g}_i) \\
=&\int_{C^{M_1}} p(\mathbf{g}_i, \mathbf{h}_{1,i})d\mathbf{h}_{1,i} \\
=&\frac{2}{\pi^M\Gamma(M_1)(\beta_{1,i}\beta_2 v^2)^M} * \int_0^\infty r^{2M_1-2M-1}\exp\left(-\frac{\|\mathbf{g}_i\|^2}{\beta_{1,i}\beta_2 v^2 r^2} - r^2\right)dr \\
=&\frac{2\|\mathbf{g}_i\|^{M_1-M}}{\pi^M\Gamma(M_1)(\beta_{1,i}\beta_2 v^2)^{\frac{M+M_1}{2}}}K_{M_1-M}\left(\frac{2}{v\sqrt{\beta_{1,i}\beta_2}}\|\mathbf{g}_i\|\right).
\end{aligned} \tag{6.7}$$

6.2.3 Asymptotic distribution of $\mathbf{g}_i$ as $M_1 \to \infty$

Let us denote $h_{imk} = h_{1,im}h_{3,mk}$. From (6.3), we can obtain that $p(h_{imk}) = \frac{2}{\pi}K_0\left(2\left|h_{imk}\right|\right)$. Since $h_{1,im}$ and $h_{3,mk}$ are both complex Gaussian random variables with zero mean and unit variance and they are independent of each other, the components of the summation in (6.1) are independent and identically distributed with zero mean and variance $\beta_{1,i}\beta_2 v^2$. Therefore, according to central limit theorem, $g_{ik} \sim \mathcal{CN}(0, M_1\beta_{1,i}\beta_2 v^2)$ as M_1 grows. As noted above, all the elements of $\mathbf{g}_i$ are uncorrelated with each other. Thus, we approximately have $\mathbf{g}_i \sim \mathcal{CN}(0, M_1\beta_{1,i}\beta_2 v^2\mathbf{I}_M)$ for sufficiently large M_1.

6.3 MMSE channel estimation

We assume that the channel estimation is performed at the BS, and N symbols are used for pilot transmission during each channel coherence interval. We denote the pilot matrix as $\mathbf{P}$, and assume that $\mathbf{P}$ is a unitary matrix. The received pilot signal at the BS can be expressed as

$$\mathbf{Y} = \mathbf{PG} + \mathbf{N} \tag{6.8}$$

where $\mathbf{N}$ denotes the complex additive white Gaussian noise at the BS whose elements are independent and identically distributed with zero mean and variance σ^2. Then, multiplying both sides of (6.8) with $\mathbf{P}^H$, we can obtain

$$\begin{aligned}\widetilde{\mathbf{Y}} &= \mathbf{P}^H\mathbf{Y} \\ &= \mathbf{G} + \mathbf{P}^H\mathbf{N}.\end{aligned} \tag{6.9}$$

Since $\mathbf{P}$ is a unitary matrix, $\widetilde{\mathbf{N}} = \mathbf{P}^H\mathbf{N}$ has the same Gaussian distribution as $\mathbf{N}$ with zero mean and variance σ^2.

We have derived the probability density function for each row of $\mathbf{G}$ in the previous section, and it is obvious that different rows of $\mathbf{G}$ are independent of each other. Besides,

the distribution of $\mathbf{g}_i$ can be equivalently represented as a GSM model [141]

$$\mathbf{g}_i = av\sqrt{\beta_{1,i}\beta_2}\mathbf{x}_i \tag{6.10}$$

where $\mathbf{x}_i \sim \mathcal{CN}(0, \mathbf{I}_M)$, and a is a scalar gamma random variable whose probability density function can be expressed as

$$p_A(a) = \frac{2a^{2M_1-1}}{\Gamma(M_1)}\exp(-a^2), \quad a > 0. \tag{6.11}$$

Then, we can obtain

$$\widetilde{\mathbf{Y}} = av\,\mathrm{diag}\left(\sqrt{\boldsymbol{\beta}_1\beta_2}\right)\mathbf{X} + \widetilde{\mathbf{N}} \tag{6.12}$$

where $\mathbf{X} = \left[\mathbf{x}_1^T, \mathbf{x}_2^T, \cdots, \mathbf{x}_N^T\right]^T$, and $\mathbf{x}_i \sim \mathcal{CN}(0, \mathbf{I}_M)$. For a given a, the two components of $\widetilde{\mathbf{Y}}$, $av\,\mathrm{diag}\left(\sqrt{\beta_1\beta_2}\right)\mathbf{X}$ and $\widetilde{\mathbf{N}}$, are both complex Gaussian random variables. Therefore, the conditional MMSE estimate of $\mathbf{G}$ is

$$\widehat{\mathbf{G}}(a) = \mathrm{diag}\left(\frac{\boldsymbol{\beta}_1\beta_2 a^2v^2}{\boldsymbol{\beta}_1\beta_2 a^2v^2 + \sigma^2}\right)\widetilde{\mathbf{Y}}. \tag{6.13}$$

Then, we can obtain the MMSE estimate of $\mathbf{G}$ which can be expressed as

$$\widehat{\mathbf{G}} = \int_0^\infty \widehat{\mathbf{G}}(a)p_A(a)da = M_1\,\mathrm{diag}\left(\left(\frac{\sigma^2}{\boldsymbol{\beta}_1\beta_2v^2}\right)^{M_1}\Gamma_1\left(-M_1, \frac{\sigma^2}{\boldsymbol{\beta}_1\beta_2v^2}\right)\exp\left(\frac{\sigma^2}{\boldsymbol{\beta}_1\beta_2v^2}\right)\right)\widetilde{\mathbf{Y}} \tag{6.14}$$

where $\Gamma_1(a, z) = \int_z^\infty t^{a-1}\exp(-t)dt$ is the upper incomplete gamma function.

Now, let us consider the mean square error of the MMSE estimate $\widehat{\mathbf{G}}$, which we denote as $\mathrm{mse}\left(\widehat{\mathbf{G}}\right)$. For brevity in the description, we denote the ith row of $\widehat{\mathbf{G}}$ as $\widehat{\mathbf{g}}_i$, and the corresponding mean square error as $\mathrm{mse}\left(\widehat{\mathbf{g}}_i\right)$. From (6.10), we know that given a, $\mathbf{g}_i$ is conditionally distributed as $\mathcal{CN}(0, \beta_{1,i}\beta_2a^2v^2\mathbf{I}_M)$, and the probability density function of a is $p_A(a)$. In this case, $\mathrm{mse}\left(\widehat{\mathbf{G}}\right)$ has no simple analytical expression, and we can only obtain

the following upper and lower bounds [142]:

$$\int_0^\infty \frac{\beta_{1,i}\beta_2 a^2 v^2 \sigma^2}{\beta_{1,i}\beta_2 a^2 v^2 + \sigma^2} p_A(a) da \leq \text{mse}\left(\widehat{\mathbf{g}}_i\right) \leq \frac{\beta_{1,i}\beta_2 v^2 \sigma^2 \int_0^\infty a^2 p_A(a) da}{\beta_{1,i}\beta_2 v^2 \int_0^\infty a^2 p_A(a) da + \sigma^2}. \tag{6.15}$$

Substituting (6.11) into (6.15), we can further obtain

$$M_1 \frac{\sigma^{2(M_1+1)}}{\left(\beta_{1,i}\beta_2 v^2\right)^{M_1}} \Gamma_1\left(-M_1, \frac{\sigma^2}{\beta_{1,i}\beta_2 v^2}\right) \exp\left(\frac{\sigma^2}{\beta_{1,i}\beta_2 v^2}\right) \leq \text{mse}\left(\widehat{\mathbf{g}}_i\right) \leq \frac{M_1 \beta_{1,i}\beta_2 v^2 \sigma^2}{M_1 \beta_{1,i}\beta_2 v^2 + \sigma^2}. \tag{6.16}$$

Finally, we can determine the upper and lower bounds of mse $\left(\widehat{\mathbf{G}}\right)$ as

$$\begin{aligned} \frac{1}{N}\text{sum}\left[M_1 \frac{\sigma^{2(M_1+1)}}{\left(\boldsymbol{\beta}_1\beta_2 v^2\right)^{M_1}} \Gamma_1\left(-M_1, \frac{\sigma^2}{\boldsymbol{\beta}_1\beta_2 v^2}\right) \exp\left(\frac{\sigma^2}{\boldsymbol{\beta}_1\beta_2 v^2}\right)\right] \leq \\ \text{mse}\left(\widehat{\mathbf{G}}\right) \leq \frac{1}{N}\text{sum}\left[\frac{M_1 \boldsymbol{\beta}_1\beta_2 v^2 \sigma^2}{M_1 \boldsymbol{\beta}_1\beta_2 v^2 + \sigma^2}\right] \end{aligned} \tag{6.17}$$

where sum $[\cdot]$ represents the summation of all the elements in a vector.

In Section 6.2.3, we have shown that $\mathbf{g}_i \sim \mathcal{CN}(0, M_1\beta_{1,i}\beta_2 v^2 \mathbf{I}_M)$ as M_1 gets larger. Therefore, the MMSE estimate of $\mathbf{G}$ for large M_1 is

$$\widetilde{\mathbf{G}} = \text{diag}\left(\frac{M_1 \boldsymbol{\beta}_1\beta_2 v^2}{M_1 \boldsymbol{\beta}_1\beta_2 v^2 + \sigma^2}\right) \widetilde{\mathbf{Y}}, \tag{6.18}$$

and the correspond mean square error is

$$\text{mse}\left(\widetilde{\mathbf{G}}\right) = \frac{1}{N}\text{sum}\left[\frac{M_1 \boldsymbol{\beta}_1\beta_2 v^2 \sigma^2}{M_1 \boldsymbol{\beta}_1\beta_2 v^2 + \sigma^2}\right] \tag{6.19}$$

which is equal to the upper bound of mse $\left(\widehat{\mathbf{G}}\right)$ in (6.17).

6.4 Numerical Analysis

We assume that there are 20 users in the IRS assisted wireless communication system, and all the users are uniformly distributed in a circular region with radius 1000m. The minimum distance between the IRS and a user is 500m. There are 20 antennas equipped at the BS. The reference distance for the path-loss between the BS, IRS and users are $d_{10} = d_{20} = 1$m, and the correspond path-loss decay exponents are $\alpha_1 = 2$ and $\alpha_2 = 2.8$, respectively. The path-loss at the reference distance is 30 dB, and the distance from the BS to the IRS is $d_2 = 100$m. The phase coefficients of the IRS reflecting elements are uniformly distributed in $[0, 2\pi]$.

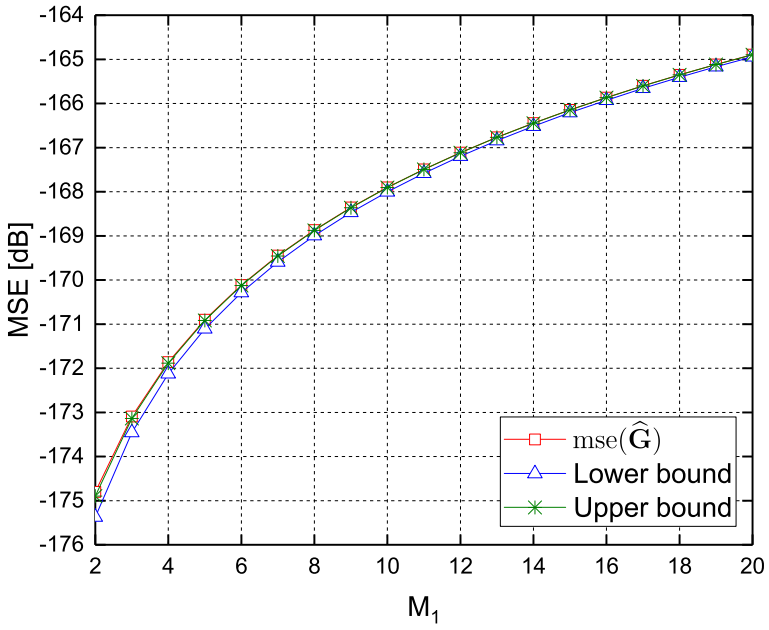

Figure 6.2: Mean square error of the MMSE estimate mse($\widehat{\mathbf{G}}$), and the corresponding analytical upper and lower bounds versus the number of reflecting elements at IRS when SNR=0 dB and $v = 1$.

Fig. 6.2 plots the mean square error of the MMSE estimate mse($\widehat{\mathbf{G}}$), and the corresponding analytical upper and lower bounds versus the number of reflecting elements at the IRS when SNR= 0 dB and $v = 1$. mse($\widehat{\mathbf{G}}$) represents the Monte Carlo simulation results of the mean square error for the MMSE estimate in (6.14). The "Upper bound" and "Lower bound" denote the numerical results of the analytical bounds presented in (6.17). These

curves show that the gap between the upper and lower bound are very small. When M_1 is larger than 4, the upper bound overlaps with mse($\widehat{\mathbf{G}}$), which matches well with our asymptotic analysis in (6.19). We also note that since larger M_1 will introduce more uncertainties in the equivalent channel matrix $\mathbf{G}$, the mean square error increases as M_1 grows.

Fig. 6.3 plots the mean square error of the MMSE estimate mse($\widehat{\mathbf{G}}$), and the corresponding analytical upper and lower bounds versus SNR when the number of reflecting elements at the IRS is 10 and $v = 1$. These curves show that the mean square error decreases dramatically as SNR increases. In particular, mean square error in dB decays linearly with increasing SNR in dB when SNR is larger than 0 dB. We also observe that the upper and lower bounds, and mse($\widehat{\mathbf{G}}$) again almost overlap with each other.

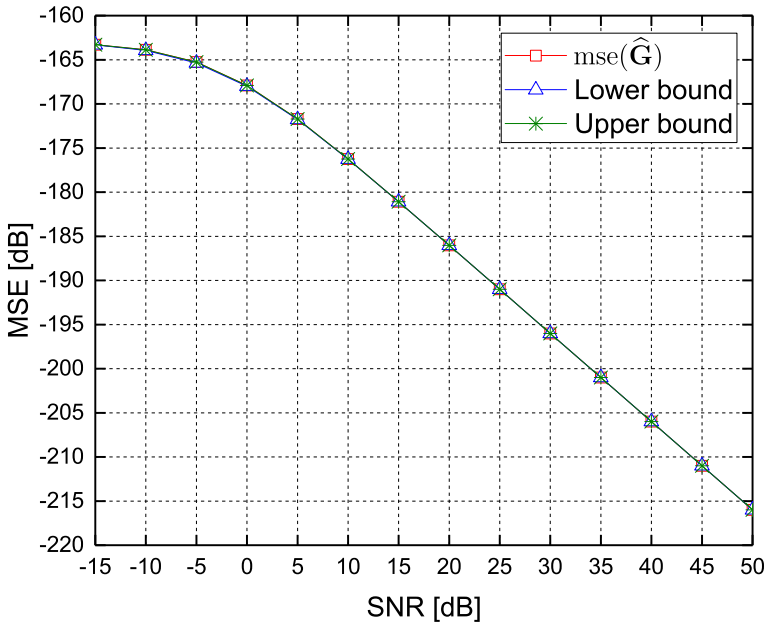

Figure 6.3: Mean square error of the MMSE estimate mse($\widehat{\mathbf{G}}$), and the corresponding analytical upper and lower bounds versus SNR when the number of reflecting elements at IRS is 10 and $v = 1$.

Fig. 6.4 plots the mean square error of the MMSE estimate mse($\widehat{\mathbf{G}}$), and the corresponding analytical upper and lower bounds versus the scattering amplitude at IRS when the number of reflecting elements at the IRS is 10 and SNR= 0 dB. These curves show that the mean square error increases as the scattering amplitude at the IRS grows, which could also

be derived considering the monotonicity of the upper and lower bound expressions in (6.17). The mean square error of the MMSE estimation overlaps with the upper bound, which further verifies our asymptotic characterization in (6.19). As noted above, these curves also exhibit a very small gap between the upper and lower bounds.

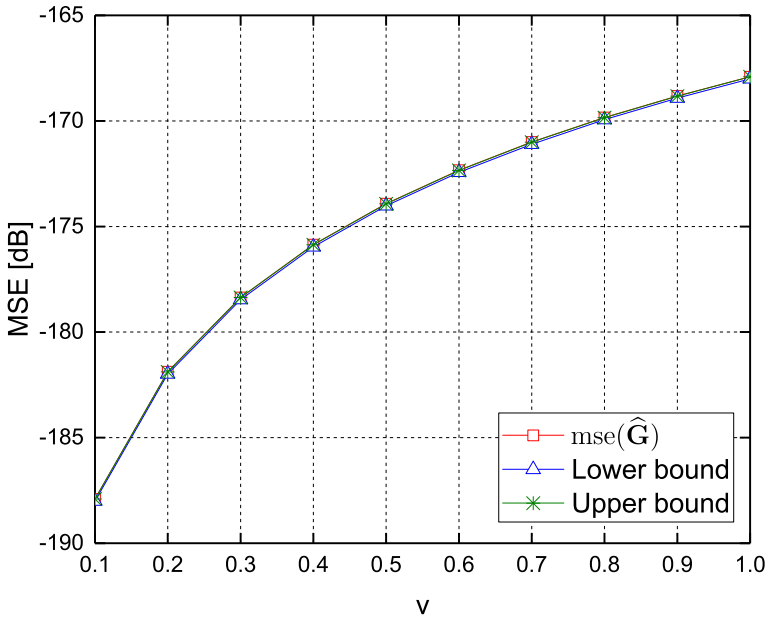

Figure 6.4: Mean square error of the MMSE estimate mse($\widehat{\mathbf{G}}$), and the corresponding analytical upper and lower bounds versus the scattering amplitude at IRS when the number of reflecting elements at IRS is 10 and SNR=0 dB.

Chapter 7

Sparse Activity Detection in Intelligent Reflecting Surface Assisted Wireless Networks

In this chapter, we consider the sparse activity detection problem in IRS assisted wireless networks. The sparse activity detection is performed by employing the GAMP algorithm [56, 143], likelihood ratio test and the optimal fusion rule. Via the GAMP algorithm, the received pilot signals at the BS are decoupled into scalar Gaussian noise corrupted versions of the effective channel coefficients. Subsequently, we obtain the MMSE estimate of the effective channel coefficients and identify a threshold detection rule. The optimal fusion rule is used to acquire final decisions on the activity of each user.

7.1 System Model

As depicted in Fig. 7.1, we consider the uplink of an IRS assisted wireless communication system, which consists of one BS equipped with M antennas, an IRS comprised of M_1 reflecting elements, and N users. All the users are uniformly distributed within a circle of radius R_r, and the IRS is located at the center of the circle. The distance

from the BS to the IRS is d_1. We assume that the scattering matrix at the IRS is reconfigurable and is known at the BS [93, 139]. We can express the scattering matrix as $\mathbf{V} = \text{diag}\left(v_1 \exp(j\theta_1), v_2 \exp(j\theta_2), \cdots, v_{M_1} \exp(j\theta_{M_1})\right)$, where $\text{diag}\left(\cdot\right)$ is the diagonalization operation for a given vector, $v_i \in [0, 1]$ and $\theta_i \in [0, 2\pi]$ represent the amplitude and the phase coefficient of the ith element of the IRS, respectively. We assume $v_1 = v_2 = \cdots = v_{M_1} = v$ in this chapter.

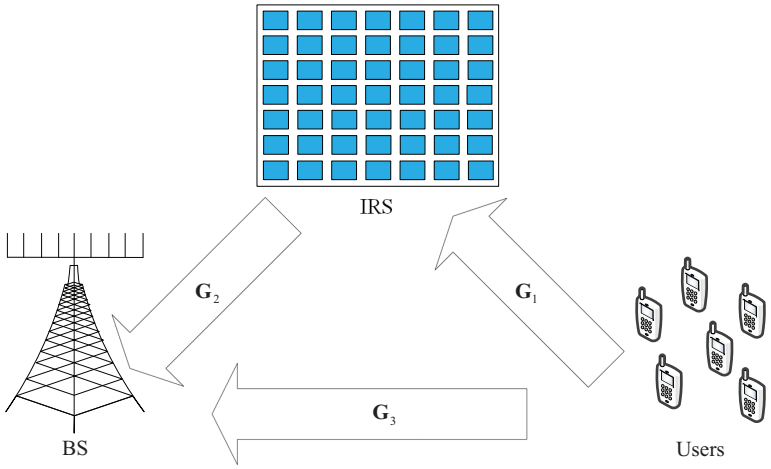

Figure 7.1: Uplink of the IRS assisted wireless network.

We further assume flat fading channels between the BS, IRS and users in this chapter, i.e., $\mathbf{G}_1 = \text{diag}\left(\sqrt{\boldsymbol{\beta}_1}\right)\mathbf{H}_1$, $\mathbf{G}_2 = \sqrt{\beta_2}\mathbf{H}_2$, and $\mathbf{G}_3 = \text{diag}\left(\sqrt{\boldsymbol{\beta}_3}\right)\mathbf{H}_3$, where $\mathbf{G}_1$, $\mathbf{G}_2$ and $\mathbf{G}_3$ are the channel coefficients from the IRS to users, from the BS to IRS, and from the BS to users, respectively. $\boldsymbol{\beta}_1 = [\beta_{1,1}, \beta_{1,2}, \cdots, \beta_{1,N}]$ is the path-loss vector from IRS to the entire set of users, where $\beta_{1,i} = \max\left(\beta_{1,0}, \beta_{1,0} + 10\alpha_1 \log_{10}\left(\frac{d_{1,i}}{d_{1,0}}\right)\right)$ is the path-loss from the IRS to the ith user in dB, where $d_{1,i}$ is the distance from the IRS to the ith user, $d_{1,0}$ is the reference distance for the path-loss between IRS and users, $\beta_{1,0}$ is the corresponding path-loss at $d_{1,0}$, and α_1 is the path-loss decay exponent. $\mathbf{H}_1$ is the small-scale fading channel coefficients from IRS to all users, and all the elements of $\mathbf{H}_1$ are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and unit variance. Similarly, β_2 and $\boldsymbol{\beta}_3 = [\beta_{3,1}, \beta_{3,2}, \cdots, \beta_{3,N}]$ denote the path-loss between the BS and IRS and the path-loss between the BS and all users, respectively. We further have $\beta_2 = \max\left(\beta_{2,0}, \beta_{2,0} + 10\alpha_2 \log_{10}\left(\frac{d_1}{d_{2,0}}\right)\right)$

and $\beta_{3,i} = \max\left(\beta_{3,0}, \beta_{3,0} + 10\alpha_3 \log_{10}\left(\frac{d_{3,i}}{d_{3,0}}\right)\right)$, where $d_{2,0}$ and $d_{3,0}$ are the reference distances for the corresponding path-loss, $\beta_{2,0}$ and $\beta_{3,0}$ are the path-loss at the corresponding reference distance, α_2 and α_3 are the corresponding path-loss decay exponents, and $d_{3,i}$ is the distance from the BS to the ith user. $\mathbf{H}_2$ and $\mathbf{H}_3$ are the corresponding small-scale fading channel coefficients in $\mathbf{G}_2$ and $\mathbf{G}_3$, and all the elements of $\mathbf{H}_2$ and $\mathbf{H}_3$ are i.i.d. complex Gaussian random variables with zero mean and unit variance as well. We assume $\boldsymbol{\beta}_1$, β_2 and $\boldsymbol{\beta}_3$ are known at the BS.

Sparsity is considered among the users in this chapter, i.e., not all the users are active simultaneously. At a given time instant, only a fraction of the users are active, while the remaining users stay inactive. We use activity probability, which is denoted as λ, to represent the sparsity of the activity of the users, and we assume that the activity probability is the same for all users, and users become active independently of each other. Moreover, we assume that the user activity remain unchanged during each channel coherence interval. At a given time instant, we use binary-valued a_i to describe the state of the ith user, namely, $a_i = 1$ represents that the ith user is active, while $a_i = 0$ denotes that it is inactive. Thus, $p(a_i = 1) = 1 - p(a_i = 0) = \lambda$.

We denote the equivalent channel coefficients from the BS to the entire set of users as $\mathbf{G}$, and use g_{ik} to represent the element on the ith row and kth column of $\mathbf{G}$. Then, it is obvious that $\mathbf{G} = \mathbf{G}_1 \mathbf{V} \mathbf{G}_2 + \mathbf{G}_3$. Besides, we denote $b_{ik} = a_i g_{ik}$ as the equivalent effective channel coefficient from the kth antenna at the BS to the ith user, and represent the aggregated equivalent effective channel coefficient from the BS to all users as $\mathbf{B}$, whose element on the ith row and kth column is b_{ik}.

Within each channel coherence interval, L symbols are used for uplink pilot transmission. Then, the received pilot signal at the BS can be expressed as

$$\mathbf{Y} = \mathbf{\Phi B} + \mathbf{N} \tag{7.1}$$

where $\mathbf{\Phi}$ is the pilot matrix whose elements are i.i.d. complex Gaussian random variables with zero mean and variance $1/N$, and $\mathbf{N} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_L)$ is the additive Gaussian noise at the BS. Following pilot transmission, the sparse activity detection is performed based on the received pilot signals at the BS.

7.2 Sparse activity detection

The GAMP algorithm, likelihood ratio test and the optimal fusion rule are employed to perform sparse activity detection in this section. With the GAMP algorithm, we determine the MMSE estimate for each element of the equivalent effective channel matrix, and acquire additive Gaussian noise corrupted versions of the equivalent effective channel coefficients. Then, multiple decisions on the activity of each user can be made using the likelihood ratio test and the Gaussian noise corrupted equivalent effective channel coefficients. At last, final decisions on the activity of all users are acquired based on the previous likelihood ratio test results, the corresponding detection reliabilities and the optimal fusion rule.

7.2.1 MMSE channel estimation with GAMP algorithm

We use $h_{1,ik}$, $h_{2,ik}$ and $h_{3,ik}$ to denote the elements on the ith row and kth column of $\mathbf{H}_1$, $\mathbf{H}_2$ and $\mathbf{H}_3$, respectively. Then, the equivalent channel coefficient g_{ik} can be expressed as

$$\begin{aligned} g_{ik} &= v\sqrt{\beta_{1,i}\beta_2}\sum_{m=1}^{M_1} h_{1,im}h_{2,mk}\exp(j\theta_m) + \sqrt{\beta_{3,i}}h_{3,ik} \\ &= \widetilde{g}_{ik} + \sqrt{\beta_{3,i}}h_{3,ik}. \end{aligned} \tag{7.2}$$

Since $h_{1,im}$ and $h_{2,mk}$ are both circularly symmetric complex Gaussian random variables with zero mean and unit variance and they are independent of each other, the components of the summation in (7.2) are independent and identically distributed with zero mean and variance $\beta_{1,i}\beta_2 v^2$. Therefore, according to the central limit theorem, the distribution of

$\frac{1}{\sqrt{M_1}}\widetilde{g}_{ik}$ tends to $\mathcal{CN}(0, v^2\beta_{1,i}\beta_2)$ as M_1 grows without bound. Thus, we approximately have $\widetilde{g}_{ik} \sim \mathcal{CN}(0, M_1 v^2\beta_{1,i}\beta_2)$ as M_1 gets large. Then, $g_{ik} \sim \mathcal{CN}(0, M_1 v^2\beta_{1,i}\beta_2+\beta_{3,i})$ as M_1 grows. In this chapter, we assume that M_1 is sufficiently large and we use $\mathcal{CN}(0, M_1 v^2\beta_{1,i}\beta_2+\beta_{3,i})$ to approximate the distribution of g_{ik}. Then, the probability density function of the equivalent effective channel coefficient is

$$p_{b_{ik}}(b) = (1-\lambda)\delta(b) + \lambda\mathcal{CN}(b; 0, M_1 v^2\beta_{1,i}\beta_2 + \beta_{3,i}). \tag{7.3}$$

With the GAMP algorithm, the received pilot signals are decomposed into scalar Gaussian noise corrupted versions which can be expressed as [56, 143]

$$r_{ik}^t = b_{ik} + \sqrt{\xi_{ik}^t}\widetilde{n} \tag{7.4}$$

where r_{ik}^t is the estimated scalar additive Gaussian noise corrupted equivalent effective channel coefficient during the tth iteration of the GAMP algorithm, ξ_{ik}^t is the corresponding noise variance and $\widetilde{n} \sim \mathcal{CN}(0,1)$ is the complex Gaussian noise.

It is obvious that $b_{ik} = 0$ when $a_i = 0$. Then, the probability density function of r_{ik}^t can be expressed as

$$p_{r_{ik}^t}(r|a_i = 0) = \frac{1}{\pi\xi_{ik}^t}\exp\left(-\frac{|r|^2}{\xi_{ik}^t}\right). \tag{7.5}$$

When $a_i = 1$, the probability density function of r_{ik}^t is

$$p_{r_{ik}^t}(r|a_i = 1) = \frac{\lambda}{\pi\left(M_1 v^2\beta_{1,i}\beta_2 + \beta_{3,i} + \xi_{ik}^t\right)}\exp\left(-\frac{|r|^2}{M_1 v^2\beta_{1,i}\beta_2 + \beta_{3,i} + \xi_{ik}^t}\right). \tag{7.6}$$

Therefore, the probability density function of r_{ik}^t can be expressed as

$$p_{r_{ik}^t}(r) = \frac{1-\lambda}{\pi\xi_{ik}^t}\exp\left(-\frac{|r|^2}{\xi_{ik}^t}\right) + \frac{\lambda}{\pi\left(M_1 v^2\beta_{1,i}\beta_2 + \beta_{3,i} + \xi_{ik}^t\right)}\exp\left(-\frac{|r|^2}{M_1 v^2\beta_{1,i}\beta_2 + \beta_{3,i} + \xi_{ik}^t}\right). \tag{7.7}$$

From (7.4), we can obtain that

$$p_{r_{ik}^t}\left(r|b_{ik}\right)=\frac{1}{\pi\xi_{ik}^t}\exp\left(-\frac{\left|r-b_{ik}\right|^2}{\xi_{ik}^t}\right). \tag{7.8}$$

Therefore, the MMSE estimate of b_{ik} is

$$\begin{aligned}\widehat{b}_{ik}^t=&E\left\{b_{ik}|r_{ik}^t\right\}\\=&\int b_{ik}p_{b_{ik}}\left(b|r_{ik}^t\right)db\\=&\frac{1}{p_{r_{ik}^t}\left(r\right)}\int b_{ik}p_{r_{ik}^t}\left(r|b\right)p_{b_{ik}}\left(b\right)db\\=&\frac{\left(M_1v^2\beta_{1,i}\beta_2+\beta_{3,i}\right)r_{ik}^t}{\left(M_1v^2\beta_{1,i}\beta_2+\beta_{3,i}+\xi_{ik}^t\right)\phi_1\left(r_{ik}^t\right)}\end{aligned} \tag{7.9}$$

where $\phi_1\left(r_{ik}^t\right)$ can be expressed as

$$\phi_1\left(r_{ik}^t\right)=1+\frac{\left(1-\lambda\right)\left(M_1v^2\beta_{1,i}\beta_2+\beta_{3,i}+\xi_{ik}^t\right)}{\lambda\xi_{ik}^t}\exp\left(-\frac{\left(M_1v^2\beta_{1,i}\beta_2+\beta_{3,i}\right)\left|r_{ik}^t\right|^2}{\xi_{ik}^t\left(M_1v^2\beta_{1,i}\beta_2+\beta_{3,i}+\xi_{ik}^t\right)}\right). \tag{7.10}$$

The first order derivative of $\widehat{b}_{ik}^t$ is

$$\frac{\partial}{\partial r_{ik}^t}\widehat{b}_{ik}^t=\frac{\left(M_1v^2\beta_{1,i}\beta_2+\beta_{3,i}\right)\omega_1\left(r_{ik}^t\right)}{\left(M_1v^2\beta_{1,i}\beta_2+\beta_{3,i}+\xi_{ik}^t\right)\phi_1\left(r_{ik}^t\right)} \tag{7.11}$$

where

$$\omega_1\left(r_{ik}^t\right)=1+\frac{\left(\phi_1\left(r_{ik}^t\right)-1\right)\left(M_1v^2\beta_{1,i}\beta_2+\beta_{3,i}\right)\left|r_{ik}^t\right|^2}{\xi_{ik}^t\left(M_1v^2\beta_{1,i}\beta_2+\beta_{3,i}+\xi_{ik}^t\right)\phi_1\left(r_{ik}^t\right)}. \tag{7.12}$$

We can perform sum-product GAMP algorithm to acquire the MMSE estimate of the equivalent effective channel coefficients by substituting (7.9) and (7.11) into [143, equations (12a) and (12b)], and acquire the additive Gaussian noise corrupted versions of the equivalent effective channel coefficients when the GAMP algorithm converges.

7.2.2 Likelihood ratio test and the optimal fusion rule

By now, we have determined the MMSE estimates and the additive Gaussian noise corrupted versions of the equivalent effective channel coefficients with the GAMP algorithm. Next, we employ a threshold detection rule to obtain decisions on the activity of all users and the corresponding reliabilities based on the likelihood ratio test and the additive Gaussian noise corrupted versions of the equivalent effective channel coefficients. Since the threshold detection rule is performed on every element of the equivalent effective channel matrix, and each row of the matrix corresponds to the equivalent effective channel coefficients from the BS to the same user, we can obtain M activity detection results for each user. Then, the final decisions on the activity of all users are made with the optimal fusion rule based on the detection results with likelihood ratio test for each user and the corresponding reliabilities.

The GAMP algorithm converges to a fixed point after sufficient number of iterations [52, 144]. We assume that the noise variance ξ_{ik}^t in (7.4) converges to ξ_0 within the GAMP algorithm. Thus, the scalar additive Gaussian noise corrupted model in (7.4) can be expressed as

$$r_{ik}^0 = b_{ik} + \sqrt{\xi_0}\widetilde{n} \tag{7.13}$$

where r_{ik}^0 stands for the converged value of r_{ik}^t via the GAMP algorithm. Then, the likelihood ratio test rule can be expressed as

$$\Lambda = \frac{p\left(r_{ik}^0 | a_i = 1\right)}{p\left(r_{ik}^0 | a_i = 0\right)} \underset{a_i=0}{\overset{a_i=1}{\gtrless}} \frac{1-\lambda}{\lambda}. \tag{7.14}$$

Substituting (7.5) and (7.6) into (7.14), we can obtain the following threshold detection rule:

$$\left|r_{ik}^0\right|^2 \underset{a_i=0}{\overset{a_i=1}{\gtrless}} r_0 \tag{7.15}$$

where r_0 can be expressed as

$$r_0 = \frac{\xi_0 \left(M_1 v^2 \beta_{1,i} \beta_2 + \beta_{3,i} + \xi_0\right)}{M_1 v^2 \beta_{1,i} \beta_2 + \beta_{3,i}} \log \frac{(1-\lambda)\left(M_1 v^2 \beta_{1,i} \beta_2 + \beta_{3,i} + \xi_0\right)}{\lambda \xi_0}. \tag{7.16}$$

Then, we can obtain the corresponding false alarm and miss detection probabilities as follows:

$$P_{ik,F} = \exp\left(-\frac{r_0}{\xi_0}\right) \tag{7.17}$$

and

$$P_{ik,M} = 1 - \exp\left(-\frac{r_0}{M_1 v^2 \beta_{1,i} \beta_2 + \beta_{3,i} + \xi_0}\right). \tag{7.18}$$

The optimal fusion rule proposed in [145] is used at the BS to acquire the final decisions on the activity of all users based on the likelihood ratio test results for each user and the corresponding reliabilities. Since the optimal fusion rule is similar for all users, for the sake of simplicity in description, we only consider the ith user here.

We denote the decisions that are made based on r_{ik}^0 by $\widehat{a}_{ik}$ $(k = 1, ..., M)$. Then, we divide the indices of the antennas into two sets, $\mathcal{S}_0 = \{k|\widehat{a}_{ik} = 0\}$ and $\mathcal{S}_1 = \{k|\widehat{a}_{ik} = 1\}$. Finally, the optimal fusion rule is expressed as [145]

$$\log\left(\frac{\lambda}{1-\lambda}\right) + \sum_{\mathcal{S}_1} \log\left(\frac{1-P_{ik,M}}{P_{ik,F}}\right) + \sum_{\mathcal{S}_0} \log\left(\frac{P_{ik,M}}{1-P_{ik,F}}\right) \underset{a_i=0}{\overset{a_i=1}{\gtrless}} 0. \tag{7.19}$$

By substituting (7.17) and (7.18) into (7.19), we can acquire the final detection rule for the activity of the ith user, which can be expressed as

$$\log\left(\frac{\lambda}{1-\lambda}\right) + \frac{r_0}{\xi_0} \sum_{\mathcal{S}_1} \frac{M_1 v^2 \beta_{1,i} \beta_2 + \beta_{3,i}}{M_1 v^2 \beta_{1,i} \beta_2 + \beta_{3,i} + \xi_0} + \sum_{\mathcal{S}_0} \log\left(\frac{1 - \exp\left(\frac{r_0}{M_1 v^2 \beta_{1,i} \beta_2 + \beta_{3,i} + \xi_0}\right)}{1 - \exp\left(-\frac{r_0}{\xi_0}\right)}\right) \underset{a_i=0}{\overset{a_i=1}{\gtrless}} 0. \tag{7.20}$$

7.3 Numerical Analysis

We assume that there are 500 users uniformly distributed in a circle with radius 500m. The IRS is located at the center of the circle, and the distance from the BS to the IRS is 100m. The reference distance for the path-loss between the BS, IRS and users is 50m. The path-loss values at the reference distance are 20dB, 15dB and 30dB for the channel coefficients between IRS and users, BS and IRS, and BS and users, respectively. And the corresponding path-loss decay exponents are 2, 2.5 and 5.

In this section, we use P_{err} to denote the error probability of the sparse activity detection method proposed in this chapter, i.e., $P_{err} = \lambda P_M + (1-\lambda)P_F$, where P_M represents the miss detection probability, and P_F stands for the false alarm probability. In certain scenarios, the direct link between the BS and users may not exist due to the unfavorable propagation conditions (e.g., blockages) [115]. Therefore, we also consider the IRS assisted network without a direct link from the BS to the users in the numerical results. In the figures, we use "without DL" to stand for numerical results of IRS assisted network without direct link, and "with DL" to represent the network scenario in which direct link exists.

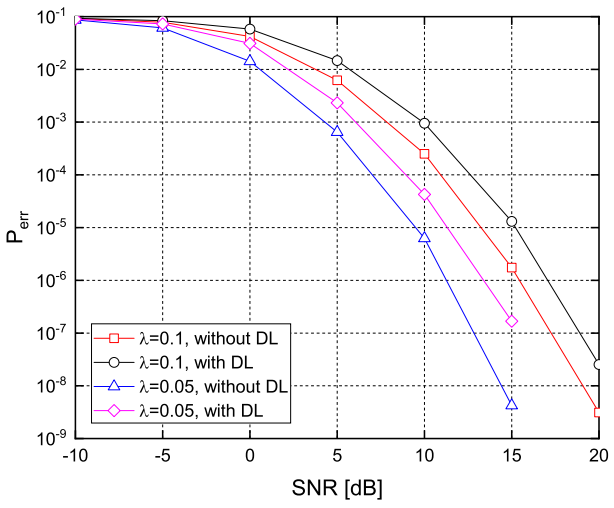

Figure 7.2: Average error probability of sparse activity detection versus SNR at the BS, $M = 50$, $M_1 = 10$, $L = 200$, $v = 1$.

Fig. 7.2 plots the curves for the average error probability of sparse activity detection versus SNR at the BS. The error probabilities expectedly diminish as the SNR increases. We note that more users become inactive on average when the activity probability decreases from 0.1 to 0.05, and the uncertainty in the equivalent effective channel coefficients becomes smaller. Consequently, as we observe in Fig. 2, the error probabilities get smaller as the activity probability decreases. Similarly, compared with the IRS assisted network without direct link, the case of IRS with direct link has larger uncertainty in the equivalent effective channel coefficients, leading to larger error probabilities in sparse activity detection.

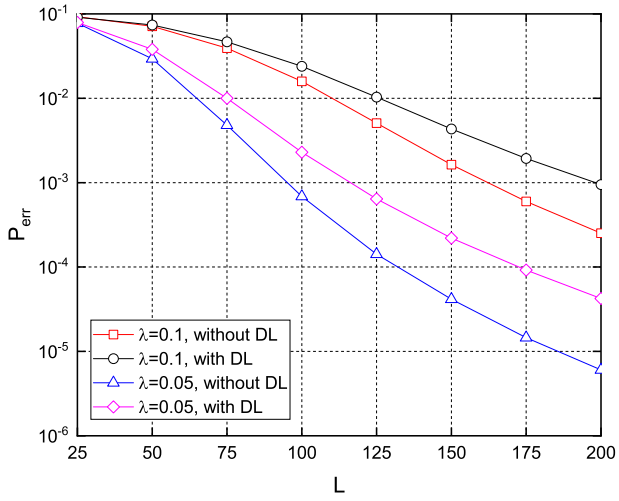

Figure 7.3: Average error probability of sparse activity detection versus number of pilots, SNR=10 dB, $M = 50$, $M_1 = 10$, $v = 1$.

Fig. 7.3 plots the average error probability versus the number of pilots used for sparse activity detection. The BS receives more pilot signals and it is more informed on the equivalent effective channel coefficients when L becomes larger. Thus, the error probabilities decrease as the number of pilots increases. When L is sufficiently large, the logarithm of the error probability diminishes linearly as L increases.

Fig. 7.4 plots the curves for the average error probability of sparse activity detection versus the number of antennas at the BS. For each user, we can obtain more activity detec-

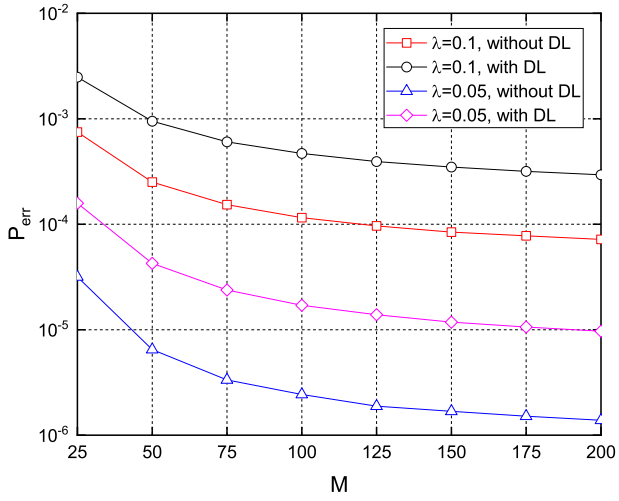

Figure 7.4: Average error probability of sparse activity detection versus number of antennas at the BS, SNR=10 dB, $M_1 = 10$, $L = 200$, $v = 1$.

tion results based on the MMSE estimate of the corresponding equivalent effective channel coefficients and the likelihood ratio test as M becomes larger. Thus, the final decision can be made with the optimal fusion rule based on increased knowledge. Therefore, the error probabilities diminish as M increases. On the other hand, more antennas at the BS will lead to more channel uncertainty which will increase the error probability. Consequently, the decay rate of error probability decreases as M increases, indicating diminishing returns as the number of antennas at the BS grows.

The curves for the average error probability of sparse activity detection versus the scattering amplitude of the elements at IRS are plotted in Fig. 7.5. We observe in this figure that the error probabilities almost keep unchanged as v increases for IRS assisted network without direct link, while the error probabilities decrease for IRS assisted network with direct link.

The curves for the error probabilities of sparse activity detection versus the number of elements at IRS are shown in Fig. 7.6. We notice that the error probabilities get smaller as M_1 increases for both cases of IRS assisted network with and without direct link. Compared

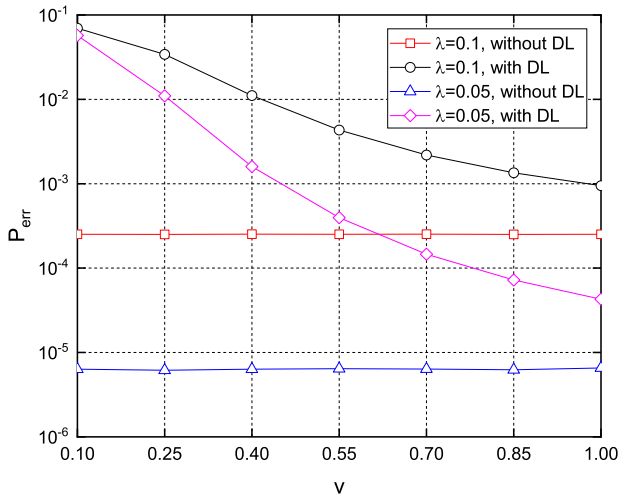

Figure 7.5: Average error probability of sparse activity detection versus the scattering amplitude of the elements at IRS, SNR=10 dB, $M = 50$, $M_1 = 10$, $L = 200$.

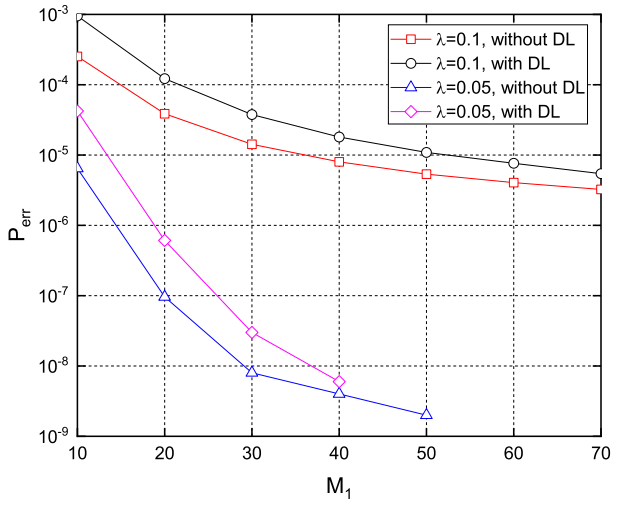

Figure 7.6: Average error probability of sparse activity detection versus the number of elements at IRS, SNR=10 dB, $M = 50$, $L = 200$, $v = 1$.

with the direct link, the IRS link becomes more and more significant as M_1 increases, and the difference between the error probabilities of these two cases becomes smaller as M_1 grows.

Chapter 8

Asymptotic Analysis of Weighted Sum Rate Maximization in Intelligent Reflecting Surface Assisted Wireless Networks

In this chapter, we conduct the asymptotic analysis of maximizing weighted sum rate under transmit power and QoS constraints in IRS assisted wireless networks. Utilizing the asymptotic results of the max-min SINR problem, we show that the transmit power constraint can be equivalently represented with an equality constraint on the SINR. Then, instead of optimizing over the beamforming matrix directly, the original problems are converted into the optimization over SINR, which dramatically promotes the solution of the original problems. After a transformation of the optimization variables, the weighted sum rate maximization problem is converted into a concave-convex optimization problem and it can be solved with the iterative approach of the concave-convex fractional programming algorithm proposed in [116]. Moreover, we show that the optimal power allocation scheme always makes full use of the available transmit power.

8.1 System Model

We consider the downlink of IRS assisted wireless network as shown in Fig. 8.1, which is comprised of one BS equipped with M antennas, one IRS and N single-antenna users. All users are uniformly distributed in a circle with radius R_r and the IRS is located at the center of the circle. The distance from the BS to the IRS is d_0. We assume that the direct link between the BS and users is neglected as a result of the unfavorable propagation conditions (e.g., due to blockages) [87].

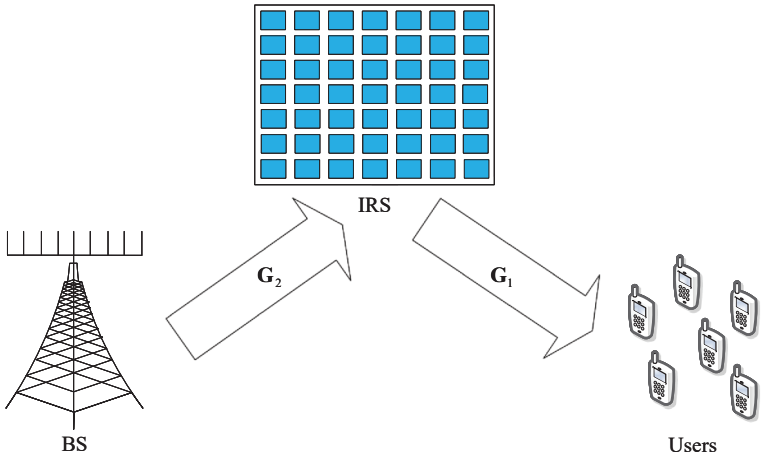

Figure 8.1: Downlink of the IRS assisted wireless network.

There are M_1 passive scattering elements at the IRS, and the scattering coefficients can be represented with a diagonal matrix $\mathbf{V} = \text{diag}(v_1 e^{j\theta_1}, v_2 e^{j\theta_2}, \cdots, v_{M_1} e^{j\theta_{M_1}})$, where $v_i \in [0, 1]$ and $\theta_i \in [0, 2\pi]$ are the amplitude and phase of the scattering coefficient at the ith element of IRS, respectively. We assume that $\mathbf{V}$ is reconfigurable and known at the BS [14, 93]. Moreover, we assume that the amplitude of the scattering coefficients at different elements are equal to each other, i.e., $v_1 = v_2 = \cdots = v_{M_1} = v$.

We assume that flat fading channels are experienced between the BS, IRS and users. $\mathbf{G}_1 = \text{diag}(\sqrt{\boldsymbol{\beta}_1})\mathbf{H}_1 \in \mathcal{C}^{N \times M_1}$ and $\mathbf{G}_2 = \sqrt{\beta_2}\mathbf{H}_2 \in \mathcal{C}^{M_1 \times M}$ stand for the channel coefficient matrices from the IRS to users and the BS to IRS, respectively. $\boldsymbol{\beta}_1 = [\beta_{1,1}, \beta_{1,2}, \cdots, \beta_{1,N}]$, where $\beta_{1,i} = \max\left(\beta_{1,0}, \beta_{1,0} + 10\alpha_1 \log_{10}\left(\frac{d_{1,i}}{d_{1,0}}\right)\right)$ stands for the path-loss (in dB) from the IRS to the ith user, $d_{1,0}$ denotes the reference distance for the path-loss between the IRS and users, $\beta_{1,0}$ is

the path-loss at $d_{1,0}$, α_1 is the corresponding path-loss decay exponent and $d_{1,i}$ is the distance from the IRS to the ith user. Similarly, $\beta_2 = \max\left(\beta_{2,0}, \beta_{2,0} + 10\alpha_2 \log_{10}\left(\frac{d_2}{d_{2,0}}\right)\right)$ represents the path-loss (in dB) from the BS to IRS, $d_{2,0}$, $\beta_{2,0}$, α_2 and d_2 denote the corresponding reference distance, minimum path-loss, path-loss decay exponent and the distance from the BS to IRS, respectively. $\mathbf{H}_1$ and $\mathbf{H}_2$ are the corresponding small-scale fading channel coefficient matrices whose elements are independent and identically distributed complex random variables with zero mean and unit variance.

We denote the equivalent channel coefficient matrix from the BS to all users as $\mathbf{G} = \mathbf{G}_1\mathbf{V}\mathbf{G}_2$. Then the received signal at the ith user is represented as

$$y_i = \mathbf{g}_i\mathbf{s} + n_i \tag{8.1}$$

where $\mathbf{g}_i \in \mathcal{C}^{1\times M}$ is the ith row of $\mathbf{G}$, which represents the equivalent channel coefficients from the BS to the ith user. $\mathbf{s} = \sum_{j=1}^{N} \mathbf{w}_j x_j \in \mathcal{C}^{M\times 1}$ is the transmitted signal at the BS, where x_j is the transmitted message intended for the jth user and $\mathbf{w}_j$ stands for the corresponding beamforming vector. $n_i \sim \mathcal{CN}(0, \sigma^2)$ denotes the additive Gaussian noise at the ith user, where σ^2 is the corresponding noise variance. We assume that the transmitted message power is $E\left\{|x_j|^2\right\} = 1$. The transmit power towards the jth user is denoted as $\frac{p_j}{N}$. The beamforming vector can be represented as $\mathbf{w}_j$, and the amplitude of $\mathbf{w}_j$ satisfies $\|\mathbf{w}_j\|^2 = \frac{p_j}{N}$, where $\|\cdot\|$ stands for the Frobenius norm. Then the aggregated beamforming matrix is $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N]$, and the total transmit power is $tr\left(\mathbf{W}\mathbf{W}^H\right)$, where $tr(\cdot)$ stands for the trace operation. We use g_{ij} to denote the element on the ith row and jth column of $\mathbf{G}$. Even for moderate number of IRS elements M_1, the distribution of g_{ij} can be approximated with a complex Gaussian distribution $\mathcal{CN}(0, \beta_i^e)$, where $\beta_i^e = M_1 v^2 \beta_{1,i}\beta_2$ can be regarded as the equivalent path-loss from the BS to the ith user [146].

The SINR at the ith user is

$$\gamma_i = \frac{\left|\mathbf{g}_i \mathbf{w}_i\right|^2}{\sum_{j=1, j\neq i}^{N} \left|\mathbf{g}_i \mathbf{w}_j\right|^2 + \sigma^2}, \tag{8.2}$$

and the corresponding achievable rate in bps/Hz can be expressed as

$$R_i = \log_2(1 + \gamma_i). \tag{8.3}$$

8.2 Preliminaries

For the completeness of this chapter, we provide several preliminary results here. The max-min weighted SINR problem under a given transmit power constraint can be expressed as follows:

$$\begin{aligned} &\max_{\mathbf{W}} \ \min_{i} \ \frac{\gamma_i}{\gamma_i^0} \\ &\quad s.t. \quad tr\left(\mathbf{W}\mathbf{W}^H\right) \leqslant P_{\max} \end{aligned} \tag{8.4}$$

where $\frac{1}{\gamma_i^0}$ $(i = 1, 2, \cdots, N)$ denotes the weight of the SINR of the ith user, and $P_{\max}$ represents the maximum transmit power.

The optimal solution of problem (8.4) is as following [114, 147]:

$$\mathbf{w}_i^* = \left(\sum_{j=1, j\neq i}^{N} \frac{q_j^*}{N} \mathbf{g}_j^H \mathbf{g}_j + \sigma^2 \mathbf{I}_M \right)^{-1} \mathbf{g}_i^H \tag{8.5}$$

where $\{q_i^*\}$ are the unique positive solutions to the following fixed point equations:

$$q_i^* = \frac{\gamma_i^0 \tau^*}{\frac{1}{N}\mathbf{g}_i \left(\sum\limits_{j=1, j\neq i}^{N} \frac{q_j^*}{N} \mathbf{g}_j^H \mathbf{g}_j + \sigma^2 \mathbf{I}_M \right)^{-1} \mathbf{g}_i^H} \tag{8.6}$$

and

$$\tau^* = \frac{NP_{\max}}{\sum_{i=1}^{N} \gamma_i^0 \left(\frac{1}{N} \mathbf{g}_i \left(\sum_{j=1, j\neq i}^{N} \frac{q_j^*}{N} \mathbf{g}_j^H \mathbf{g}_j + \sigma^2 \mathbf{I}_M \right)^{-1} \mathbf{g}_i^H \right)^{-1}}. \tag{8.7}$$

The optimal objective function value in (8.4) satisfies

$$\frac{\gamma_i^*}{\gamma_i^0} = \tau^*, \; i = 1, 2, \cdots N \tag{8.8}$$

with

$$\gamma_i^* = \frac{\frac{p_i^*}{N} \frac{\left| \mathbf{g}_i \mathbf{w}_i^* \right|^2}{\left\| \mathbf{w}_i^* \right\|^2}}{\sum_{j=1, j\neq i}^{N} \frac{p_j^*}{N} \frac{\left| \mathbf{g}_i \mathbf{w}_j^* \right|^2}{\left\| \mathbf{w}_j^* \right\|^2} + \sigma^2}. \tag{8.9}$$

Let us denote $\mathbf{p}^* = [p_1^*, p_2^*, \cdots, p_N^*]^T$, then

$$\mathbf{p}^* = \sigma^2 \tau^* (\mathbf{I}_N - \tau^* \mathbf{\Gamma F})^{-1} \mathbf{\Gamma 1}_N \tag{8.10}$$

where $\mathbf{\Gamma} = \text{diag}\left(\frac{N\gamma_i^0 \left\| \mathbf{w}_1^* \right\|^2}{\left| \mathbf{g}_1 \mathbf{w}_1^* \right|^2}, \frac{N\gamma_2^0 \left\| \mathbf{w}_2^* \right\|^2}{\left| \mathbf{g}_2 \mathbf{w}_2^* \right|^2}, \cdots, \frac{N\gamma_N^0 \left\| \mathbf{w}_N^* \right\|^2}{\left| \mathbf{g}_N \mathbf{w}_N^* \right|^2} \right)$, $\mathbf{F} \in \mathcal{C}^{N \times N}$ whose elements are given by:

$$[\mathbf{F}]_{i,j} = \begin{cases} 0 & \text{if } i = j \\ \frac{1}{N} \frac{\left| \mathbf{g}_i \mathbf{w}_j^* \right|^2}{\left\| \mathbf{w}_j^* \right\|^2} & \text{if } i \neq j \end{cases} \tag{8.11}$$

and $\mathbf{1}_N$ stands for a $N \times 1$ vector whose elements equal to 1.

We assume that the estimated channel after the channel estimation phase can be expressed as

$$\widehat{\mathbf{g}}_i = \sqrt{1 - \eta_{er}^2} \mathbf{g}_i + \sqrt{\beta_i^e} \eta_{er} \mathbf{e}_i \tag{8.12}$$

where $\eta_{er} \in [0, 1]$ describes the quality of channel estimation, and $\mathbf{e}_i \sim \mathcal{CN}(0, \mathbf{I}_M)$ is the channel estimation error which is independent of $\mathbf{g}_i$.

As described in [114], when only imperfect CSI is available at the BS, we cannot obtain the optimal solutions of the original problem. Instead, we simply replace $\{\mathbf{g}_i\}$ in the solutions

of (8.4) with $\{\widehat{\mathbf{g}}_i\}$, and denote the corresponding results of $\mathbf{w}_i^*$, q_i^*, γ_i^*, p_i^* and τ^* as $\widehat{\mathbf{w}}_i$, $\widehat{q}_i$, $\widehat{\gamma}_i$, $\widehat{p}_i$ and $\widehat{\tau}$.

Let us introduce the following assumptions for the asymptotic analysis performed in this chapter:

Assumption 1: We assume that both M and N grow without bound while their ratio is kept as a constant that is larger than 1, i.e., $\lim_{M,N\to\infty} \frac{M}{N} = \delta > 1$. Hence, this assumption addresses the massive MIMO setting with a large number of users.

Assumption 2: The equivalent path-loss of all users are upper bounded, i.e., $0 < \beta_i^e < \infty$.

Theorem 1 [114]: Denote the asymptotic results of $\widehat{\tau}$, $\widehat{p}_i$, $\widehat{q}_i$, $\widehat{\gamma}_i$ and $\widehat{\mathbf{w}}_i$ as $\overline{\tau}$, $\overline{p}_i$, $\overline{q}_i$, $\overline{\gamma}_i$ and $\overline{\mathbf{w}}_i$, respectively. Then, under the settings of Assumptions 1 and 2, $|\widehat{\tau} - \overline{\tau}| \to 0$, $|\widehat{p}_i - \overline{p}_i| \to 0$, $|\widehat{q}_i - \overline{q}_i| \to 0$, $|\widehat{\gamma}_i - \overline{\gamma}_i| \to 0$ and $|\widehat{\mathbf{w}}_i - \overline{\mathbf{w}}_i| \to 0$, where $\overline{\tau}$ is the unique positive solution to the following fixed point equation:

$$\overline{\tau} = \frac{P_{\max}}{\frac{\sigma^2}{N}\sum_{i=1}^{N}\frac{\gamma_i^0}{\beta_i^e}} \left(\delta - \frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^0\overline{\tau}}{1+\gamma_i^0\overline{\tau}}\right), \tag{8.13}$$

where $\overline{p}_i$ can be expressed as

$$\overline{p}_i = \frac{\gamma_i^0}{\beta_i^e}\frac{\overline{\tau}}{\xi}\left(\frac{\beta_i^e P_{\max}}{\left(1+\gamma_i^0\overline{\tau}\right)^2} + \sigma^2\right) \tag{8.14}$$

with

$$\xi = \delta - \frac{1}{N}\sum_{i=1}^{N}\frac{\left(\gamma_i^0\overline{\tau}\right)^2}{\left(1+\gamma_i^0\overline{\tau}\right)^2}, \tag{8.15}$$

$\overline{q}_i$ can be expressed as

$$\overline{q}_i = \frac{\gamma_i^0}{\beta_i^e}\frac{P_{\max}}{\frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^0}{\beta_i^e}}, \tag{8.16}$$

$\overline{\gamma}_i$ can be represented with

$$\overline{\gamma}_i = \frac{\overline{p}_i(1-\eta_{er}^2)\xi}{\mu_i P_{\max} + \frac{\sigma^2}{\beta_i^e}} \tag{8.17}$$

with

$$\mu_i = \frac{1 + 2\eta_{er}^2 \gamma_i^0 \overline{\tau} + \left(\eta_{er} \gamma_i^0 \overline{\tau}\right)^2}{\left(1 + \gamma_i^0 \overline{\tau}\right)^2}, \tag{8.18}$$

and $\overline{\mathbf{w}}_i$ can be represented as

$$\overline{\mathbf{w}}_i = \sqrt{\frac{\overline{p}_i}{N}} \frac{\overline{\mathbf{v}}_i}{\|\overline{\mathbf{v}}_i\|} \tag{8.19}$$

with

$$\overline{\mathbf{v}}_i = \left(\sum_{j=1}^{N} \frac{\overline{q}_j}{N} \widehat{\mathbf{g}}_j^H \widehat{\mathbf{g}}_j + \sigma^2 \mathbf{I}_M \right)^{-1} \widehat{\mathbf{g}}_i^H. \tag{8.20}$$

Theorem 1 provides the asymptotic results for the weighted max-min SINR problem under a given transmit power constraint, and establishes the theoretical foundation of our asymptotic analysis of maximizing the weighted sum rate under transmit power and QoS constraints in IRS assisted wireless networks.

8.3 Asymptotic Analysis of Weighted Sum Rate Maximization in IRS Assisted Wireless Networks

As it is difficult to solve the original problem (given in (8.21) below) by optimizing over the beamforming matrix directly, we seek to find an equivalent representation for the achievable SINR region under the transmit power constraint in this section, and then address the original problem via maximization over the SINR region.

8.3.1 Asymptotic Achievable SINR Region

The asymptotic expression of τ^*, $\overline{\tau}$, is the solution of the fixed point equation in (8.13), and the achievable SINR region under the transmit power constraint can be equivalently represented with the inverse weight sets, $\{\gamma_i^0\}$, under which the optimal value of the objective function in (8.4) satisfies $\tau^* \geq 1$, i.e., for a given set of $\{\gamma_i^0\}$, SINR region is achievable if $\tau^* \geq 1$, and otherwise, it is not achievable [111]. In this section, we will show that the

achievable SINR region under the transmit power constraint can be further equivalently represented with the weight sets of the weighted max-min SINR problem under the same constraint, leading to the result that the optimal value of the objective function in (8.4) satisfies $\overline{\tau} = 1$, as illustrated in Theorem 2 below.

Theorem 2: The achievable SINR region with joint beamforming and power allocation under a given transmit power constraint can be equivalently represented with the inverse weight sets of the max-min problem in (8.4) under the same constraint that makes its optimal objective function value satisfy $\overline{\tau} = 1$.

Proof 8.1 *The proof of Theorem 2 is conducted with the following three different cases:* $\overline{\tau} < 1$, $\overline{\tau} = 1$ *and* $\overline{\tau} > 1$.

First, since the minimum value of $\frac{\gamma_i}{\gamma_i^0}$ $(i = 1, 2, \cdots, N)$ *is maximized in (8.4), it is obvious that there exists at least one* γ_i *(minimum value of* $\frac{\gamma_i}{\gamma_i^0}$*) which could not achieve* γ_i^0 *when* $\overline{\tau} < 1$*, i.e., the weighted sets* $\{\gamma_i^0\}$ *could not be achieved by joint beamforming and power allocation under the same average transmit power constraint when* $\overline{\tau} < 1$.

Second, since the optimal objective function value in (8.4) satisfies $\frac{\gamma_i^*}{\gamma_i^0} = \overline{\tau}$ $(i = 1, 2, \cdots, N)$*, it is obvious that* $\gamma_i^* = \gamma_i^0$ *when* $\overline{\tau} = 1$*, i.e., the weighted sets* $\{\gamma_i^0\}$ *could be achieved by joint beamforming and power allocation under the same average transmit power constraint when* $\overline{\tau} = 1$.

Third, let us denote $\gamma_i^{00} = \gamma_i^0 \overline{\tau}$ $(i = 1, 2, \cdots, N)$ *when* $\overline{\tau} > 1$*. Since* $\overline{\tau}$ *satisfies the fixed point equation in (8.13), which can be expressed as*

$$\frac{\sigma^2}{N} \sum_{i=1}^{N} \frac{\gamma_i^0 \overline{\tau}}{\beta_i^e} = P_{\max} \left(\delta - \frac{1}{N} \sum_{i=1}^{N} \frac{\gamma_i^0 \overline{\tau}}{1 + \gamma_i^0 \overline{\tau}} \right),$$

i.e.,

$$\frac{\sigma^2}{N} \sum_{i=1}^{N} \frac{\gamma_i^{00}}{\beta_i^e} = P_{\max} \left(\delta - \frac{1}{N} \sum_{i=1}^{N} \frac{\gamma_i^{00}}{1 + \gamma_i^{00}} \right),$$

and it is obvious that $\overline{\tau}' = 1$ *is one solution of the fixed point equation in (8.13) when we regard* $\{\gamma_i^{00}\}$ *as the weighted set in (8.4).*

Denote the fixed point equation in (8.13) as a function of $\overline{\tau}$, namely,

$$f(\overline{\tau}) = \frac{P_{\max}}{\frac{\sigma^2}{N}\sum\limits_{i=1}^{N}\frac{\gamma_i^0}{\beta_i^e}}\left(\delta - \frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^0\overline{\tau}}{1+\gamma_i^0\overline{\tau}}\right) - \overline{\tau}.$$

Then, the solution of the fixed point equation in (8.13) satisfies $f(\overline{\tau}) = 0$. Since the derivative of $f(\overline{\tau})$ is negative, i.e.,

$$f'(\overline{\tau}) = -\frac{P_{\max}}{\sigma^2\sum\limits_{i=1}^{N}\frac{\gamma_i^0}{\beta_i^e}}\sum_{i=1}^{N}\frac{\gamma_i^0}{\left(1+\gamma_i^0\overline{\tau}\right)^2} - 1 < 0,$$

there exists at most one solution for the fixed point equation in (8.13). Therefore, the unique solution of the fixed point equation in (8.13) is $\overline{\tau}' = 1$ when we regard $\{\gamma_i^{00}\}$ as the weighted set in (8.4).

Above all, we can find another set of weights $\{\gamma_i^{00}\}$ which makes the optimal objective function value of (8.4) equals to one to equivalently represent the results leading by the weighted set $\{\gamma_i^0\}$ when $\overline{\tau} > 1$.

In summary, the achievable SINR region with joint beamforming and power allocation under average transmit power constraint could be equivalently represented with the weighted sets of the max-min problem in (8.4) under the same constraint that makes its optimal objective function value satisfies $\overline{\tau} = 1$. Theorem 2 is proved.

The condition under which the solution of the fixed point equation in (8.13) satisfies $\overline{\tau} = 1$ is given in the following Theorem.

Theorem 3: Under the settings of Assumption 1 and 2, the solution of the fixed point equation in (8.13) satisfies $\overline{\tau} = 1$ if and only if $\frac{1}{N}\sum\limits_{i=1}^{N}\frac{\gamma_i^0}{\beta_i^e} = \frac{P_{\max}}{\sigma^2}\left(\delta - \frac{1}{N}\sum\limits_{i=1}^{N}\frac{\gamma_i^0}{1+\gamma_i^0}\right)$.

Proof 8.2 *Sufficiency: From (8.13), it is easy to obtain that*

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^0}{\beta_i^e} = \frac{P_{\max}}{\sigma^2}\left(\delta - \frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^0\overline{\tau}}{1+\gamma_i^0\overline{\tau}}\right)$$

when $\overline{\tau}=1$. *Then, we can acquire*

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^0}{\beta_i^e}=\frac{P_{\max}}{\sigma^2}\left(\delta-\frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^0}{1+\gamma_i^0}\right).$$

Necessity: If $\frac{1}{N}\sum\limits_{i=1}^{N}\frac{\gamma_i^0}{\beta_i^e}=\frac{P_{\max}}{\sigma^2}\left(\delta-\frac{1}{N}\sum\limits_{i=1}^{N}\frac{\gamma_i^0}{1+\gamma_i^0}\right)$, *we can acquire*

$$\begin{aligned}\overline{\tau}&=\frac{P_{\max}}{\frac{\sigma^2}{N}\sum\limits_{i=1}^{N}\frac{\gamma_i^0}{\beta_i^e}}\left(\delta-\frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^0\overline{\tau}}{1+\gamma_i^0\overline{\tau}}\right)\\&=\left(\delta-\frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^0\overline{\tau}}{1+\gamma_i^0\overline{\tau}}\right)\left(\delta-\frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^0}{1+\gamma_i^0}\right)^{-1}.\end{aligned}$$

Then,

$$\overline{\tau}\left(\delta-\frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^0}{1+\gamma_i^0}\right)=\delta-\frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^0\overline{\tau}}{1+\gamma_i^0\overline{\tau}}.$$

Therefore,

$$(\overline{\tau}-1)\left(\delta-\frac{1}{N}\sum_{i=1}^{N}\frac{(\gamma_i^0)^2\overline{\tau}}{(1+\gamma_i^0)(1+\gamma_i^0\overline{\tau})}\right)=0.$$

Since we assume that $\delta>1$ *(Assumption 1) in this chapter,*

$$\delta-\frac{1}{N}\sum_{i=1}^{N}\frac{(\gamma_i^0)^2\overline{\tau}}{(1+\gamma_i^0)(1+\gamma_i^0\overline{\tau})}>0.$$

Thus,

$$\overline{\tau}-1=0,\ \textit{i.e.,}\ \overline{\tau}=1.$$

Theorem 3 provides an explicit condition for the achievable SINR sets $\{\gamma_i^0\}$ under a given transmit power constraint, and the corresponding transmit power and beamforming vectors for the achievable target SINR set $\{\gamma_i^0\}$ could be acquired by substituting $\{\gamma_i^0\}$ into (8.14) and (8.19), respectively.

Theorem 4: Under the settings of Assumption 1 and 2, the optimal average transmit

power of joint beamforming and power allocation equals $P_{\max}$, i.e., $\frac{1}{N}\sum_{i=1}^{N}\overline{p}_i = P_{\max}$.

Proof 8.3 *From (8.16), we can obtain that*

$$\frac{1}{N}\sum_{k=1}^{N}\frac{\gamma_k^0}{\beta_k^e} = \frac{\gamma_i^0 P_{\max}}{\beta_i^e \overline{q}_i}.$$

Substituting into (8.13), we can acquire

$$\frac{\gamma_i^0\overline{\tau}\sigma^2}{\beta_i^e} = \overline{q}_i\left(\delta - \frac{1}{N}\sum_{k=1}^{N}\frac{\gamma_k^0\overline{\tau}}{1+\gamma_k^0\overline{\tau}}\right).$$

Then,

$$\begin{aligned}\overline{p}_i &= \frac{P_{\max}}{\xi}\frac{\gamma_i^0\overline{\tau}}{\left(1+\gamma_i^0\overline{\tau}\right)^2} + \frac{1}{\xi}\frac{\gamma_i^0\overline{\tau}\sigma^2}{\beta_i^e}\\ &= \frac{P_{\max}}{\xi}\frac{\gamma_i^0\overline{\tau}}{\left(1+\gamma_i^0\overline{\tau}\right)^2} + \frac{\overline{q}_i}{\xi}\left(\delta - \frac{1}{N}\sum_{k=1}^{N}\frac{\gamma_k^0\overline{\tau}}{1+\gamma_k^0\overline{\tau}}\right).\end{aligned}$$

Finally,

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\overline{p}_i\\ =&\frac{P_{\max}}{N\xi}\sum_{i=1}^{N}\frac{\gamma_i^0\overline{\tau}}{\left(1+\gamma_i^0\overline{\tau}\right)^2} + \frac{1}{\xi}\left(\delta - \frac{1}{N}\sum_{k=1}^{N}\frac{\gamma_k^0\overline{\tau}}{1+\gamma_k^0\overline{\tau}}\right)\frac{1}{N}\sum_{i=1}^{N}\overline{q}_i\\ =&\frac{P_{\max}}{\xi}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^0\overline{\tau}}{\left(1+\gamma_i^0\overline{\tau}\right)^2} + \delta - \frac{1}{N}\sum_{k=1}^{N}\frac{\gamma_k^0\overline{\tau}}{1+\gamma_k^0\overline{\tau}}\right)\\ =&\frac{P_{\max}}{\xi}\cdot\xi\\ =&P_{\max}.\end{aligned}$$

In this section, we have shown that the asymptotic results for the average transmit power constraint for joint beamforming and power allocation could be equivalently represented with a correspond nonlinear equality constraint. Then, with the help of these results, we will per-

form some asymptotic analysis for the maximizing weighted sum rate by joint beamforming and power allocation under average transmit power and QoS constraints for both perfect and imperfect CSI at BS in the following sections.

8.3.2 Asymptotic Analysis of Weighted Sum Rate Maximization with Perfect CSI

In this subsection, we assume the availability of perfect CSI at the BS. As noted above, the transmit power constraint could be equivalently represented with a nonlinear equality constraint, and the variables of the optimization problem considered in this chapter are converted from the beamforming matrix to the SINR region, which dramatically promotes the solution process of the original problem.

The weighted sum rate maximization by joint beamforming and power allocation under transmit power and QoS constraints can be expressed as

$$\max_{\mathbf{W}} \sum\nolimits_{i=1}^{N} u_i R_i \tag{8.21a}$$

$$s.t.\ tr\left(\mathbf{W}\mathbf{W}^H\right) \leq P_{\max} \tag{8.21b}$$

$$\gamma_i \geq \gamma_i^{\min}, i = 1, 2, \cdots, N \tag{8.21c}$$

where u_i stands for the weight of the rate of the ith user.

As noted before, the transmit power constraint in (8.21b) can be equivalently represented with a nonlinear equality constraint as described in Theorem 3. Then, the optimization problem (8.21) can be converted into

$$\max_{\boldsymbol{\gamma}} \sum\nolimits_{i=1}^{N} u_i \log_2(1+\gamma_i) \tag{8.22a}$$

$$s.t.\ \frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i}{\beta_i^e} = \frac{P_{\max}}{\sigma^2}\left(\delta - \frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i}{1+\gamma_i}\right) \tag{8.22b}$$

$$\gamma_i \geq \gamma_i^{\min}, i = 1, 2, \cdots, N \tag{8.22c}$$

where $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \cdots, \gamma_N]^T$ denotes the set of SINR values for all users. By now, we have converted the optimization variable from the beamforming matrix $\mathbf{W}$ to the SINR set $\boldsymbol{\gamma}$. By changing the optimization variable to $x_i = \frac{1}{1+\gamma_i}$, problem (8.22) can be converted to the following:

$$\max_{\mathbf{x}} \sum\nolimits_{i=1}^{N} u_i \log_2 \left(\frac{1}{x_i}\right) \tag{8.23a}$$

$$s.t. \frac{1}{N}\sum_{i=1}^{N} \frac{1-x_i}{x_i \beta_i^e} - \frac{P_{\max}}{\sigma^2}\left(\delta - 1 + \frac{1}{N}\sum_{i=1}^{N} x_i\right) = 0 \tag{8.23b}$$

$$x_i - \frac{1}{1+\gamma_i^{\min}} \leq 0, i = 1, 2, \cdots, N. \tag{8.23c}$$

It is easy to determine that (8.23b) and (8.23c) are convex constraints. Since (8.23a) is the summation of a sequence of non-decreasing functions of $\frac{1}{x_i}$, the optimization problem (8.23) is equivalent to [116]

$$\max_{\mathbf{x},\mathbf{y}} \sum\nolimits_{i=1}^{N} u_i \log_2 \left(2y_i - x_i y_i^2\right) \tag{8.24a}$$

$$s.t. \frac{1}{N}\sum_{i=1}^{N} \frac{1-x_i}{x_i \beta_i^e} - \frac{P_{\max}}{\sigma^2}\left(\delta - 1 + \frac{1}{N}\sum_{i=1}^{N} x_i\right) = 0 \tag{8.24b}$$

$$x_i - \frac{1}{1+\gamma_i^{\min}} \leq 0, i = 1, 2, \cdots, N \tag{8.24c}$$

$$y_i \in \mathbb{R}, i = 1, 2, \cdots, N. \tag{8.24d}$$

Now, the maximization problem in (8.24) can be efficiently solved with the iterative approach of the concave-convex fractional programming algorithm proposed in [116]. We denote the optimal set of SINR values for (8.22) as $\widetilde{\boldsymbol{\gamma}}$.

By now, we have obtained the optimal set of SINR values for (8.22) as $\widetilde{\boldsymbol{\gamma}}$. Then, we need to find the optimal transmit power and beamforming vectors which could achieve the SINR set $\widetilde{\boldsymbol{\gamma}}$ under the given transmit power constraint. The problem of finding the optimal transmit power and beamforming vectors under the given transmit power constraint can be regarded

as a max-min weighted SINR problem under the same transmit power constraint, and the correspond solutions for the optimal transmit power and beamforming vectors are given in (8.14) and (8.19). Therefore, we can acquire the optimal transmit power and beamforming vectors for all users by substituting $\widetilde{\boldsymbol{\gamma}}$ into (8.14) and (8.19), respectively.

8.3.3 Asymptotic Analysis of Weighted Sum Rate Maximization with Imperfect CSI

Heretofore, we have investigated the asymptotic performance of the weighted sum rate maximization under transmit power and QoS constraints with perfect CSI at the BS in IRS assisted wireless networks. However, it is difficult to obtain the perfect CSI in practice. Therefore, we consider imperfect CSI at the BS in this subsection, and the estimated channel is described as in (8.12).

Theorem 1 shows that $\overline{\tau}$, $\overline{p}_i$, $\overline{q}_i$ and $\overline{\mathbf{w}}_i$ are independent of η_{er}, while $\overline{\gamma}_i$ is determined by η_{er}. When perfect CSI is available at the BS, we can obtain the optimal set of SINR values which maximize the weighted sum rate by solving the corresponding optimization problem in (8.24), and then acquire the corresponding optimal transmit power and beamforming vectors. However, as noted in Section III, we simply replace $\{\mathbf{g}_i\}$ with $\{\widehat{\mathbf{g}}_i\}$ to acquire $\widehat{\mathbf{w}}_i$, $\widehat{q}_i$, $\widehat{\tau}$, $\widehat{\gamma}_i$ and $\widehat{p}_i$ when there is channel estimation error at the BS, which leads to the result that $\overline{\gamma}_i \neq \gamma_i^0 \overline{\tau}$ in this situation. Therefore, instead of substituting the solution of (8.24) into (8.14) directly to obtain the optimal transmit power, we need to solve an optimization problem which regards the transmit power levels as the optimization variables when imperfect CSI is acquired at the BS. Then the optimal beamforming vectors could be obtained by substituting the optimal transmit power into (8.19).

Assuming that the optimal solution of (8.24) is $\widetilde{\boldsymbol{\gamma}} = [\widetilde{\gamma}_1, \widetilde{\gamma}_2, \cdots, \widetilde{\gamma}_N]^T$, and substituting it into (8.15) and (8.18), we can obtain the corresponding $\widetilde{\xi}$ and $\widetilde{\mu}_i$. Then, substituting $\widetilde{\xi}$ and $\widetilde{\mu}_i$ into (8.17), we can acquire the SINR of the ith user under imperfect CSI which can be

expressed as

$$\widehat{\gamma}_i = \frac{\widehat{p}_i(1-\eta_{er}^2)\widetilde{\xi}}{\widetilde{\mu}_i P_{\max} + \frac{\sigma^2}{\beta_i^e}}. \tag{8.25}$$

Then, the weighted sum rate maximization under transmit power and QoS constraints can be formulated as

$$\max_{\widehat{\mathbf{p}}} \sum\nolimits_{i=1}^{N} u_i \log_2\left(1 + \frac{\widehat{p}_i(1-\eta_{er}^2)\widetilde{\xi}}{\widetilde{\mu}_i P_{\max} + \frac{\sigma^2}{\beta_i^e}}\right) \tag{8.26a}$$

$$s.t.\ \frac{1}{N}\sum_{i=1}^{N}\widehat{p}_i \le P_{\max} \tag{8.26b}$$

$$\frac{\widehat{p}_i(1-\eta_{er}^2)\widetilde{\xi}}{\widetilde{\mu}_i P_{\max} + \frac{\sigma^2}{\beta_i^e}} \ge \gamma_i^{\min}, i = 1, 2, \cdots, N \tag{8.26c}$$

where $\widehat{\boldsymbol{p}} = [\widehat{p}_1, \widehat{p}_2, \cdots, \widehat{p}_N]^T$ stands for the transmit power vector for all users. It is easy to determine that (8.26a) is a concave function, while (8.26b) and (8.26c) are convex constraints. If we regard $\widehat{p}_i$ as a fractional function whose denominator equals to 1, (8.26a) can be regarded as the summation of non-decreasing functions of $\widehat{p}_i$. Similarly, with the results in [116], (8.26) is equivalent to

$$\max_{\widehat{\mathbf{p}},\mathbf{y}} \sum\nolimits_{i=1}^{N} u_i \log_2\left[1 + \frac{(1-\eta_{er}^2)\widetilde{\xi}}{\widetilde{\mu}_i P_{\max} + \frac{\sigma^2}{\beta_i^e}}\left(2y_i\sqrt{\widehat{p}_i} - y_i^2\right)\right] \tag{8.27a}$$

$$s.t.\ \frac{1}{N}\sum_{i=1}^{N}\widehat{p}_i \le P_{\max} \tag{8.27b}$$

$$\frac{\widehat{p}_i(1-\eta_{er}^2)\widetilde{\xi}}{\widetilde{\mu}_i P_{\max} + \frac{\sigma^2}{\beta_i^e}} \ge \gamma_i^{\min}, i = 1, 2, \cdots, N \tag{8.27c}$$

$$y_i \in \mathbb{R}, i = 1, 2, \cdots, N. \tag{8.27d}$$

Then, the maximization problem in (8.27) can be efficiently solved with the iterative approach of the concave-convex fractional programming algorithm proposed in [116]. Similar to the case for perfect CSI, the optimal beamforming vector under imperfect CSI can be acquired by substituting the optimal transmit power vector, $\widehat{\boldsymbol{p}}^*$, into (8.19).

8.4 Numerical Analysis

In this section, we conduct a numerical analysis. We assume that all users are uniformly distributed in a circle with radius 500m. There are 10 passive scattering elements at the IRS, and the reflecting coefficient amplitudes of all elements are equal to 1. The minimum SINR requirement for each user is -20 dB. The maximum transmit power is $P_{\max} = 10$ Watts. The reference distance for the channel coefficients is $d_{1,0} = d_{2,0} = 50$m, and the path-loss at the reference distance is 20 dB and 15 dB for the channel coefficients from the IRS to the users and from the BS to the IRS, respectively. The path-loss decay exponents for the channel coefficients from the IRS to the users and from the BS to the IRS are 2 and 2.5, respectively. The weights for each user are assumed to be one. We consider the performance in terms of the weighted sum rate here, and the sum rate is normalized over the bandwidth (and has the units of bps/Hz) in this chapter.

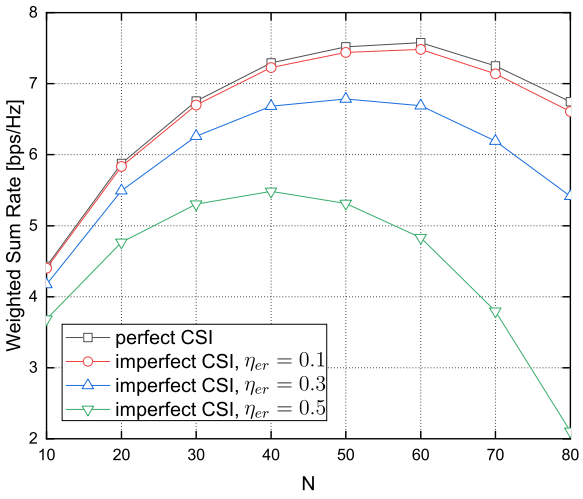

Figure 8.2: Weighted sum rate versus the number of users under perfect and imperfect CSI, $M = 100$ and SNR=20 dB.

Fig. 8.2 depicts the relationship between the weighted sum rate and the number of users under perfect and imperfect CSI[1]. We observe that the weighted sum rate initially

[1]Note that in Fig. 8.2, M is fixed and N varies. In such a case, analytical characterizations are applied

increases at small number of users and then starts decreasing as N grows beyond a certain threshold. For small values of N, the given transmit power is large enough to support all users and the weighted sum rate increases at first as N grows. However, the limited available transmit power presents a challenge when the user number is large. In order to fulfill the QoS requirements of all users, transmit power is partially allocated to the users with worse channel conditions instead of being assigned to users with better channels, leading to the decrease in the weighted sum rate when the user number is large. Additionally, these curves also show that under the same settings, the achievable weighted sum rate decreases as the channel estimation error grows. Indeed, the loss in the weighted sum rate is seen to get worse as η_{er} increases.

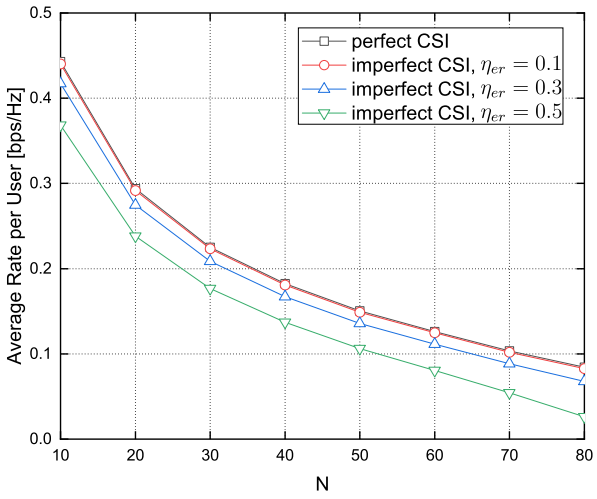

Figure 8.3: Average rate per user versus the number of users under perfect and imperfect CSI, $M = 100$ and SNR=20 dB.

Fig. 8.3 plots the curves illustrating the relationship between the achievable average rate per user and the number of users under perfect and imperfect CSI. In order to fulfill the QoS requirements of all users, the transmit power allocation becomes more and more inefficient and the achievable average rate per user diminishes as the number of users grows. Under

with different $\delta = \frac{M}{N}$ to each point in the plot.

the same settings, we again note that the achievable average rate per user decreases as the channel estimation error increases.

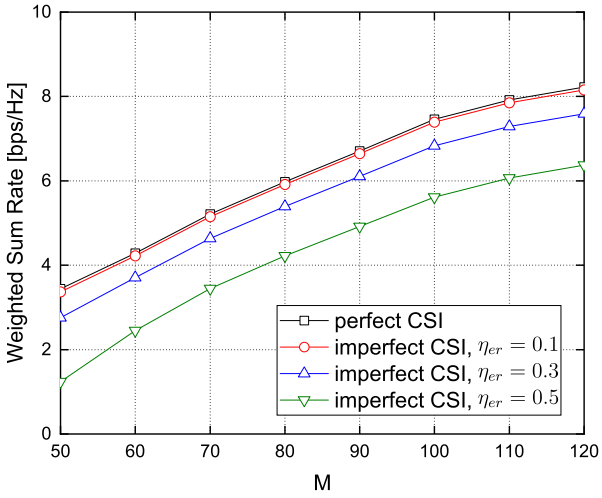

Figure 8.4: Weighted sum rate versus the number of antennas at the BS under perfect and imperfect CSI, $N = 40$ and SNR=20 dB.

Fig. 8.4 plots the weighted sum rate vs. the number of antennas at the BS under perfect and imperfect CSI. Higher degrees of freedom and larger diversity gains are obtained and the achievable weighted sum rate increases as the number of antennas at the BS increases. Similarly as before, the achievable weighted sum rate decreases with the increasing channel estimation error, and degradation in the weighted sum rate is observed to accelerate as the estimation quality drops.

Fig. 8.5 depicts the relationship between the weighted sum rate and SNR under perfect and imperfect CSI. These curves show that the weighted sum rate increases with growing SNR while it expectedly decreases as the channel estimation error increases.

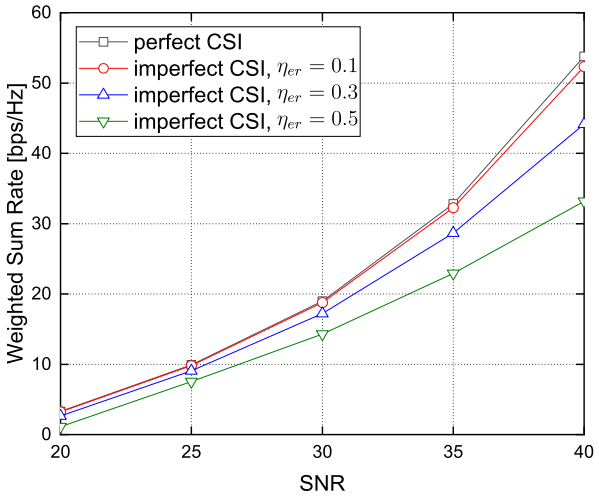

Figure 8.5: Weighted sum rate versus SNR at user side under perfect and imperfect CSI, $N = 40$ and $M = 50$.

Chapter 9

Conclusion

9.1 Summary

In this thesis, we have studied the performance of wireless networks with massive connectivity. The contributions of this thesis are summarized below.

In Chapter 2, we proposed a novel joint antenna and user selection method for single-cell massive MIMO communication systems. With the two-step iterative procedure, we can obtain the EE-optimal subset of antennas and users, based on the bisection search, random selection and cross-entropy algorithms. We have demonstrated that with joint antenna and user selection, the EE of the system could be improved significantly, especially when M and K are large.

In Chapter 3, Algorithm 3.2 is proposed to solve the joint uplink and downlink EE maximization problem with joint antenna selection and user scheduling in single-cell massive MIMO systems, under a limitation on the number of available RF chains. With Jensen's inequality and the power consumption model, the original joint antenna selection and user scheduling problem is converted into a combinatorial optimization problem, and we have shown that it can be solved efficiently with the developed learning-based stochastic gradient descent algorithm. We have also employed the rare event simulation method in the learning-

based stochastic gradient descent algorithm to generate samples with very small probabilities. We have considered both perfect and imperfect CSI at the BS. Via numerical results, we have provided insightful observations that are poised to be beneficial in the design of practical single-cell massive MIMO systems.

In Chapter 4, we have derived a decoupling principle for SMV based MMSE estimation of sparse signal vectors with i.n.i.d. non-zero components in cell-free massive MIMO networks with massive connectivity. With this decoupling principle and the likelihood ratio test, we have obtained a detection rule for the activity of users based on the received pilot signals at only one AP. Subsequently, with the decoupling principle of the SMV based MMSE estimation, likelihood ratio test and the optimal fusion rule, we have determined detection rules for the activity of users based on the cooperation with the received pilot signals at all APs for centralized and distributed detection. We have also demonstrated that the error probabilities of both centralized and distributed detection tend to zero when the number of AP tends to infinity while the asymptotic ratio between the number of users and pilots is kept constant. Moreover, we have analyzed oracle estimation in cell-free massive MIMO networks with massive connectivity, and the asymptotic behavior of oracle estimation is identified via random matrix theory. Via numerical analysis, we have investigated the impact of the number of pilots, SNR, and the number of APs on mean square error and error probabilities of different schemes. We have also observed that the theoretical analysis with the decoupling principle of SMV based MMSE estimation matches well with the numerical results of the CB-AMP algorithm.

Performance of cell-free massive MIMO systems with massive connectivity is analyzed in Chapter 5. In the joint activity detection and channel estimation phase, MMSE estimates of the effective channel coefficients from all users to the entire set of APs are obtained via the GAMP algorithm. Then, the effective noise variance is determined with the state evolution equations of the GAMP algorithm. Following this characterization, the variances of both the estimated channel and channel estimation error are identified. Finally, the achievable uplink

rates in the data transmission phase with ZF detector deployed at the CPU are analyzed and the impact of the number of pilots, the number of APs, SNR, and the user activity detection and channel estimation results on the performance is determined.

In Chapter 6, we have analyzed the MMSE channel estimation in IRS assisted wireless communication systems. We have first identified statistics of the end-to-end channel matrix. Specifically, we have first shown that each row vector of the equivalent channel matrix from the BS to the users has a Bessel K distribution, and all the rows are independent of each other. Following this characterization, we have employed a GSM model, and obtained an analytical closed-form expression of the MMSE estimate of the equivalent channel. Furthermore, we have derived analytical upper and lower bounds of the mean square error. We have also provided an asymptotic analysis of the MMSE estimation, and shown that the upper bound of the mean square error of the MMSE estimate equals the asymptotic mean square error of the MMSE estimation as M_1 gets large. Furthermore, from the derived expressions of the MMSE estimate of the equivalent channel and the corresponding upper and lower bound expressions, we notice that the mean square error is independent of the number of antennas at the BS. Moreover, since the row vectors of the equivalent channel are independent of each other and orthogonal pilots are used during the channel estimation process, the mean square error does not depend on the number of users either. Via numerical analysis, we have identified how the mean square error varies as a function of SNR, the number of elements and the scattering amplitudes at the IRS. We have also demonstrated that the upper and lower bounds lead to very accurate approximations of the mean square error.

The sparse activity detection in IRS assisted wireless network is considered in Chapter 7. Via the GAMP algorithm, we have obtained the MMSE estimates and the additive Gaussian noise corrupted versions of the equivalent effective channel coefficients. Subsequently, we have obtained multiple decisions on the activity of each user based on the Gaussian noise corrupted equivalent effective channel coefficients and the likelihood ratio test. Finally, the decisions on the activity of all users are made by employing the optimal fusion rule. Following

the theoretical characterizations, we conducted a numerical analysis. Via numerical results, we have investigated the impact of system and network parameters, such as SNR, the number of pilots, number of BS antennas, number of IRS elements, on the average error probability of sparse activity detection. In particular, we observe that the average error probability diminishes as the SNR, number of pilots, number of antennas at the BS or number of elements at the IRS increases.

In Chapter 8, we have conducted the asymptotic analysis of weighted sum rate maximization under transmit power and QoS constraints in IRS assisted wireless networks. Making use of the asymptotic results on the max-min weighted SINR problem, the transmit power constraint is equivalently represented with a nonlinear equality constraint, and the original problem is optimized over the SINR region instead of being optimized over the beamforming vectors directly, which dramatically promotes the solution process of the original problems. Finally, the optimization problem is converted into a concave-convex optimization problem via a transformation of the variables, after which it can be efficiently solved with the iterative method of the concave-convex fractional programming algorithm.

9.2 Future Research Directions

9.2.1 Joint Antenna Selection and User Scheduling in Multi-Cell Massive MIMO Networks

In Chapters 2 and 3, we have studied the EE maximization problem by joint antenna selection and user scheduling in single-cell massive MIMO systems. However, since a single cell could support only a limited area, it is more practical to consider the joint antenna selection and user scheduling in multi-cell massive MIMO networks. When multiple cells are considered, inter-cell interference will be introduced, potentially leading to additional power consumption for multi-cell cooperation. These aspects will make the analysis more challenging.

9.2.2 Joint Activity Detection and Channel Estimation in Cell-Free Massive MIMO Networks With User-Dependent and Temporally Correlated Activity Probabilities

In Chapters 4 and 5, we assume that the users' activity probabilities are independent and equal to each other to perform the joint activity detection and channel estimation. With these assumptions, the CB-AMP algorithm is employed to obtain the MMSE estimate of the effective channel coefficients. Finally, users' activities are obtained with the likelihood ratio test and the optimal fusion rule, and the channel coefficients are acquired with the results of users' activities and the MMSE estimates of the effective channel coefficients.

In practice, there might be different types of users in IoT networks, and the users' activity probabilities are not equal to each other. Moreover, some of the users' activity probabilities might be correlated as a result of their characteristics or being physically close to each other. Furthermore, users might have temporal correlations in their activities, i.e., activity probability can depend on the previous state of the user, and can be described via a Markovian model. These are some practical settings for the users' activity probabilities in IoT networks, and the joint activity detection and channel estimation will become more challenging when these conditions are considered.

9.2.3 Performance Analysis of IRS Assisted Wireless Networks with Different Reflecting Amplitudes at the Scattering Elements

In Chapters 6, 7 and 8, we have studied the performance analysis of IRS assisted wireless networks, e.g., MMSE channel estimation, asymptotic analysis and joint activity detection and channel estimation. In these works, we assumed that the reflecting amplitudes at the passive scattering elements of IRS are equal to each other. Then, we find that the equivalent channel coefficients follow a Bessel K distribution, and it can be equivalently represented

with a GSM model. However, when the passive scattering elements at IRS have different reflecting amplitudes, it is challenging to find the distribution of the equivalent channel coefficients, which is a fundamental result needed for the further performance analysis in IRS assisted wireless network. We leave this as a topic of our future research work.

Bibliography

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

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

[3] M. M. Rana and M. K. Hosain, "Adaptive channel estimation techniques for mimo ofdm systems," *International Journal of Advanced Computer Science and Applications*, vol. 1, no. 6, 2010. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2010.010620

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

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

[6] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive MIMO: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99 878–99 888, 2019.

[7] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO: Uniformly great service for everyone," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2015, pp. 201–205.

[8] "Cisco annual internet report (2018–2023) white paper," 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[9] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1890–1904, April 2018.

[10] L. Liu and W. Yu, "Massive connectivity with massive MIMO—part I: Device activity detection and channel estimation," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2933–2946, June 2018.

[11] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Reconfigurable intelligent surfaces: Three myths and two critical questions," *IEEE Communications Magazine*, vol. 58, no. 12, pp. 90–96, December 2020.

[12] J. Zhao, "A survey of intelligent reflecting surfaces (IRSs): Towards 6G wireless communication networks," 2019, arXiv:1907.04789. [Online]. Available: https://arxiv.org/abs/1907.04789

[13] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Transactions on Communications*, pp. 1–39, 2021.

[14] S. Gong, X. Lu, D. T. Hoang, D. Niyato, L. Shu, D. I. Kim, and Y. C. Liang, "Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Communications Surveys Tutorials*, vol. 22, no. 4, pp. 2283–2314, 2020.

[15] M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 3, pp. 990–1002, 2020.

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

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

[18] Z. Liu, W. Du, and D. Sun, "Energy and spectral efficiency tradeoff for Massive MIMO systems with transmit antenna selection," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4453–4457, May 2017.

[19] M. T. A. Rana, R. Vesilo, and I. B. Collings, "Antenna selection in Massive MIMO using non-central principal component analysis," in *2016 26th International Telecommunication Networks and Applications Conference (ITNAC)*, Dec 2016, pp. 283–288.

[20] M. Benmimoune, E. Driouch, W. Ajib, and D. Massicotte, "Feedback reduction and efficient antenna selection for Massive MIMO system," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, Sept 2015, pp. 1–6.

[21] S. Qin, G. Li, G. Lv, G. Zhang, and H. Hui, "$L_{1/2}$-Regularization based antenna selection for RF-chain limited Massive MIMO systems," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Sept 2016, pp. 1–5.

[22] T. H. Tai, W. H. Chung, and T. S. Lee, "A low complexity antenna selection algorithm for energy efficiency in Massive MIMO systems," in *2015 IEEE International Conference on Data Science and Data Intensive Systems*, Dec 2015, pp. 284–289.

[23] M. Arash, E. Yazdian, M. S. Fazel, G. G. de Oliveira Brante, and M. A. Imran, "Employing antenna selection to improve energy efficiency in Massive MIMO systems," *Trans. Emerging Telecommunications Technologies*, vol. 28, 2017.

[24] X. Gao, O. Edfors, J. Liu, and F. Tufvesson, "Antenna selection in measured Massive MIMO channels using convex optimization," in *2013 IEEE Globecom Workshops (GC Wkshps)*, Dec 2013, pp. 129–134.

[25] M. Arash, E. Yazdian, and M. Fazel, "Antenna selection: A novel approach to improve energy efficiency in Massive MIMO systems," in *2016 6th International Conference on Computer and Knowledge Engineering (ICCKE)*, Oct 2016, pp. 106–110.

[26] B. M. Lee, J. Choi, J. Bang, and B. C. Kang, "An energy efficient antenna selection for large scale green MIMO systems," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, May 2013, pp. 950–953.

[27] K. Elkhalil, A. Kammoun, T. Y. Al-Naffouri, and M. S. Alouini, "A blind antenna selection scheme for single-cell uplink Massive MIMO," in *2016 IEEE Globecom Workshops (GC Wkshps)*, Dec 2016, pp. 1–6.

[28] H. Liu, H. Gao, S. Yang, and T. Lv, "Low-complexity downlink user selection for Massive MIMO systems," *IEEE Systems Journal*, vol. 11, no. 2, pp. 1072–1083, June 2017.

[29] G. Alyami and I. Kostanic, "A low complexity user selection scheme with linear precoding for Massive MIMO systems," in *International Journal of Computer Science*, 2016.

[30] S. Maimaiti, G. Chuai, W. Gao, K. Zhang, X. Liu, and Z. Si, "A low-complexity algorithm for the joint antenna selection and user scheduling in multi-cell multi-user downlink massive mimo systems," *EURASIP Journal on Wireless Communications*

and Networking, vol. 2019, no. 1, p. 208, 2019. [Online]. Available: https://doi.org/10.1186/s13638-019-1529-7

[31] Y. Dong, Y. Tang, and K. Z. Shenzhen, "Improved joint antenna selection and user scheduling for Massive MIMO systems," in *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, May 2017, pp. 69–74.

[32] M. Benmimoune, E. Driouch, W. Ajib, and D. Massicotte, "Joint transmit antenna selection and user scheduling for Massive MIMO systems," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2015, pp. 381–386.

[33] G. Xu, A. Liu, W. Jiang, H. Xiang, and W. Luo, "Joint user scheduling and antenna selection in distributed Massive MIMO systems with limited backhaul capacity," *China Communications*, vol. 11, no. 5, pp. 17–30, May 2014.

[34] M. Rana, L. Li, and S. W. Su, "Distributed state estimation over unreliable communication networks with an application to smart grids," *IEEE Transactions on Green Communications and Networking*, vol. 1, no. 1, pp. 89–96, March 2017.

[35] S. U. Pillai, T. Suel, and Seunghun Cha, "The perron-frobenius theorem: some of its applications," *IEEE Signal Processing Magazine*, vol. 22, no. 2, pp. 62–75, March 2005.

[36] B. M. Lee and H. Yang, "Massive MIMO with massive connectivity for industrial internet of things," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 6, pp. 5187–5196, June 2020.

[37] J. Yuan, Q. He, M. Matthaiou, T. Q. S. Quek, and S. Jin, "Toward massive connectivity for IoT in Mixed-ADC distributed massive MIMO," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1841–1856, March 2020.

[38] V. Boljanovic, D. Vukobratovic, P. Popovski, and C. Stefanovic, "User activity detection in massive random access: Compressed sensing vs. coded slotted ALOHA," in

IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2017, pp. 1–6.

[39] F. Monsees, C. Bockelmann, and A. Dekorsy, "Compressed sensing Neyman-Pearson based activity detection for sparse multiuser communications," in *10th International ITG Conference on Systems, Communications and Coding*, Feb 2015, pp. 1–6.

[40] A. Fengler, G. Caire, P. Jung, and S. Haghighatshoar, "Massive MIMO unsourced random access," 2019, arXiv:1901.00828. [Online]. Available: https://arxiv.org/abs/1901.00828.

[41] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the internet of things," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 88–99, Sep. 2018.

[42] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity in cellular networks: Multi-cell cooperation VS. large-scale antenna arrays," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 6618–6622.

[43] P. Han, R. Niu, M. Ren, and Y. C. Eldar, "Distributed approximate message passing for sparse signal recovery," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2014, pp. 497–501.

[44] P. Han, R. Niu, and M. Ren, "Distributed approximate message passing for compressed sensing," 2014, arXiv:1404.3766. [Online]. Available: https://arxiv.org/abs/1404.3766.

[45] R. Hayakawa, A. Nakai, and K. Hayashi, "Distributed approximate message passing with summation propagation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 4104–4108.

[46] T. Jiang, Y. Shi, J. Zhang, and K. B. Letaief, "Joint activity detection and channel estimation for IoT networks: Phase transition and computation-estimation tradeoff," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6212–6225, Aug 2019.

[47] S. Xia and Y. Shi, "Intelligent reflecting surface for massive device connectivity: Joint activity detection and channel estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5175–5179.

[48] S. Xia, Y. Shi, Y. Zhou, and X. Yuan, "Reconfigurable intelligent surface for massive connectivity," 2021, arXiv:2101.10322. [Online]. Available: https://arxiv.org/abs/2101.10322v1.

[49] T. Ding, X. Yuan, and S. C. Liew, "Sparsity learning-based multiuser detection in grant-free massive-device multiple access," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3569–3582, 2019.

[50] X. Yang, S. Jin, C.-K. Wen, X. Li, and J. Xue, "Joint activity detection and channel estimation for massive connectivity network with 1-bit DAC," in *11th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2019, pp. 1–6.

[51] X. Bian, Y. Mao, and J. Zhang, "Joint activity detection, channel estimation, and data decoding for grant-free massive random access," 2021, arXiv:2107.05246. [Online]. Available: https://arxiv.org/abs/2107.05246.

[52] F. Caltagirone, L. Zdeborov, and F. Krzakala, "On convergence of approximate message passing," in *IEEE International Symposium on Information Theory*, June 2014, pp. 1812–1816.

[53] S. Rangan, P. Schniter, A. K. Fletcher, and S. Sarkar, "On the convergence of approximate message passing with arbitrary matrices," 2014, arXiv:1402.3210. [Online]. Available: https://arxiv.org/abs/1402.3210.

[54] S. C. Birgmeier and N. Goertz, "Robust approximate message passing for nonzero-mean sensing matrices," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 4898–4902.

[55] A. Manoel, F. Krzakala, E. Tramel, and L. Zdeborov, "Swept approximate message passing for sparse estimation," in *Proceedings of the 32nd International Conference on Machine Learning*, Jul 2015, pp. 1123–1132.

[56] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *2011 IEEE International Symposium on Information Theory Proceedings*, July 2011, pp. 2168–2172.

[57] X. Meng, S. Wu, L. Kuang, and J. Lu, "Concise derivation of complex Bayesian approximate message passing via expectation propagation," 2015, arXiv:1509.08658. [Online]. Available: https://arxiv.org/abs/1509.08658.

[58] J. Vila, P. Schniter, S. Rangan, F. Krzakala, and L. Zdeborov, "Adaptive damping and mean removal for the generalized approximate message passing algorithm," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 2021–2025.

[59] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, Feb 2011.

[60] G. Hannak, A. Perelli, N. Goertz, G. Matz, and M. E. Davies, "Performance analysis of approximate message passing for distributed compressed sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 5, pp. 857–870, Oct 2018.

[61] D. Guo and S. Verdu, "Randomly spread CDMA: asymptotics via statistical physics," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1983–2010, June 2005.

[62] J. Zhu, D. Baron, and F. Krzakala, "Performance limits for noisy multimeasurement vector problems," *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2444–2454, May 2017.

[63] W. Huleihel and N. Merhav, "Asymptotic MMSE analysis under sparse representation modeling," *Signal Processing*, vol. 131, pp. 320 – 332, 2017.

[64] A. M. Tulino, G. Caire, S. Verdu, and S. Shamai, "Support recovery with sparsely sampled free random matrices," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4243–4271, July 2013.

[65] Z. Chen and E. Bjrnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5205–5219, Nov 2018.

[66] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and M. Debbah, "Cell-free massive MIMO with limited backhaul," in *IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–7.

[67] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and H. V. Poor, "Mixed quality of service in cell-free massive MIMO," *IEEE Communications Letters*, vol. 22, no. 7, pp. 1494–1497, July 2018.

[68] G. Interdonato, H. Q. Ngo, E. G. Larsson, and P. Frenger, "On the performance of cell-free massive MIMO with short-term power constraints," in *IEEE 21st International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD)*, Oct 2016, pp. 225–230.

[69] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, March 2017.

[70] S. Buzzi and A. Zappone, "Downlink power control in user-centric and cell-free massive MIMO wireless networks," in *IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct 2017, pp. 1–6.

[71] M. N. Boroujerdi, A. Abbasfar, and M. Ghanbari, "Cell free massive MIMO with constrained coverage," *Wireless Personal Communications*, vol. 97, no. 1, pp. 333–348, 2017.

[72] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and H. Yang, "Cell-free massive MIMO systems," in *49th Asilomar Conference on Signals, Systems and Computers*, Nov 2015, pp. 695–699.

[73] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Communications Letters*, vol. 21, no. 8, pp. 1871–1874, Aug 2017.

[74] H. Q. Ngo, L. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 1, pp. 25–39, March 2018.

[75] 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.

[76] P. Liu, K. Luo, D. Chen, and T. Jiang, "Spectral efficiency analysis of cell-free massive MIMO systems with zero-forcing detector," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 795–807, Feb 2020.

[77] J. Zhang, Y. Wei, E. Bjrnson, Y. Han, and X. Li, "Spectral and energy efficiency of cell-free massive MIMO systems with hardware impairments," in *9th International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct 2017, pp. 1–6.

[78] M. Protter, I. Yavneh, and M. Elad, "Closed-form MMSE estimation for signal denoising under sparse representation modeling over a unitary dictionary," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3471–3484, July 2010.

[79] E. G. Larsson and Y. Selen, "Linear regression with a sparse parameter vector," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 451–460, Feb 2007.

[80] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided MIMO communication," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1823–1838, 2020.

[81] T. Van Chien, L. T. Tu, S. Chatzinotas, and B. Ottersten, "Coverage probability and ergodic capacity of intelligent reflecting surface-enhanced communication systems," *IEEE Communications Letters*, vol. 25, no. 1, pp. 69–73, 2021.

[82] B. Di, H. Zhang, L. Song, Y. Li, Z. Han, and H. V. Poor, "Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1809–1822, 2020.

[83] H. Yu, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "Joint design of reconfigurable intelligent surfaces and transmit beamforming under proper and improper gaussian signaling," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2589–2603, 2020.

[84] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5849–5863, 2020.

[85] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.

[86] ——, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 7830–7833.

[87] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, Aug 2019.

[88] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1719–1734, 2020.

[89] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Latency minimization for intelligent reflecting surface aided mobile edge computing," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2666–2682, 2020.

[90] A. S. Abdalla, T. F. Rahman, and V. Marojevic, "UAVs with reconfigurable intelligent surfaces: Applications, challenges, and opportunities," 2020, arXiv:2012.04775. [Online]. Available: https://arxiv.org/abs/2012.04775.

[91] Y. Zhu, G. Zheng, and K. K. Wong, "Stochastic geometry analysis of large intelligent surface-assisted millimeter wave networks," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1749–1762, 2020.

[92] H. Guo, Y. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for intelligent reflecting surface enhanced wireless networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[93] Q. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M. Alouini, "Intelligent reflecting surface-assisted multi-user MISO communication: Channel esti-

mation and beamforming design," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 661–680, 2020.

[94] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user MIMO systems," 2019, arXiv:1912.03619. [Online]. Available: https://arxiv.org/abs/1912.03619

[95] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Intelligent reflecting surface assisted wireless communication: Modeling and channel estimation," 2019, arXiv:1906.02360. [Online]. Available: https://arxiv.org/abs/1906.02360.

[96] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Transactions on Communications*, pp. 1–1, 2021.

[97] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5000–5004.

[98] T. Lin, X. Yu, Y. Zhu, and R. Schober, "Channel estimation for intelligent reflecting surface-assisted millimeter wave MIMO systems," 2020, arXiv:2005.04720. [Online]. Available: https://arxiv.org/abs/2005.04720.

[99] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Processing Letters*, vol. 27, pp. 905–909, 2020.

[100] H. Liu, X. Yuan, and Y.-J. A. Zhang, "Matrix-calibration-based cascaded channel estimation for reconfigurable intelligent surface assisted multiuser MIMO," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2621–2636, Nov 2020.

[101] Z. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 210–214, 2020.

[102] M. Jung, W. Saad, Y. Jang, G. Kong, and S. Choi, "Reliability analysis of large intelligent surfaces (LISs): Rate distribution and outage probability," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1662–1666, 2019.

[103] ——, "Performance analysis of large intelligent surfaces (LISs): Asymptotic data rate and channel hardening effects," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2052–2065, 2020.

[104] B. Al-Nahhas, Q.-U.-A. Nadeem, and A. Chaaban, "Intelligent reflecting surface assisted MISO downlink: Channel estimation and asymptotic analysis," in *IEEE Global Communications Conference*, 2020, pp. 1–6.

[105] L.-N. Tran, M. Juntti, M. Bengtsson, and B. Ottersten, "Weighted sum rate maximization for MIMO broadcast channels using dirty paper coding and zero-forcing methods," *IEEE Transactions on Communications*, vol. 61, no. 6, pp. 2362–2373, 2013.

[106] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3064–3076, 2020.

[107] W. Tabikh, D. Slock, and Y. Yuan-Wu, "Weighted sum rate maximization of MISO interference broadcast channels via difference of convex functions programming: A large system analysis," in *IEEE Statistical Signal Processing Workshop (SSP)*, 2016, pp. 1–5.

[108] ——, "A large system analysis of weighted sum rate maximization of single stream MIMO interference broadcast channels under linear precoding," in *International Symposium on Wireless Communication Systems (ISWCS)*, 2016, pp. 6–10.

[109] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, 2011.

[110] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, 2008.

[111] L. Liu, R. Zhang, and K.-C. Chua, "Achieving global optimality for weighted sum-rate maximization in the K-user gaussian interference channel with multiple antennas," *IEEE Transactions on Wireless Communications*, vol. 11, no. 5, pp. 1933–1945, 2012.

[112] C. W. Tan, M. Chiang, and R. Srikant, "Maximizing sum rate and minimizing MSE on multiuser downlink: Optimality, fast algorithms and equivalence via max-min SINR," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 6127–6143, 2011.

[113] D. W. H. Cai, T. Q. S. Quek, and C. W. Tan, "A unified analysis of max-min weighted SINR for MIMO downlink system," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3850–3862, 2011.

[114] H. Sifaou, A. Kammoun, L. Sanguinetti, M. Debbah, and M.-S. Alouini, "Max-min SINR in large-scale single-cell MU-MIMO: Asymptotic analysis and low-complexity transceivers," *IEEE Transactions on Signal Processing*, vol. 65, no. 7, pp. 1841–1854, 2017.

[115] Q. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M. Alouini, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2020.

[116] K. Shen and W. Yu, "Fractional programming for communication systems–Part I: Power control and beamforming," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2616–2630, 2018.

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

[118] Z. I. Botev, D. P. Kroese, R. Y. Rubinstein, and P. L'Ecuyer, "The cross-entropy method for optimization," in *Handbook of statistics*. Elsevier, 2013, vol. 31, pp. 35–59.

[119] A. Berny, *Statistical Machine Learning and Combinatorial Optimization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 287–306.

[120] S.-K. Au and J. L. Beck, "Estimation of small failure probabilities in high dimensions by subset simulation," *Probabilistic Engineering Mechanics*, vol. 16, no. 4, pp. 263 – 277, 2001.

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

[122] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated rician mimo channels with rician disturbance," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1807–1820, March 2010.

[123] K. M. Zuev, *Subset Simulation Method for Rare Event Estimation: An Introduction*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2021, pp. 1–25.

[124] A. Berny, "Selection and reinforcement learning for combinatorial optimization," in *International Conference on Parallel Problem Solving from Nature*. Springer, 2000, pp. 601–610.

[125] M. Gallagher and M. Frean, "Population-based continuous optimization, probabilistic modelling and mean shift," *Evolutionary Computation*, vol. 13, no. 1, pp. 29–42, March 2005.

[126] K. Zhang, A. Koppel, H. Zhu, and T. Basar, "Global convergence of policy gradient methods to (almost) locally optimal policies," 2019.

[127] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1902–1923, March 2012.

[128] M. Afshang and H. S. Dhillon, "Fundamentals of modeling finite wireless networks using binomial point process," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 3355–3370, May 2017.

[129] A. M. Mathai, *An introduction to geometrical probability: distributional aspects with applications*. Amsterdam, Holland: Gordon and Breach, 1999.

[130] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 08, p. P08009, Aug 2012.

[131] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends® in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[132] D. Guo, Y. Wu, S. S. Shitz, and S. Verdu, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2371–2385, April 2011.

[133] M. Guo, M. C. Gursoy, and P. K. Varshney, "Sparse activity detection in cell-free massive MIMO systems," in *IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 1177–1182.

[134] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, no. 1, pp. 98–101, Jan 1986.

[135] M. Guo, M. C. Gursoy, and P. K. Varshney, "Sparse activity detection in cell-free massive MIMO systems," in *2020 IEEE International Symposium on Information Theory (ISIT)*, June 2020, pp. 1177–1182.

[136] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[137] L. Liu and W. Yu, "Massive connectivity with massive MIMO—part II: Achievable rate characterization," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2947–2959, June 2018.

[138] D. Kudathanthirige, D. Gunasinghe, and G. Amarasuriya, "Performance analysis of intelligent reflective surfaces for wireless communication," 2020, arXiv:2002.05603. [Online]. Available: https://arxiv.org/abs/2002.05603

[139] S. Gong, X. Lu, D. T. Hoang, D. Niyato, L. Shu, D. I. Kim, and Y. C. Liang, "Towards smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Communications Surveys & Tutorials (Early Access)*, 2020, doi: 10.1109/COMST.2020.3004197.

[140] R. K. Mallik and N. C. Sagias, "Distribution of inner product of complex Gaussian random vectors and its applications," *IEEE Transactions on Communications*, vol. 59, no. 12, pp. 3353–3362, 2011.

[141] P. A. Khazron and I. W. Selesnick, "Bayesian estimation of Bessel K form random vectors in AWGN," *IEEE Signal Processing Letters*, vol. 15, pp. 261–264, 2008.

[142] J. T. Flam, S. Chatterjee, K. Kansanen, and T. Ekman, "On MMSE estimation: A linear model under Gaussian mixture statistics," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3840–3845, 2012.

[143] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," 2010, arXiv:1010.5141. [Online]. Available: https://arxiv.org/abs/1010.5141.

[144] S. Rangan, P. Schniter, A. K. Fletcher, and S. Sarkar, "On the convergence of approximate message passing with arbitrary matrices," 2018, arXiv:1402.3210. [Online]. Available: https://arxiv.org/abs/1402.3210.

[145] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, no. 1, pp. 98–101, 1986.

[146] M. Guo and M. C. Gursoy, "Performance analysis of cell-free massive MIMO systems with massive connectivity," in *IEEE 18th Annual Consumer Communications Networking Conference (CCNC)*, 2021, pp. 1–6.

[147] D. W. H. Cai, T. Q. S. Quek, C. W. Tan, and S. H. Low, "Max-min SINR coordinated multipoint downlink transmission-duality and algorithms," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5384–5395, 2012.

Vita

Mangqing Guo received the B.S. degree in measuring and control technology and instrumentation program from Jilin University, Changchun, China, in 2013 and the M.S. degree in electronics science and technology from Beijing University of Posts and Telecommunications, Beijing, China, in 2016. He received the Ph.D. degree in the Department of Electrical Engineering and Computer Science, Syracuse University, in 2022. His research interests include massive MIMO networks, Internet-of-Things (IoT) and intelligent reflecting surface (IRS) assisted wireless communications.