Performance Analyses for Large-Scale Antennas Equipped Two-Way AF Relaying and Heterogeneous Networks

by

Yongyu Dai
B.Sc., Southeast University, China, 2010
M.Sc., Southeast University, China, 2013

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Electrical and Computer Engineering

Performance Analyses for Large-Scale Antennas Equipped Two-Way AF Relaying and Heterogeneous Networks

by

Yongyu Dai
B.Sc., Southeast University, China, 2010
M.Sc., Southeast University, China, 2013

Supervisory Committee

Dr. Xiaodai Dong, Supervisor
(Department of Electrical and Computer Engineering)

Dr. Wu-Sheng Lu, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Julie Zhou, Outside Member
(Department of Mathematics and Statistics)

Supervisory Committee

Dr. Xiaodai Dong, Supervisor
(Department of Electrical and Computer Engineering)

Dr. Wu-Sheng Lu, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Julie Zhou, Outside Member
(Department of Mathematics and Statistics)

ABSTRACT

In this dissertation, performance analyses for large-scale antennas equipped two-way amplify-and-forward (AF) relaying and heterogeneous network (HetNet) are carried out. Energy-efficiency oriented design becomes more important for the next generation of wireless systems, which motivates us to study the strong candidates, such as massive multiple-input multiple-output (MIMO) combined with cooperative relaying and HetNet. Based on the achievable rate analyses for both massive MIMO two-way AF relaying, effective power allocation schemes are presented to further improve system performance. Focusing on the MIMO downlinks in the HetNet, mean square error (MSE) based precoding schemes are designed and employed by the macro base station (BS) and the small cell (SC) nodes. Considering a HetNet where both macro BS and SC nodes are equipped with large-scale antenna arrays, the capacity lower bounds are derived, followed by the proposed user scheduling algorithms.

The work on multi-pair two-way AF relaying with linear processing considers a system where multiple sources exchange information via a relay equipped with massive antennas. Given that channel estimation is non-ideal, and that the relay employs either maximum-ratio combining/maximum-ratio transmission (MRC/MRT) or zero-forcing reception/zero-forcing transmission (ZFR/ZFT) beamforming, we derive two corresponding closed-form lower bound expressions for the ergodic achievable rate of each pair sources. The closed-form expressions enable us to design an optimal power allocation (OPA) scheme that maximizes the sum spectral

efficiency under certain practical constraints. As the antenna array size tends to infinity and the signal to noise ratios become very large, asymptotically optimal power allocation schemes in simple closed-form are derived. The capacity lower bounds are verified to be accurate predictors of the system performance by simulations, and the proposed OPA outperforms equal power allocation (EPA). It is also found that in the asymptotic regime, when MRC/MRT is used at the relay and the link end-to-end large-scale fading factors among all pairs are equal, the optimal power allocated to a user is inverse to the large-scale fading factor of the channel from the user to the relay, while OPA approaches EPA when ZFR/ZFT is adopted.

The work on the MSE-based precoding design for MIMO downlinks investigates a HetNet system consisting of a macro tier overlaid with a second tier of SCs. First, a new sum-MSE of all users based minimization problem is proposed aiming to design a set of macro cell (MC) and SC transmit precoding matrices or vectors. To solve it, two different algorithms are presented. One is via a relaxed-constraints based alternating optimization (RAO) realized by efficient alternating optimization and relaxing non-convex constraints to convex ones. The other is via an unconstrained alternating optimization with normalization (UAON) implemented by introducing the constraints into the iterations with the normalization operation. Second, a separate MSE minimization based two-level precoder is proposed by considering the signal and interference terms corresponding to the macro tier and the individual SCs separately. Furthermore, robust precoders are designed correspondingly with estimated imperfect channel. Simulation results show that the sum-MSE based RAO algorithm provides the best MSE performance among the proposed schemes under a number of system configurations. When the number of antennas at the macro-BS is sufficiently large relative to the number of MUEs, the MSE of the separate MSE-based precoding is found to approach those of RAO and UAON. Together, this thesis provides a suite of three new precoding techniques that is expected to meet the need in a broad range of HetNet environments with balance between performance and complexity.

The work on a large-scale HetNet studies the performance for MIMO downlink systems where both macro BS and SC nodes are equipped with large-scale antenna arrays. Suppose that the large-scale antenna arrays at both macro BS and SC nodes employ MRT or ZFT precoding, and transmit data streams to the served users simultaneously. A new pilot reuse pattern among small cells is proposed for channel estimation. Taking into account imperfect CSI, lower capacity bounds for MRT and ZFT are derived, respectively, in closed-form expressions involving only statistical CSI. Then asymptotic analyses for massive arrays are presented, from which we obtain the optimal antenna number ratio between BS and SCs under specific

power scaling laws. Subsequently, two user scheduling algorithms, that is, greedy scheduling algorithm and asymptotical scheduling algorithm (ASA), are proposed based on the derived capacity lower bounds and asymptotic analyses, respectively. ASA is demonstrated to be a near optimal user scheduling scheme in the asymptotic regime and has low complexity. Finally, the derived closed-form achievable rate expressions are verified to be accurate predictors of the system performance by Monte-Carlo simulations. Numerical results demonstrate the effectiveness of the asymptotic analysis and the proposed user scheduling schemes.

Contents

Supervisory Committee ii

Abstract iii

List of Tables ix

List of Figures x

Notations xii

Abbreviations xiii

Acknowledgements xv

Dedication xvi

1 Overview 1

1.1 Research Areas Overview 1

1.1.1 Multi-Pair Two-Way Relaying with Large-Scale Antennas . . 1

1.1.2 Interference Mitigation for Heterogeneous Networks 2

1.1.3 Heterogeneous Networks with Large-Scale Antennas 3

1.2 Dissertation Organization 3

2 Power Allocation for Multi-Pair Massive MIMO Two-Way AF Relaying with Linear Processing 4

2.1 Introduction 4

2.2 Related Work 5

2.3 Contributions 6

2.4 System Model 6

2.4.1 Channel Estimation 8

2.4.2 Data Transmission 9

2.4.3 MRC/MRT Processing 10

2.4.4 ZFR/ZFT Processing . . . 11
2.5 Achievable Rate Analysis . . . 11
2.5.1 Asymptotic Analysis with Massive Arrays . . . 14
2.6 Power Allocation Schemes . . . 15
2.6.1 Optimal Power Allocation (OPA) . . . 16
2.6.2 Asymptotically Optimal Power Allocation (AOPA) . . . 18
2.7 Numerical Results . . . 20
2.7.1 Validation of Achievable Rate Results . . . 20
2.7.2 Power Allocation . . . 24
2.8 Conclusion . . . 26
2.9 Appendices . . . 26
2.9.1 Proof of Equation (2.10) . . . 26
2.9.2 Proof of Equation (2.13) . . . 29
2.9.3 Proof of Theorem 2.1 . . . 30
2.9.4 Proof of Theorem 2.2 . . . 32

3 MSE-based Precoding for MIMO Downlinks in Heterogeneous Networks 35
3.1 Introduction . . . 35
3.2 Related Work . . . 36
3.3 Contributions . . . 37
3.4 System Model . . . 38
3.5 Sum-MSE Minimization Based Precoding in HetNet . . . 40
3.5.1 Relaxed-constraints based Alternating Optimization (RAO) 42
3.5.2 Unconstrained Alternating Optimization with Normalization (UAON) . . . 47
3.6 Separate MSE Minimization Based Two-level Precoding in HetNet . 49
3.6.1 MSE Minimization at the BS . . . 49
3.6.2 MSE Minimization at each SC . . . 52
3.7 Robust Precoding Design With Imperfect CSI in HetNet . . . 52
3.7.1 Robust RAO With Imperfect CSI . . . 53
3.7.2 Robust UAON With Imperfect CSI . . . 54
3.7.3 Robust Non-iterative Algorithm With Imperfect CSI . . . 54
3.8 Simulation Results . . . 56
3.9 Conclusion . . . 62
3.10 Appendices . . . 62
3.10.1 Proof of Equation (3.25) . . . 62

4 Performance Analysis for Pilot-Reused HetNets with Large-Scale Antenna Arrays in Downlink Systems **64**
4.1 Introduction . . . 64
4.2 Related Work . . . 65
4.3 Contributions . . . 65
4.4 System Model . . . 67
4.4.1 Channel Estimation with Pilot Reuse . . . 68
4.4.2 Data Transmission . . . 70
4.4.3 MRT Precoding . . . 70
4.4.4 ZFT Precoding . . . 71
4.5 Achievable Rate Analysis . . . 71
4.5.1 MRT Precoding . . . 71
4.5.2 ZFT Precoding . . . 75
4.5.3 Asymptotic Analysis with Massive Arrays . . . 77
4.6 User Scheduling Algorithms . . . 80
4.6.1 Greedy Scheduling Algorithm (GSA) . . . 80
4.6.2 Asymptotic Scheduling Algorithm (ASA) . . . 81
4.7 Numerical Results . . . 84
4.7.1 Comparison between One-Tier and Two-Tier Network Topologies . . . 84
4.7.2 Validation of Lower Capacity Bounds and Pilot Reuse Pattern 85
4.7.3 User Scheduling . . . 88
4.8 Conclusion . . . 91
4.9 Appendices . . . 92
4.9.1 Proof of (4.10) . . . 92
4.9.2 Proof of Theorem 4.1 . . . 93
4.9.3 Proof of Theorem 4.2 . . . 97

5 Conclusions and Further Research Issues **100**
5.1 Conclusions . . . 100
5.2 Further Research Issues . . . 102

Bibliography **105**

List of Tables

3.1 Simulation Parameters 56

4.1 Simulation Parameters 84

List of Figures

2.1 System diagram of multi-pair two-way AF relaying. 7
2.2 Spectral efficiency versus SNR for lower bounds and Monte-Carlo results ($p_{\mathrm{P}} = 10$ dB, EPA, $P_{\mathrm{R}} = 2KP_{\mathrm{S}}$). 21
2.3 Spectral efficiency versus K for lower bounds and Monte-Carlo results ($p_{\mathrm{P}} = 10$ dB, EPA, SNR $= P_{\mathrm{R}} = 2KP_{\mathrm{S}}$, $N = 128$). 22
2.4 Spectral efficiency versus K for lower bounds and Monte-Carlo results ($p_{\mathrm{P}} = 10$ dB, EPA, SNR $= P_{\mathrm{R}} = 2KP_{\mathrm{S}} = 0$ dB). 22
2.5 Required user power to achieve 1 bit/s/Hz per user for MRC/MRT and ZFR/ZFT (EPA, $P_{\mathrm{R}} = 2KP_{\mathrm{S}}$). 23
2.6 Spectral efficiency versus p_{P} ($P_0 = 10$ dB, $P_{\mathrm{R},0} = 23$ dB, $P = 23$ dB). 25
2.7 Spectral efficiency versus N for different power allocation schemes ($P_{\mathrm{R}} = P/2$). 25

3.1 System model for HetNet with SCs deployment. 38
3.2 The average MSE per data stream learning curve over 100 runs ($K = 8$, $P_{\mathrm{BS}} = 46$ dBm, $\bar{\sigma}_h^2 = 1$). 57
3.3 The average MSE per data stream for MUE/SUE versus transmit power at BS ($K = 8$, Perfect CSI). 57
3.4 The BER per data stream for MUE/SUE versus transmit power at BS ($K = 8$, Perfect CSI). 58
3.5 The average MSE per data stream versus the number of MUEs K ($P_{\mathrm{BS}} = 46$ dBm, $\bar{\sigma}_h^2 = 1$). 58
3.6 The BER per data stream for MUE/SUE versus the number of MUEs K ($P_{\mathrm{BS}} = 46$ dBm, $\bar{\sigma}_h^2 = 1$). 59
3.7 The average MSE per data stream versus normalized channel estimation error $\bar{\sigma}_h^2$ ($K = 8$, $P_{\mathrm{BS}} = 46$ dBm, Imperfect CSI). 59
3.8 The BER per data stream for MUE/SUE versus normalized channel estimation error $\bar{\sigma}_h^2$ ($K = 8$, $P_{\mathrm{BS}} = 46$ dBm, Imperfect CSI). 60

4.1 System model for HetNet with SCs deployment. 67

4.2 Spectral efficiency versus p_{BS} for one-tier and two-tier network topologies (RSA, $p_\tau = 0$ dBm, $S = 8$, $\gamma = 8$). 85
4.3 Cell boundary SC user rate versus p_{BS} for one-tier and two-tier network topologies (RSA, $p_\tau = 0$ dBm, $S = 8$, $\gamma = 8$). 86
4.4 Spectral efficiency versus p_{BS} for Monte-Carlo results and lower bounds (RSA, $p_\tau = 0$ dBm, $S = 8$, $\gamma = 8$). 86
4.5 Spectral efficiency versus N_{BS} for different PR factors (RSA, $p_{\mathrm{BS}} = 46$ dBm, $p_\tau = 0$ dBm, $S = 20$). 87
4.6 Transmit power required to achieve 1 bit/s/Hz per user for case I (normalized $p_\tau = 0$ dB, $S = 8$). 89
4.7 Transmit power required to achieve 1 bit/s/Hz per user for case I ($S = 8$, $\gamma = 8$). 89
4.8 Transmit power required to achieve 1 bit/s/Hz per user for case II (normalized $E_\tau = 0$ dB, $S = 8$, $\gamma = 8$). 90
4.9 Spectral efficiency versus p_{BS} for different user scheduling algorithms (MRT, $p_\tau = 5$ dBm, $S = 8$, $\gamma = 8$). 90
4.10 Spectral efficiency versus p_{BS} for different user scheduling algorithms (ZFT, $p_\tau = 5$ dBm, $S = 8$, $\gamma = 8$). 91

Notations

Unless stated otherwise, boldface upper-case and lower-case letters denote matrices and vectors respectively.

$\mathbf{X}^*$	the conjugate of matrix $\mathbf{X}$
$\mathbf{X}^{\mathrm{T}}$	the transpose of matrix $\mathbf{X}$
$\mathbf{X}^{\mathrm{H}}$	the conjugate transpose (Hermitian) of matrix $\mathbf{X}$
$\mathrm{tr}\{\mathbf{X}\}$	the trace of $\mathbf{X}$
$\mathbf{I}_M$	the identity matrix of dimension $M \times M$
$\mathbf{0}$	a zero vector or matrix
$\lvert x \rvert$	the absolute value of a real scalar x or magnitude of a complex scalar x
$\lVert \mathbf{x} \rVert$	the Euclidean norm of a vector $\mathbf{x}$
$\lVert \mathbf{X} \rVert_{\mathrm{F}}$	the Frobenius norm of a matrix $\mathbf{X}$
$\mathrm{E}[\cdot]$	the expectation operator
$\mathrm{Var}[\cdot]$	the variance operator
$\mathrm{vec}(\cdot)$	the vector composed of all columns of a matrix
$\mathbb{R}$	the real number set
$\mathbb{C}$	the complex number set
$\mathbb{C}^m$	a set of complex column vectors with size m
$\mathbb{C}^{m \times n}$	a set of complex matrices with size $m \times n$
$\sim$	distributed according to
$\mathcal{CN}(\mathbf{m}, \mathbf{\Sigma})$	complex Gaussian distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{\Sigma}$
$\xrightarrow{a.s.}$	almost sure convergence
$\triangleq$	defined as
max	maximize
min	minimize
$\max\{x, y\}$	the minimum one of x and y
$[a]^+$	$\max\{a, 0\}$
$\mathrm{diag}\{\mathbf{x}\}$	diagonalization of $\mathbf{x}$
$\mathrm{bd}\{\mathbf{X}_1, \ldots, \mathbf{X}_K\}$	block diagonal matrix with the main diagonal blocks as $\mathbf{X}_1, \ldots, \mathbf{X}_K$

Abbreviations

ACO	Alternating Convex Optimization
AF	Amplify and Forward
ASA	Asymptotically Scheduling Algorithm
AWGN	Additive White Gaussian Noise
BC	Broadcast
BD	Block Diagonalization
BER	Bit Error Rate
BS	Base Station
CoMP	Coordinated Multi-point
CSI	Channel State Information
DF	Decode and Forward
EPA	Equal Power Allocation
GP	Geometric Programing
GSA	Greedy Scheduling Algorithm
HetNet	Heterogeneous Network
IC	Interference Control
JP	Joint Processing
KKT	Karush-Kuhn-Tucker
LP	Linear Programming
MA	Multi-Access
MC	Macro Cell
MMSE	Minimum Mean-Square-Error
MRC	Maximum-Ratio Combining
MRT	Maximum-Ratio Transmission
MUE	Macro-cell User Equipment
OPA	Optimal Power Allocation
PA	Power Allocation
PR	Pilot Reuse
RAO	Relaxed-constraints based Alternating Optimization

RSA	Random Scheduling Algorithm
SC	Small Cell
SCSI	Statistical Channel State Information
SIC	Self-Interference Cancellation
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SUE	Small-cell User Equipment
UAON	Unconstrained Alternating Optimization with Normalization
UE	User Equipment
ZFR	Zero-Forcing Receiption
ZFT	Zero-Forcing Transmission

ACKNOWLEDGEMENTS

First and foremost, I would like to attribute my greatest gratitude to my supervisor, Prof. Xiaodai Dong, for the continuous support of my Ph.D study and research. I am grateful to her for her patient guidance, insightful comments, inspired instructions and thesis revision every time, and providing me with an excellent research atmosphere.

I am also indebted to my departmental committee member, Prof. Wu-Sheng Lu, for his meticulous guidance and insightful suggestions and advice.

I also thank my thesis outside committee member, Prof. Julie Zhou, for offering me valuable comments and suggestions.

My sincere thanks also go to Prof. Hai Lin for the insightful comments on the research work and paper revision.

I also thank my fellow classmates in the whole laboratory: Zheng Xu, Ming Lei, Ping Cheng, Yunlong Shao, Lan Xu, Le Liang, Weiheng Ni, Yiming Huo, Farnoosh Talaei, Jun Zhou, Guang Zeng, Hongrui Wang, Yuejiao Hui, Guowei Zhang, Wanbo Li, for the research discussions and for all the fun we had in the last three years. I would like to thank my best friends: Fang Chen, Mengyue Cai, Xiao Feng, Xiao Xie, Zhu Ye, Simin Yu, Yue Gu, Xiaoyan Shen, Juan Yang, for the support and happy time they gave me in my life.

Last but not least, I express my endless gratitude to my husband, Leyuan Pan, for his unconditional support, help, patience, sacrifices and love. He stands by me through the good times and the bad. And my parents, who are always supporting me and encouraging me through my whole life.

Yongyu Dai, Victoria, BC, Canada

DEDICATION

To my parents,
My husband,
And all my friends,
For everything.

Chapter 1

Overview

1.1 Research Areas Overview

MASSIVE multiple-input multiple-output (MIMO) transmission, in which a base station is equipped with hundreds of antennas for multiuser operation, is an emerging technology that enables enormous capacity enhancements and significant total power reduction. As one of the key enabling technologies for the next generation of wireless systems, not surprisingly, massive MIMO has become one of the most attractive technologies. This dissertation is focused on massive MIMO two-way relaying and heterogeneous network (HetNet). Capacity lower bounds for massive MIMO two-way amplify-and-forward (AF) relaying will be first derived with maximum-ratio combining/maximum-ratio transmission (MRC/MRT) and zero-forcing reception/zero-forcing transmission (ZFR/ZFT) beamforming, respectively. Based on the derived closed-form achievable rate expressions, optimal power allocation (OPA) strategies will be proposed to further improve system performance. Then, the mean square error (MSE) based precoding will be proposed and then be employed by the macro base station (BS) and the small cell (SC) nodes for MIMO downlinks in the HetNet. Finally, the performance analysis for HetNet with large-scale antenna arrays will be provided and effective user scheduling algorithms will be presented.

1.1.1 Multi-Pair Two-Way Relaying with Large-Scale Antennas

Massive MIMO combined with cooperative relaying is addressed as a strong candidate for the development of future energy-efficient networks. In the field of cooperative relaying, the two-way relaying technique outperforms one-way relaying

in terms of spectral efficiency, since it employs the principle of network coding at the relay to extract the desired information. As to the multi-pair two-way relaying with massive MIMO in the literature, the asymptotic analyses of the system were obtained with both MRC/MRT and ZFR/ZFT beamforming supposing that the number of relay antennas approaches to infinity and the transmit powers of all users are the same. Besides, asymptotic approximations were always involved even in the derivation of the closed-form expression for the ergodic achievable rate with finite number of relay antennas. Neither exact closed-form expressions or power allocation schemes have been addressed in a multi-pair massive MIMO two-way relaying in the literature. This motivates us to derive closed-form achievable rate expressions without resorting to asymptotic approximations for a multi-pair massive MIMO two-way AF relaying, and to propose the corresponding optimal PA schemes.

MRC/MRT beamforming is a kind of much simpler signal processing technique, which aims at maximizing the average output signal-to-noise ratio (SNR). In the massive MIMO two-way relaying system, large antenna arrays can substantially reduce the effects of the noise, small-scale fading and inter-user interference, using only straightforward MRC/MRT. While ZFR/ZFT beamforming is able to null multi-user interference signals and further improve the achievable rate in the massive MIMO two-way relaying system with finite number of antennas. And both of them will be focused on in Chapter 2.

1.1.2 Interference Mitigation for Heterogeneous Networks

As a viable and cost-effective way to increase network capacity, HetNets that embed low-power nodes, called small cells, into an existing macro network has emerged aiming to offload traffic from the macro cell (MC) to small cells because of the practical limitations of conventional homogeneous networks. In a typical HetNet consisting of a MC and several SCs, the MC serves its user equipments (UEs) in a large region by a high-power base station, while each SC serves its UEs in its own coverage region by a low-power SC node if there is no cooperative transmission between the BSs and SCs. Due to the large number of potential interfering nodes in the network, mitigating both the inter-cell and intra-cell interference becomes a crucial issue facing HetNet. There are some effective interference control methods and coordinated multi-point transmission schemes that can be applied in HetNet to overcome the interference introduced effect. Nevertheless, these schemes with closed-form expressions are only available in certain cases, such as a two-user MIMO interference channel. Hence, in Chapter 3 we will investigate and design the precoding schemes for the MIMO downlink systems in the HetNet.

1.1.3 Heterogeneous Networks with Large-Scale Antennas

Since massive MIMO becomes more and more popular, its coexistence with small cells has been put forth with large-scale antenna arrays at the BSs and limited antennas at the SCs[1] due to their smaller form factor. As variable structure of antenna arrays, such as a cylindrical array, requires less space, large-scale antenna arrays set at SCs becomes realizable. This motivates us to consider a large-scale HetNet with massive antennas at both BSs and SCs. Similarly, both MRT and ZFT will be focused on in Chapter 4.

1.2 Dissertation Organization

In Chapter 2, capacity lower bounds for multi-pair massive MIMO two-way AF relaying with linear processing are derived, based on which optimal power allocation strategies are proposed to further improve the system performance. In Chapter 3, a HetNet system consisting of a macro tier overlaid with a second tier of SCs are studied, and the MSE based precoding schemes are designed and employed by the macro base station and the SC nodes for MIMO downlinks. Chapter 4 provides performance analysis for pilot-reused HetNet with large-scale antenna arrays set at both BS and SC nodes and proposes corresponding user scheduling schemes. Finally, the research contributions on each topic are concluded in Chapter 5.

[1] In this dissertation, we use SC to denote the SC node for simplicity.

Chapter 2

Power Allocation for Multi-Pair Massive MIMO Two-Way AF Relaying with Linear Processing

Employing relay techniques is a common method in communications to extend network coverage. Currently, two-way relaying has gained increasing attention as it is able to circumvent the inherent spectral loss of unidirectional protocols caused by half-duplex constraint [1–3]. It is also well known that by extending single antenna relay to multi-antenna relay can introduce significant performance improvement for the system. Since massive multiple-input multiple-output (MIMO) transmission was proposed and then became more and more popular, we will focus on the two-way amplify-and-forward (AF) relaying with large-scale antennas utilizing linear processing in this chapter.

2.1 Introduction

MASSIVE multiple-input multiple-output (MIMO) transmission, in which a base station is equipped with hundreds of antennas for multiuser operation, is considered as one of the key enabling technologies for 5G [4]. In [5], it was first proposed for multi-cell noncooperative scenarios. Such large antenna arrays can substantially reduce the effects of noise, small-scale fading and inter-user interference, using only simple signal processing techniques with reduced total transmit power, and only inter-cell interference caused by pilot contamination remains [5, 6]. Subsequently, the energy and spectral efficiency of very large multiuser MIMO systems were investigated in the single cell scenarios in [7], which showed that the power radiated by the terminals could be made inversely proportional to the square-root of the

number of base station antennas with no reduction in performance when considering imperfect channel state information (CSI), and that the power could be made inversely proportional to the number of antennas if perfect CSI were available.

2.2 Related Work

Currently, massive MIMO combined with cooperative relaying is considered as a strong candidate for the development of future energy-efficient networks and has received increasing attention [8–13]. In the field of cooperative relaying, two-way relaying technique outperforms one-way relaying in terms of spectral efficiency, since it employs the principle of network coding at the relay in order to mix the signals received simultaneously from two links for subsequent forwarding, and then applies the self-interference cancellation (SIC) at each user to extract the desired information [1, 2]. For the multi-pair two-way relaying with massive MIMO, [10] obtained the asymptotic spectral and energy efficiencies of the system analytically with both maximum-ratio combining/maximum-ratio transmission (MRC/MRT) and zero-forcing reception/zero-forcing transmission (ZFR/ZFT) beamforming, supposing that the number of relay antennas approaches to infinity and the transmit power of all users is equal. However, only asymptotic cases with perfect CSI and perfect SIC were studied and no closed-form expression for the ergodic achievable rate with finite number of relay antennas was derived in [10]. In [12], the ergodic achievable rates were investigated with perfect CSI based MRC/MRT used at the relay, providing a capacity lower bound, the derivation of which involved asymptotic approximations. Neither [10] nor [12] considers imperfect CSI or power allocation (PA) problems.

In the literature, instantaneous power allocation schemes based on instantaneous rate for regular scale MIMO rather than massive MIMO were presented for one way or two way AF wireless relay systems to improve system performance [14–16]. In massive MIMO systems, ergodic rate is usually used in power allocation because the instantaneous rate approaches the ergodic rate as the number of antennas tends to infinity due to the law of large numbers, and such PA schemes are more practical with lower complexity than instantaneous rate based ones. In [13], an ergodic rate based optimal power allocation (OPA) scheme was proposed for a multi-pair decode-and-forward (DF) one-way relaying with massive arrays. Nevertheless, power allocation has not been addressed in a massive MIMO two-way relaying system. Besides, there is no closed-form ergodic rate expressions derived for massive MIMO two-way relaying with ZFR/ZFT in the literature.

2.3 Contributions

This chapter considers a multi-pair two-way amplify-and-forward (AF) relaying system where multiple sources exchange information via a relay node equipped with large-scale arrays [17]. Assuming imperfect CSI estimation, the relay station employs the MRC/MRT and ZFR/ZFT beamforming to process the signals, respectively. First, utilizing the technique in [18, 19], we derive for the first time two statistical CSI (SCSI) based closed-form lower bounds for the ergodic achievable rate in the case of arbitrary number of relay antennas (without resorting to asymptotic approximations) with MRC/MRT and ZFR/ZFT processing, respectively, based on the properties of Wishart and inverse Wishart matrices. Having obtained the closed-form expressions, we are able to design an OPA scheme that maximizes the sum spectral efficiency under certain practical constraints. The proposed OPA scheme is based on geometric programming (GP) [14, 20], which can be solved by conventional optimization tools, such as CVX [21]. Considering the massive MIMO properties, an asymptotically OPA is presented for the asymptotic regimes with closed-form solutions. The derived closed-form expressions for the achievable rate are verified to be accurate predictors of the system performance by Monte-Carlo simulations. Furthermore, in order to demonstrate the effectiveness of the developed OPA schemes, simulations of spectral efficiency are conducted under different system configurations, respectively, in comparison to the equal power allocation (EPA) schemes.

The rest of the chapter is organized as follows. We briefly describe the system model for the multi-pair two-way AF relaying in Section 4.4. In Section 2.5, two closed-form expressions for the achievable rate are derived for MRC/MRT and ZFR/ZFT, respectively, followed by asymptotic analysis. Then, an OPA and an asymptotically OPA are proposed by solving the sum-rate maximization based optimization problem in Section 2.6. Furthermore, simulation results under different system configurations are given in Section 2.7 to demonstrate the effectiveness of both derived rate expressions and developed OPAs. Finally, we draw our conclusions in Section 2.8.

2.4 System Model

Fig. 2.1 shows the considered multi-pair two-way AF relaying network, where K pairs of users communicate with the help of a common relay station by sharing the same time-frequency resources. In this system, two single-antenna users in the lth user pair denoted by $(2l-1, 2l)$ or $(2l, 2l-1)$ (for $l = 1, \cdots, K$) want

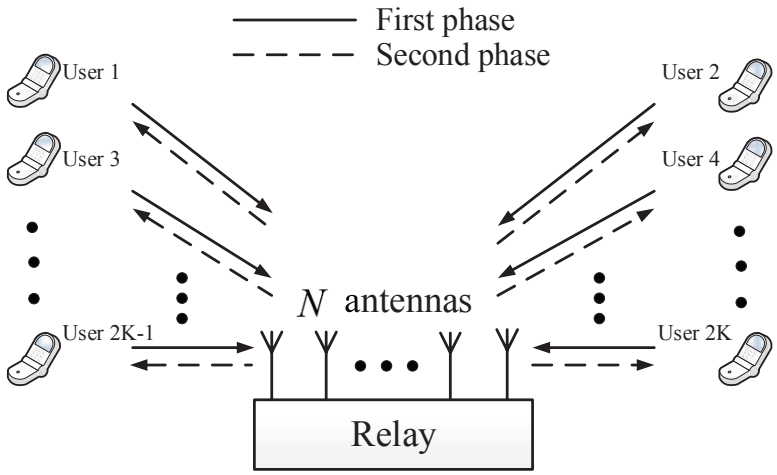

Figure 2.1: System diagram of multi-pair two-way AF relaying.

to exchange information with each other via the relay equipped with N ($N \gg 2K \gg 1$) antennas. Notably, the direct links between the corresponding users are assumed non-existing in the two-way relaying system. Typically, a two-way network is divided into two phases, namely the multiple-access (MA) phase and the broadcast (BC) phase [1]. In the MA phase, information is sent from the user pairs to the relay; while in the BC phase, the relay broadcasts the processed information.

Let p_i ($i = 1, 2, \cdots, 2K$) and P_{R} denote the power transmitted by user i and the relay corresponding to the MA and BC phases, respectively. We assume that all the channels between the users and the relay follow independent and identically distributed (i.i.d.) Rayleigh fading and time division duplex (TDD) is adopted in all transceivers. Thus, supposing that $\mathbf{g}_i \in \mathbb{C}^{N\times 1}$ ($i = 1, 2, \cdots, 2K$) is the channel between the ith user and the relay, $\mathbf{g}_i$ contains the i.i.d. $\mathcal{CN}(0, \sigma_i^2)$ elements, where σ_i^2 represents the corresponding large-scale fading coefficient. In this way, we can denote the channel matrix between all the users and the relay accounting for both small-scale fading and large-scale fading by

$$\mathbf{G} = \mathbf{H}\mathbf{D}^{1/2} = [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_{2K}] \in \mathbb{C}^{N\times 2K} \tag{2.1}$$

where $\mathbf{H} \in \mathbb{C}^{N\times 2K}$ includes the i.i.d. $\mathcal{CN}(0, 1)$ small-scale fading coefficients, and $\mathbf{D}$ is the large-scale fading diagonal matrix with the ith diagonal elements denoted by σ_i^2 ($i = 1, 2, \cdots, 2K$).

2.4.1 Channel Estimation

Practically, the channel matrices in both the MA and BC phases have to be estimated for relay processing. However, due to the large-scale antenna array at the relay, channel estimation at the user side becomes rather impractical. Thus, time division duplex (TDD) is adopted here and channel reciprocity can be utilized, i.e., only channel matrix $\mathbf{G}$ between all the users and the relay has to be estimated based on the uplink training. The relay then has the estimated CSIs of all uplink and downlink channels. The required channel related information at the user side can be calculated by the relay and fed back to the users, as will be explained later. At the beginning of each coherence interval T, all users simultaneously transmit pilot sequences of length τ symbols. The pilot sequences of all the $2K$ users are pairwisely orthogonal, i.e., $\tau \geq 2K$ is required. Then the training matrix received at the relay is

$$\mathbf{Y}_{\mathrm{R}} = \sqrt{\tau p_{\mathrm{P}}}\mathbf{G}\mathbf{\Phi} + \mathbf{N}_{\mathrm{R}} \tag{2.2}$$

where p_{P} is the transmit power of each pilot symbol, $\mathbf{N}_{\mathrm{R}}$ is the additive white Gaussian noise (AWGN) matrix with i.i.d. components following $\mathcal{CN}(0, \sigma_n^2)$, the training vector transmitted by the ith ($i = 1, \cdots, 2K$) user is denoted by the ith row of $\mathbf{\Phi} \in \mathbb{C}^{2K\times\tau}$, satisfying $\mathbf{\Phi}\mathbf{\Phi}^H = \mathbf{I}_{2K}$. Moreover, since the rows of pilot sequence matrices are pairwisely orthogonal, we have $\phi^{(i)}\left(\phi^{(j)}\right)^H = 0$ ($\forall i \neq j \in \{1, \ldots, 2K\}$).

In order to estimate the channel matrices $\mathbf{G}$, we employ the minimum mean-square-error (MMSE) estimation at the relay. The MMSE channel estimates are given by [22]

$$\hat{\mathbf{G}} = \frac{1}{\sqrt{\tau p_{\mathrm{P}}}}\mathbf{Y}_{\mathrm{R}}\mathbf{\Phi}^H\tilde{\mathbf{D}} = \mathbf{G}\tilde{\mathbf{D}} + \frac{1}{\sqrt{\tau p_{\mathrm{P}}}}\tilde{\mathbf{N}}_{\mathrm{R}}\tilde{\mathbf{D}} \tag{2.3}$$

where we define $\tilde{\mathbf{D}} \triangleq \left(\frac{\mathbf{D}^{-1}\sigma_n^2}{\tau p_{\mathrm{P}}} + \mathbf{I}_{2K}\right)^{-1}$ and $\tilde{\mathbf{N}}_{\mathrm{R}} \triangleq \mathbf{N}_{\mathrm{R}}\mathbf{\Phi}^H$. According to the property of $\mathbf{\Phi}$, we conclude that $\tilde{\mathbf{N}}_{\mathrm{R}}$ is composed of i.i.d. $\mathcal{CN}(0, \sigma_n^2)$ elements. Then,

$$\hat{\mathbf{G}} = \mathbf{G} + \mathbf{\Xi} = [\hat{\mathbf{g}}_1, \hat{\mathbf{g}}_2, \ldots, \hat{\mathbf{g}}_{2K}] \in \mathbb{C}^{N\times 2K} \tag{2.4}$$

where $\mathbf{\Xi} = [\xi_1, \xi_2, \ldots, \xi_{2K}]$ denotes the estimation error matrix which is independent of $\hat{\mathbf{G}}$ from the property of MMSE channel estimation [22]. Hence, we have $\hat{\mathbf{G}} \sim CN\left(0, \hat{\mathbf{D}}\right)$ with $\hat{\mathbf{D}} = diag\left\{\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_{2K}^2\right\}$, and $\mathbf{\Xi} \sim CN\left(0, \mathbf{D} - \hat{\mathbf{D}}\right)$ with $\mathbf{D} - \hat{\mathbf{D}} = diag\left\{\sigma_{\xi_1}^2, \sigma_{\xi_2}^2, \ldots, \sigma_{\xi_{2K}}^2\right\}$. The diagonal elements satisfy $\hat{\sigma}_i^2 = \frac{\tau p_{\mathrm{P}}\sigma_i^4}{\tau p_{\mathrm{P}}\sigma_i^2 + \sigma_n^2}$ and $\sigma_{\xi_i}^2 \triangleq \sigma_i^2 - \hat{\sigma}_i^2 = \frac{\sigma_i^2}{\tau p_{\mathrm{P}}\sigma_i^2 + \sigma_n^2}$ with $i = 1, \ldots, 2K$.

2.4.2 Data Transmission

Since the relay station estimates all the channels, it employs linear processing MRC/MRT and ZFR/ZFT based on the imperfect CSI. While each user only has the knowledge of its pairwise effective channel coefficient for data detection and self-interference cancellation coefficient for SIC, which are calculated and sent out by the relay. In the MA phase, all the users transmit their signals simultaneously to the relay. That is, the received signal at the relay station is given by

$$\mathbf{r} = \sum_{i=1}^{2K} \sqrt{p_i}\mathbf{g}_i x_i + \mathbf{n}_{\mathrm{r}} = \tilde{\mathbf{G}}\mathbf{x} + \mathbf{n}_{\mathrm{r}} \tag{2.5}$$

where $\tilde{\mathbf{G}} = \mathbf{GP}$, $\mathbf{P} = diag\left\{\sqrt{p_1}, \sqrt{p_2} \ldots \sqrt{p_{2K}}\right\}$ with each power satisfying $0 \leq p_i \leq P_0$[1], $\mathbf{x} = [x_1, x_2, \ldots, x_{2K}]^T$ with the ith element x_i representing the transmitted signal by the ith user and $\mathrm{E}\left[\mathbf{x}\mathbf{x}^H\right] = \mathbf{I}_{2K}$, $\mathbf{r} \in \mathbb{C}^{N\times 1}$, and $\mathbf{n}_{\mathrm{r}}$ is the additive white Gaussian noise (AWGN) vector at the relay with zero mean and the variance of σ_n^2.

Then, in the BC phase, the relay multiplies the received signal by a linear receiving and precoding matrix to yield the relay transmitted signal given by $\hat{\mathbf{r}} = \mathbf{Fr} \in \mathbb{C}^{N\times 1}$, where $\mathbf{F}$ is the combined beamforming matrix at the relay and its expression will be given in the next subsection. The transmitted signal satisfies the expected transmit power constraint at the relay [23], i.e.,

$$P_{\mathrm{R}} = \mathrm{E}\left[\|\hat{\mathbf{r}}\|^2\right] = \mathrm{tr}\left\{\mathrm{E}\left[\mathbf{F}\left(\tilde{\mathbf{G}}\tilde{\mathbf{G}}^H + \sigma_n^2\mathbf{I}_N\right)\mathbf{F}^H\right]\right\} \tag{2.6}$$

with a total power constraint $\sum_{i=1}^{2K} p_i + P_{\mathrm{R}} \leq P$[1], where the expectation is performed with respect to all the involved random variables. In the BC phase, the received signal at the k'th user can be expressed as

$$\begin{aligned} y_{k'} &= \mathbf{g}_{k'}^T\hat{\mathbf{r}} + n_{k'} \\ &= \underbrace{\sqrt{p_k}\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k x_k}_{\text{desired signal}} + \underbrace{\sqrt{p_{k'}}\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_{k'}x_{k'}}_{\text{self-interference}} + \underbrace{\mathbf{g}_{k'}^T\mathbf{F}\sum_{i\neq k,k'}^{2K}\sqrt{p_i}\mathbf{g}_i x_i}_{\text{inter-pair interference}} + \underbrace{\mathbf{g}_{k'}^T\mathbf{F}\mathbf{n}_{\mathrm{r}}}_{\text{noise from relay}} + \underbrace{n_{k'}}_{\text{noise at user}} \end{aligned} \tag{2.7}$$

where (k, k') is defined to indicate the $\lceil k/2 \rceil$th[2] user pair, and $n_{k'}$ represents the AWGN noise at the k'th user side with zero mean and variance of σ_n^2.

[1]Here, P and P_0 are two constants preset for the total power constraint and individual power constraint, respectively.

[2]It is clear that $\lceil k/2 \rceil = \lceil k'/2 \rceil$.

Using the estimated CSI, the relay calculates and sends out the SIC coefficient $\hat{\mathbf{g}}_k^T\mathbf{F}\hat{\mathbf{g}}_k$ ($k \in \{1, \cdots, 2K\}$) for each user. Hence, the received signal at the k'th user after SIC is rewritten as

$$\tilde{y}_{k'} = \underbrace{\sqrt{p_k}\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k x_k}_{\text{desired signal}} + \underbrace{\sqrt{p_{k'}}\lambda_{k'}x_{k'}}_{\text{residual self-interference}} + \underbrace{\mathbf{g}_{k'}^T\mathbf{F}\sum_{i\neq k,k'}^{2K}\sqrt{p_i}\mathbf{g}_i x_i}_{\text{inter-pair interference}} + \underbrace{\mathbf{g}_{k'}^T\mathbf{F}\mathbf{n}_{\mathrm{r}}}_{\text{noise from relay}} + \underbrace{n_{k'}}_{\text{noise at user}}. \tag{2.8}$$

where the residual self-interference involves $\lambda_{k'} = \mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_{k'} - \hat{\mathbf{g}}_{k'}^T\mathbf{F}\hat{\mathbf{g}}_{k'}$, since the SIC coefficient $\hat{\mathbf{g}}_{k'}^T\mathbf{F}\hat{\mathbf{g}}_{k'}$ for user k' is obtained from the estimated CSI. Here, we suppose that there is no error during the SIC coefficients transmission from the relay.

2.4.3 MRC/MRT Processing

In this subsection, the simple and widely used MRC/MRT beamforming is adopted. According to [24], the imperfect CSI based MRC/MRT beamforming is given by

$$\mathbf{F} = \alpha_1\hat{\mathbf{G}}^*\mathbf{T}\hat{\mathbf{G}}^H \tag{2.9}$$

where $\mathbf{T} = diag\{\mathbf{T}_1, \mathbf{T}_2, \ldots \mathbf{T}_K\}$ is the block diagonal permutation matrix indicating the user pairing format with $\mathbf{T}_1 = \mathbf{T}_2 = \cdots = \mathbf{T}_K = [0\ 1;\ 1\ 0]$, and α_1 is a normalization constant, chosen to satisfy the power constraint at the relay station in (2.6).

By substituting (2.9) into (2.6), we have

$$\begin{aligned}\alpha_1 &\overset{(a)}{=} \sqrt{\frac{P_{\mathrm{R}}}{\mathrm{E}\left[\left\|\hat{\mathbf{G}}^*\mathbf{T}\hat{\mathbf{G}}^H\tilde{\mathbf{G}}\mathbf{x}\right\|^2\right] + \mathrm{E}\left[\left\|\hat{\mathbf{G}}^*\mathbf{T}\hat{\mathbf{G}}^H\mathbf{n}_{\mathrm{r}}\right\|^2\right]}} \\ &\overset{(b)}{=} \sqrt{\frac{P_{\mathrm{R}}}{N\left(N+1\right)\left[2\left(\Psi+\sigma_n^2\right)\hat{\Phi} + \left(N+1\right)\sum_{i=1}^{K}\hat{\psi}_i\hat{\phi}_i\right]}}\end{aligned} \tag{2.10}$$

where

$$\begin{aligned}\hat{\Phi} &= \sum_{i=1}^{K}\hat{\phi}_i, \hat{\phi}_i = \hat{\sigma}_{2i-1}^2\hat{\sigma}_{2i}^2, \hat{\psi}_i = p_{2i-1}\hat{\sigma}_{2i-1}^2 + p_{2i}\hat{\sigma}_{2i}^2 \\ \Psi &= \sum_{i=1}^{K}\left(p_{2i-1}\sigma_{2i-1}^2 + p_{2i}\sigma_{2i}^2\right).\end{aligned} \tag{2.11}$$

The detailed derivation of the equation is given in Appendix 2.9.1.

2.4.4 ZFR/ZFT Processing

When employing ZFR/ZFT with imperfect CSI, in which the pseudo-inverse of the estimated channels in (2.4) are needed for processing, the linear beamforming is given by [24]

$$\mathbf{F} = \alpha_2 \hat{\mathbf{G}}^* \left(\hat{\mathbf{G}}^T \hat{\mathbf{G}}^* \right)^{-1} \mathbf{T} \left(\hat{\mathbf{G}}^H \hat{\mathbf{G}} \right)^{-1} \hat{\mathbf{G}}^H = \alpha_2 \hat{\bar{\mathbf{G}}}^* \mathbf{T} \hat{\bar{\mathbf{G}}}^H \tag{2.12}$$

where $\hat{\bar{\mathbf{G}}} = \hat{\mathbf{G}} \left(\hat{\mathbf{G}}^H \hat{\mathbf{G}} \right)^{-1}$ and α_2 is the normalization constant, chosen to satisfy the transmit power constraints at the relay. Notably, SIC is not necessary as ZFR/ZFT leads to $\hat{\mathbf{g}}_{k'}^T \mathbf{F} \hat{\mathbf{g}}_{k'} = 0$ $(\forall k' \in \{1, \cdots, 2K\})$. On the basis of (2.12) and $\mathrm{tr}\{\mathbf{AB}\} = \mathrm{tr}\{\mathbf{BA}\}$, we have

$$\begin{aligned} \alpha_2 &\overset{(a)}{=} \sqrt{\frac{P_{\mathrm{R}}}{\mathrm{E}\left[\left\| \hat{\bar{\mathbf{G}}}^* \mathbf{T} \hat{\bar{\mathbf{G}}}^H \tilde{\mathbf{G}} \mathbf{x} \right\|^2 \right] + \mathrm{E}\left[\left\| \hat{\bar{\mathbf{G}}}^* \mathbf{T} \hat{\bar{\mathbf{G}}}^H \mathbf{n}_{\mathrm{r}} \right\|^2 \right]}} \\ &\overset{(b)}{=} \sqrt{\frac{P_{\mathrm{R}}}{\sum_{i=1}^{2K} \frac{p_{i'}}{(N-2K-1)\hat{\sigma}_i^2} + \hat{\eta} \left(\sum_{j=1}^{2K} p_j \sigma_{\xi_j}^2 + \sigma_n^2 \right)}} \end{aligned} \tag{2.13}$$

where $\hat{\eta} = \sum_{j=1}^{2K} \frac{1}{(N-2K)(N-2K-3)\hat{\sigma}_j^2 \hat{\sigma}_{j'}^2}$. The detailed derivation of (2.13) is given in Appendix 2.9.2.

2.5 Achievable Rate Analysis

In this section, a general form of the ergodic achievable rate of the transmission link $k \to k'$ for MRC/MRT processing is given first, followed by a rate expression for ZFR/ZFT. In order to obtain a basic and insightful expression that can be used for power allocation optimization, a simplified capacity lower bound is derived utilizing the technique of [18, 19], in which the received signal is rewritten as a known mean times the desired symbol, plus an uncorrelated effective noise. The worst-case uncorrelated effective noise, where each additive term is treated as independent Gaussian noise of the same variance, is employed to derive a lower bound.

From (2.8), the ergodic achievable rate of transmission link $k \to k'$ is expressed

as

$$\gamma_{k'}^{\mathrm{ICSI}} = \mathrm{E}\left[\log_2\left(1+\frac{p_k\left|\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right|^2}{p_{k'}|\lambda_{k'}|^2+\sum\limits_{i\neq k,k'}^{2K}p_i\left|\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_i\right|^2+\left\|\mathbf{g}_{k'}^T\mathbf{F}\right\|^2\sigma_n^2+\sigma_n^2}\right)\right]. \quad (2.14)$$

Remark 2.1: Here, the ergodic achievable rate is valid based on the assumption that the receiving user k' knows perfectly $\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k$ in the detection process. To demonstrate the accuracy of the derived lower bounds, we compare the lower bounds with Monte-Carlo realized (2.14) in Section 2.7. The normalization constant for $\mathbf{F}$ in (2.14) is assumed to be calculated based on instantaneous CSI by satisfying $P_{\mathrm{R}} = \|\hat{\mathbf{r}}\|^2$.

Further derivation of (2.14) is difficult because of the intractability to carry out the ensemble average analytically. Instead, we adopt the technique in [18] to derive a worst-case lower bound of the achievable rate. The first step is to rewrite $\sqrt{p_k}\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k x_k$ in (2.8) as the sum of $\sqrt{p_k}\mathrm{E}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right]x_k$ and $\sqrt{p_k}\left(\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k-\mathrm{E}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right]\right)x_k$, where the first part is now considered as the "desired signal". That is, (2.8) can be expressed as

$$\tilde{y}_{k'} = \underbrace{\sqrt{p_k}\mathrm{E}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right]x_k}_{\text{desired signal}} + \underbrace{\tilde{n}_{k'}}_{\text{effective noise}} \quad (2.15)$$

where $\tilde{n}_{k'}$ is considered as the effective noise and given by

$$\begin{aligned}\tilde{n}_{k'} \triangleq& \sqrt{p_k}\left(\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k-\mathrm{E}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right]\right)x_k+\sqrt{p_{k'}}\lambda_{k'}x_{k'}\\&+\mathbf{g}_{k'}^T\mathbf{F}\sum_{i\neq k,k'}^{2K}\sqrt{p_i}\mathbf{g}_i x_i+\mathbf{g}_{k'}^T\mathbf{F}\mathbf{n}_{\mathrm{r}}+n_{k'}.\end{aligned} \quad (2.16)$$

It is straightforward to show that the first term "desired signal" and the second term "effective noise" in (2.15) are uncorrelated. The exact pdf of $\tilde{n}_{k'}$ is not easy to obtain, but we know that the worst-case is to approximate the effective noise as independently Gaussian distributed [18, 25]. Since the relay is equipped with large-scale antenna arrays by assuming $N \gg 2K \gg 1$, the central limit theorem provides a tight statistical CSI based lower bound for the achievable rate. Then, the statistical CSI based achievable rate lower bound of the transmission link $k \to k'$ can be obtained as

$$\gamma_{k'}^{\mathrm{SCSI}} = \log_2\left(1+\frac{p_k\left|\mathrm{E}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right]\right|^2}{p_k\mathrm{Var}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right]+\mathrm{SI}_{k'}+\mathrm{IP}_{k'}+\mathrm{NR}_{k'}+\mathrm{NU}_{k'}}\right) \quad (2.17)$$

where $\mathrm{SI}_{k'}$, $\mathrm{IP}_{k'}$, $\mathrm{NR}_{k'}$ and $\mathrm{NU}_{k'}$ denote the residual self-interference after SIC, the inter-pair interference, the amplified noise from relay and the noise at user, respectively, i.e.,

$$\mathrm{SI}_{k'} \triangleq p_{k'}\mathrm{E}\left[\left|\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_{k'} - \hat{\mathbf{g}}_{k'}^T\mathbf{F}\hat{\mathbf{g}}_{k'}\right|^2\right] \tag{2.18a}$$

$$\mathrm{IP}_{k'} \triangleq \sum_{i\neq k,k'}^{2K} p_i\mathrm{E}\left[\left|\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_i\right|^2\right] \tag{2.18b}$$

$$\mathrm{NR}_{k'} \triangleq \mathrm{E}\left[\left|\mathbf{g}_{k'}^T\mathbf{F}\mathbf{n}_{\mathrm{r}}\right|^2\right],\ \mathrm{NU}_{k'} \triangleq \mathrm{E}\left[|n_{k'}|^2\right]. \tag{2.18c}$$

When MRC/MRT beamforming is employed, further mathematical derivation of (2.17) leads to the following theorem:

Theorem 2.1: With imperfect CSI based MRC/MRT, the ergodic achievable rate of the transmission link $k \to k'$, for a finite number of antennas at the relay, is lower bounded by

$$\text{MRC/MRT}: \ \gamma_{k'}^{\mathrm{SCSI}} = \log_2\left(1 + \frac{a_{k'}p_k}{\sum_{i=1}^{2K}\left(b_{k',i}^{(1)} + b_{k',i}^{(2)}P_{\mathrm{R}}^{-1}\right)p_i + c_{k'}p_{k'} + \left(d_{k'}^{(1)} + d_{k'}^{(2)}P_{\mathrm{R}}^{-1}\right)}\right) \tag{2.19}$$

where $a_{k'} = N(N+1)\hat{\sigma}_k^4\hat{\sigma}_{k'}^4$, $b_{k',i}^{(1)} = (N+1)(\sigma_i^2\hat{\sigma}_{k'}^4\hat{\sigma}_k^2 + \sigma_{k'}^2\hat{\sigma}_i^4\hat{\sigma}_{i'}^2) + 2\sigma_i^2\sigma_{k'}^2\hat{\Phi}$, $b_{k',i}^{(2)} = \sigma_n^2\left[2\hat{\Phi}\sigma_i^2\sigma_{i'}^2 + (N+1)\hat{\sigma}_i^4\hat{\sigma}_{i'}^2\right]$, $c_{k'} = 2\left[(N+1)(\sigma_{k'}^2 - 2\hat{\sigma}_{k'}^2)\hat{\sigma}_k^2\hat{\sigma}_{k'}^4 + (\sigma_{k'}^4 - 2\hat{\sigma}_{k'}^4)\hat{\Phi}\right]$, $d_{k'}^{(1)} = \sigma_n^2\left[(N+1)\hat{\sigma}_k^2\hat{\sigma}_{k'}^4 + 2\sigma_{k'}^2\hat{\Phi}\right]$, and $d_{k'}^{(2)} = 2\sigma_n^4\hat{\Phi}$.

Proof: See Appendix 2.9.3.

For imperfect CSI based ZFR/ZFT processing, a closed-form expression for the achievable rate in (2.17) is derived as follows:

Theorem 2.2: With imperfect CSI based ZFR/ZFT beamforming, the achievable rate of the transmission link $k \to k'$, for a finite number of antennas at the relay, is lower bounded by

$$\text{ZFR/ZFT}: \ \gamma_{k'}^{\mathrm{SCSI}} = \log_2\left(1 + \frac{e_{k'}p_k}{\sum_{i=1}^{2K}\left(f_{k',i}^{(1)} + f_{k',i}^{(2)}P_{\mathrm{R}}^{-1}\right)p_i + m_{k'}p_{k'} + \left(n_{k'}^{(1)} + n_{k'}^{(2)}P_{\mathrm{R}}^{-1}\right)}\right) \tag{2.20}$$

where $e_{k'} = e_k = 1$, $f_{k',i}^{(1)} = \frac{\sigma_{\xi_i}^2}{(N-2K-1)\hat{\sigma}_k^2} + \frac{\sigma_{\xi_{k'}}^2}{(N-2K-1)\hat{\sigma}_{i'}^2} + \sigma_{\xi_{k'}}^2\sigma_{\xi_i}^2\hat{\eta}$, $f_{k',i}^{(2)} = \sigma_n^2\left(\frac{1}{(N-2K-1)\hat{\sigma}_{i'}^2} + \sigma_{\xi_i}^2\hat{\eta}\right)$, $m_{k'} = \sigma_{\xi_{k'}}^4\hat{\eta}$, $n_{k'}^{(1)} = \frac{\sigma_n^2}{(N-2K-1)\hat{\sigma}_k^2} + \sigma_n^2\sigma_{\xi_{k'}}^2\hat{\eta}$, and $n_{k'}^{(2)} = \sigma_n^4\hat{\eta}$.

Proof: See Appendix 2.9.4.

Theorems 2.1 and 2.2 are also valid for conventional MIMO systems, while the

bounds become less tight as the antenna scale goes down. The capacity lower bounds for perfect CSI can always be obtained by setting $\sigma_{\xi_k}^2 = 0$ and $\hat{\sigma}_k^2 = \sigma_k^2$ ($k \in \{1, \cdots, 2K\}$) in (2.19) and (2.20). Moreover, it can be observed from (2.19) that when the estimation error is severe, the residual SI occupies the major part of the imperfect CSI effect in comparison to other terms. On the other hand, if channel estimation is rather accurate, the residual SI has slight effects in comparison to other terms. While for ZFR/ZFT, both the residual SI and inter-pair interference are determined by the channel estimation accuracy.

2.5.1 Asymptotic Analysis with Massive Arrays

Based on the derived closed-form expressions for the achievable rate in (2.19) and (2.20), this section provides the asymptotic analysis under two different cases when the number of relay antennas approaches to infinity. Suppose that all users have the same transmit power, i.e., $p_1 = p_2 \cdots = p_{2K} = P_{\mathrm{S}}$.

Proposition 2.1: In case I where p_{P} is fixed, $p_i = P_{\mathrm{S}} = \frac{E_{\mathrm{S}}}{N^{\rho}}$ ($i = 1, 2, \cdots, 2K$), $P_{\mathrm{R}} = \frac{E_{\mathrm{R}}}{N^{\theta}}$, and E_{S} and E_{R} are fixed, to achieve non-vanishing user rate as $N \to \infty$, the user and relay transmit power scaling factor ρ and θ must satisfy $0 \leq \rho \leq 1$ and $0 \leq \theta \leq 1$. When $\rho = 1$ and $\theta = 1$, the asymptotic achievable rate expressions of the transmission link $k \to k'$ for imperfect CSI based MRC/MRT and ZFR/ZFT are

$$\text{MRC/MRT}: \ \gamma_{k'}^{\text{SCSI}} \xrightarrow[N\to\infty]{a.s.} \log_2 \left(1 + \frac{E_{\mathrm{S}} E_{\mathrm{R}} \hat{\sigma}_k^4 \hat{\sigma}_{k'}^4}{E_{\mathrm{S}} \sigma_n^2 \sum\limits_{i=1}^{2K} \hat{\sigma}_i^4 \hat{\sigma}_{i'}^2 + E_{\mathrm{R}} \sigma_n^2 \hat{\sigma}_k^2 \hat{\sigma}_{k'}^4 + 2\hat{\Phi} \sigma_n^4} \right) \tag{2.21}$$

$$\text{ZFR/ZFT}: \ \gamma_{k'}^{\text{SCSI}} \xrightarrow[N\to\infty]{a.s.} \log_2 \left(1 + \frac{E_{\mathrm{R}} E_{\mathrm{S}} \hat{\sigma}_k^2}{E_{\mathrm{R}} \sigma_n^2 + \hat{\sigma}_k^2 \sigma_n^2 \sum\limits_{i=1}^{2K} \left(\frac{E_{\mathrm{S}}}{\hat{\sigma}_i^2} + \frac{\sigma_n^2}{\hat{\sigma}_i^2 \hat{\sigma}_{i'}^2} \right)} \right) \tag{2.22}$$

respectively, which show that the transmit powers at both users and relay sides can be scaled down by up to $\frac{1}{N}$ to maintain a given rate in case I. When $\rho < 1$ and $\theta < 1$, the asymptotic achievable rate of each user approaches to infinity as $N \to \infty$.

Proposition 2.2: In case II where $p_{\mathrm{P}} = \frac{E_{\mathrm{P}}}{N^{\varsigma}}$, $p_i = P_{\mathrm{S}} = \frac{E_{\mathrm{S}}}{N^{\rho}}$ ($i = 1, 2, \cdots, 2K$), $P_{\mathrm{R}} = \frac{E_{\mathrm{R}}}{N^{\theta}}$, and E_{S} and E_{R} are fixed, to achieve non-vanishing user rate as $N \to \infty$, the pilot, user and relay transmit power scaling factors ς, ρ and θ must satisfy $0 < \varsigma \leq 1$, $0 \leq \rho \leq 1 - \varsigma$ and $0 \leq \theta \leq 1 - \varsigma$. When $0 \leq \varsigma < 1$, $\rho = 1 - \varsigma$ and $\theta = 1 - \varsigma$, the asymptotic achievable rate of the transmission link $k \to k'$ for

imperfect CSI based MRC/MRT and ZFR/ZFT are

$$\text{MRC/MRT}: \ \gamma_{k'}^{\text{SCSI}} \xrightarrow[N\to\infty]{a.s.} \log_2\left(1+\frac{\tau^2 E_{\text{P}}^2 E_{\text{S}} E_{\text{R}} \sigma_k^8 \sigma_{k'}^8}{\tau E_{\text{P}} \sigma_n^4 \left[E_{\text{S}} \sum_{i=1}^{2K} \sigma_i^8 \sigma_{i'}^4 + E_{\text{R}} \sigma_k^4 \sigma_{k'}^8\right] + 2\sigma_n^8 \sum_{j=1}^{K} \sigma_j^4 \sigma_{j'}^4}\right) \tag{2.23}$$

$$\text{ZFR/ZFT}: \ \gamma_{k'}^{\text{SCSI}} \xrightarrow[N\to\infty]{a.s.} \log_2\left(1+\frac{\tau^2 E_{\text{P}}^2 E_{\text{S}} E_{\text{R}} \sigma_k^4}{\tau E_{\text{P}} E_{\text{R}} \sigma_n^4 + \sigma_k^4 \sigma_n^4 \sum_{i=1}^{2K} \left(\frac{\tau E_{\text{P}} E_{\text{S}}}{\sigma_i^4} + \frac{\sigma_n^4}{\sigma_i^4 \sigma_{i'}^4}\right)}\right) \tag{2.24}$$

respectively, from which we conclude that the transmit powers of each user and the relay can only be reduced by up to $\frac{1}{N^{1-\varsigma}}$ when the pilot transmit power is set as $p_{\text{P}} = \frac{E_{\text{P}}}{N^{\varsigma}}$, in order to maintain a given spectral efficiency. Similarly, when $\rho < 1-\varsigma$ and $\theta < 1-\varsigma$, the asymptotic achievable rate of each user approaches to infinity as $N \to \infty$.

Remark 2.2: When the pilot power scaling factor $\varsigma = 1$, which means that the pilot power scales down by $\frac{1}{N}$, to guarantee user rate there is $\rho = \theta = 0$, which means the relay and user transmit power must stay constant and do not scale down with N. The achievable rate in this case can be derived from (2.19) and (2.20), but not shown here due to space limitation. It is found that channel estimation error induced interference and inter-pair interference cannot be eliminated when p_{P} is scaled down proportionally to $\frac{1}{N}$ with fixed p_i and P_R in case II.

It can be observed from Theorems 2.1 and 2.2 that for fixed σ_i $(i = 1, \cdots, 2K)$, σ_n, and p_{P}, the achievable rate of each pair-wise user transmission link depends on the user power, i.e., the values of p_i $(i = 1, \cdots, 2K)$, and the relay power. Next we propose the optimal power allocation for the studied system.

2.6 Power Allocation Schemes

In this section, a power allocation problem is first formulated and solved for multi-pair users in the MA phase transmission and the relay in the BC phase, which maximizes the sum spectral efficiency[3]. The achievable rate of a transmission link γ_k for $k \in \{1, \cdots, 2K\}$ used in the optimization refers to the SCSI based achievable rate, given in (2.19) and (2.20). Power allocation can be performed at the relay side according to the SCSI, and then the relay notifies the user pairs their allocated power

[3]The objective of power allocation can also be minimizing the total power consumption or maximizing the minimum achievable rate, which can be formulated and solved using the similar method.

values. Moreover, closed-form asymptotic power allocation solutions are presented for MRC/MRT and ZFR/ZFT, respectively, for the asymptotic regimes with high SNR and $N \to \infty$.

2.6.1 Optimal Power Allocation (OPA)

Most power optimization in communications aims to maximize the sum spectral efficiency, which is defined as the sum-rate (in bits) per channel use. Assuming that T is the length of the coherent interval (in symbols), in which τ symbols are used for channel estimation, the sum spectral efficiency[4] denoted as $S \triangleq \frac{T-\tau-2}{T} \sum_{k=1}^{2K} \gamma_k = \frac{T-\tau-2}{T} \log_2 \prod_{k=1}^{2K} (1+\chi_k)$, where $\chi_k = \frac{a_k p_{k'}}{\sum_{i=1}^{2K} \left(b_{k,i}^{(1)} + b_{k,i}^{(2)} P_{\mathrm{R}}^{-1}\right) p_i + c_k p_k + \left(d_k^{(1)} + d_k^{(2)} P_{\mathrm{R}}^{-1}\right)}$ is the signal-to-interference-plus-noise ratio (SINR). Then, the power allocation problem to maximize the sum spectral efficiency can be formulated as

$$\max_{p_i, P_{\mathrm{R}}} \quad \frac{T-\tau-2}{T} \log_2 \prod_{k=1}^{2K} (1+\chi_k) \tag{2.25a}$$

$$\text{s.t.} \quad \sum_{i=1}^{2K} p_i + P_{\mathrm{R}} \leq P \tag{2.25b}$$

$$0 \leq p_k \leq P_0, \ 0 \leq P_{\mathrm{R}} \leq P_{\mathrm{R},0}, \ k = 1, \cdots, 2K \tag{2.25c}$$

where P in constraint (2.25b) is the total power allocated to all the users and relay, and constraints (2.25c) specify the peak power limits P_0 and $P_{\mathrm{R},0}$ for each user k and relay, respectively. The objective in (2.25a) can be equivalently rewritten as $\min_{p_i, P_{\mathrm{R}}} \left[\prod_{k=1}^{2K} (1+\chi_k)\right]^{-1}$, as $\log_2(x)$ is a monotonic increasing function of x. We can see that the constraints are posynomial functions. If the objective function is a monomial or posynomial function, the problem (2.25) becomes a GP which can be reformulated as a convex problem, and thus, can be solved efficiently by convex optimization tools, such as CVX [21]. However, the rewritten objective function for (2.25) is still neither a monomial nor posynomial, making solving the problem directly by the convex optimization tools impossible. To solve this problem, an approximation for the objective function can be efficiently found by using the technique in [26]. Specifically, according to [26, Lemma 1], we can use a monomial function $\kappa_k \chi_k^{\eta_k}$ to approximate $(1+\chi_k)$ near an arbitrary point $\hat{\chi}_k > 0$, where

[4]Here, the loss due to relay sending out pairwise effective channel coefficients and SIC coefficients is taken into account. The transmission is assumed to be perfect and the overhead is 2 (symbols) for the pairwise effective channel coefficient and SIC coefficient per user.

$\eta_k \triangleq \hat{\chi}_k(1+\hat{\chi}_k)^{-1}$ and $\kappa_k \triangleq \hat{\chi}_k^{-\eta_k}(1+\hat{\chi}_k)$. Consequently, the objective function can be approximated as $\prod_{k=1}^{2K}(1+\chi_k) \approx \prod_{k=1}^{2K}\kappa_k\chi_k^{\eta_k}$, which is a monomial function. In this way, the problem is transformed into a GP problem by the approximation.

Similar to [26], a successive approximation algorithm for the power allocation problem in (2.25) is proposed as Algorithm 2.1. Notably, the parameter β here is utilized to control the desired approximation accuracy. The accuracy is high when β is close to 1, but the convergence rate is low, and vice versa. As shown in [26], $\beta = 1.1$ is an option that introduces a good accuracy trade off in most practical cases.

Algorithm 2.1: Successive approximation algorithm for power allocation

---

Initialization: Given tolerance $\varepsilon > 0$, the maximum number of iterations L, and parameter $\beta > 1$. Set $m = 1$. Select the initial values $\chi_{k,1}$ for χ_k $(k = 1, 2, \cdots, 2K)$.

Repeat:

1) Compute $\eta_{k,m} = \chi_{k,m}(1+\chi_{k,m})^{-1}$ and $\kappa_{k,m} = \chi_{k,m}^{-\eta_{k,m}}(1+\chi_{k,m})$;

2) Solve the GP:

$$\min_{p_i, P_{\mathrm{R}}, \chi_k} \quad \left[\prod_{k=1}^{2K}\kappa_{k,m}\chi_k^{\eta_{k,m}}\right]^{-1} \tag{2.26a}$$

$$\text{s.t.} \quad \frac{a_k p_{k'}}{\sum_{i=1}^{2K}\left(b_{k,i}^{(1)} + b_{k,i}^{(2)}P_{\mathrm{R}}^{-1}\right)p_i + c_k p_k + \left(d_k^{(1)} + d_k^{(2)}P_{\mathrm{R}}^{-1}\right)} \le \chi_k \tag{2.26b}$$

$$(2.25\text{b}),\ (2.25\text{c}),\ \beta^{-1}\chi_{k,m} \le \chi_k \le \beta\chi_{k,m} \tag{2.26c}$$

3) Set $m = m + 1$, and update $\chi_{k,m} = \chi_k^*$, where χ_k^* $(k = 1, 2, \cdots, 2K)$ are obtained based on the solutions p_i^* $(i = 1, 2, \cdots, 2K)$ and P_{R}^* of the GP;

Until: Stop if $\max_k \left|\frac{\chi_{k,m}-\chi_{k,m-1}}{\chi_{k,m}}\right| < \varepsilon$ or $m = L$;

Output: Output p_i^* $(i = 1, 2, \cdots, 2K)$ and P_{R}^* as the solutions.

2.6.2 Asymptotically Optimal Power Allocation (AOPA)

Obviously, the optimal power allocation scheme in Algorithm 2.1 is an iterative numerical solution with no closed-form. However, more tractable expressions can be found for MRC/MRT and ZFR/ZFT, respectively, when we consider the asymptotic regimes with high SNR and $N \to \infty$.

AOPA for MRC/MRT:

Suppose that the SNRs at both the relay and user sides are very high, i.e., $P_{\mathrm{R}} \gg \sigma_n^2$, $p_{\mathrm{P}} \gg \sigma_n^2$ and $p_i \gg \sigma_n^2$ $(i = 1, \cdots, 2K)$, and $\frac{N}{K} = \lambda$ with fixed λ and $K \to \infty$. Then the lower bound can be simplified as given by the following Lemma.

Lemma 2.1: When $P_{\mathrm{R}} \gg \sigma_n^2$, $p_{\mathrm{P}} \gg \sigma_n^2$, $p_i \gg \sigma_n^2$ $(i = 1, \cdots, 2K)$, and $N \gg K \to \infty$, the rate of the transmission link $k \to k'$ can be approximated as

$$\gamma_{k'}^{\mathrm{SCSI}} \approx \log_2 \left(1 + \frac{N \sigma_k^4 \sigma_{k'}^2 p_k}{\sum\limits_{i \neq k'}^{2K} \left(\sigma_i^2 \sigma_{k'}^2 \sigma_k^2 + \sigma_i^4 \sigma_{i'}^2 \right) p_i} \right). \tag{2.27}$$

Proof: Firstly, we have $\hat{\sigma}_i^2 \approx \sigma_i^2$ for $i = 1, \cdots, 2K$ due to $p_{\mathrm{P}} \gg \sigma_n^2$. Then, we divide both the denominator and numerator of the SINR in (2.19) by $(N+1)$. Each item with $\frac{1}{N+1}$ in the denominator is able to be ignored based on $N \to \infty$. Then, according to $\frac{\sigma_n^2}{P_{\mathrm{R}}} \approx 0$ and $p_i \gg \sigma_n^2$ for $i = 1, \cdots, 2K$, (2.27) can be obtained.

In order to obtain a closed-form solution for asymptotically optimal power allocation, we set the fixed link condition $\sigma_i^2 \sigma_{i'}^2 = C$ for $i = 1, \cdots, 2K$. Since the approximated rate expressions involve no P_{R}, we do not need to find the optimal solution for P_{R}. For a fixed P_{R}, (2.25) can be rewritten as

$$\max_{p_i} \quad \prod_{k=1}^{2K} \left(1 + \frac{N \sigma_{k'}^2 p_{k'}}{2 \sum\limits_{i \neq k}^{2K} \sigma_i^2 p_i} \right) \tag{2.28a}$$

$$\text{s.t.} \quad \sum_{i=1}^{2K} p_i \leq P - P_{\mathrm{R}} \tag{2.28b}$$

$$0 \leq p_k, \ k = 1, 2, \cdots, 2K. \tag{2.28c}$$

where the peak power constraints for each user are ignored for analysis simplicity by assuming that the channel large-scale factors are on the same order of magnitude. Then, we have the following *Theorem 3* with regard to the optimal allocated power

for each user.

Theorem 2.3: The optimal solution to (2.28) is obtained as

$$p_i^* = \frac{P - P_{\mathrm{R}}}{\sigma_i^2 \sum\limits_{k=1}^{2K} \frac{1}{\sigma_k^2}}. \tag{2.29}$$

Proof: Since $\left| \frac{1}{2K-1} \sum\limits_{i \neq k'}^{2K} \sigma_i^2 p_i - \frac{1}{2K} \sum\limits_{i=1}^{2K} \sigma_i^2 p_i \right| \xrightarrow{p} 0$ [27] in probability when $K \to \infty$, according to the inequality of arithmetic and geometric means, the objective in (2.28a) is upper bounded by

$$\begin{aligned} &\prod_{k=1}^{2K} \frac{(2K-1) \sum\limits_{i=1}^{2K} \sigma_i^2 p_i + NK\sigma_{k'}^2 p_{k'}}{(2K-1) \sum\limits_{i=1}^{2K} \sigma_i^2 p_i} \\ \leq &\left(\sum_{k=1}^{2K} \frac{(2K-1) \sum\limits_{i=1}^{2K} \sigma_i^2 p_i + NK\sigma_{k'}^2 p_{k'}}{2K\,(2K-1) \sum\limits_{i=1}^{2K} \sigma_i^2 p_i} \right)^{2K} \\ = &\left(\frac{1}{2K} + \frac{N}{2\,(2K-1)} \right)^{2K} \end{aligned} \tag{2.30}$$

where the equality is achieved if and only if $\sigma_i^2 p_i = A$ for $\forall i \in \{1, \cdots, 2K\}$. To maximize the sum rate, it is obvious that the total user power should reach the largest value $P - P_{\mathrm{R}}$ in (2.28b). Thus (2.29) can be obtained to satisfy $\sum\limits_{i=1}^{2K} p_i = P - P_{\mathrm{R}}$, indicating that the optimal allocated power for each user is inverse to its corresponding large-scale fading factor of the channel from the user to the relay, and proportional to the channel from the relay to its destination when MRC/MRT beamforming is used at the relay under the condition that the link end-to-end large-scale fading factors among all pairs are equal. Furthermore, the asymptotic sum-rate is independent of the allocated power.

AOPA for ZFR/ZFT:

Similarly, we make the same assumption for ZFR/ZFT that the SNRs at both the relay and user sides are very high and $N \gg K \to \infty$. Then the following Lemma can be obtained.

Lemma 2.2: When $P_{\mathrm{R}} \gg \sigma_n^2$, $p_{\mathrm{P}} \gg \sigma_n^2$, $p_i \gg \sigma_n^2$ $(i = 1, \cdots, 2K)$, and $N \gg$

$K \to \infty$, the rate of the transmission link $k \to k'$ can be approximated as

$$\gamma_{k'}^{\mathrm{SCSI}} \approx \log_2 \left(1 + \frac{(N - 2K - 1)\,\sigma_k^2 p_k}{\sigma_n^2}\right). \tag{2.31}$$

Therefore, without any assumptions on link conditions, (2.25) can be rewritten as

$$\max_{p_i} \prod_{k=1}^{2K} \left(1 + \frac{(N - 2K - 1)\,\sigma_{k'}^2 p_{k'}}{\sigma_n^2}\right) \tag{2.32a}$$

$$\text{s.t.} \quad (2.28b), (2.28c) \tag{2.32b}$$

Theorem 2.4 states the optimal allocated power for each user.

Theorem 2.4: The optimal solution to (2.32) is given by

$$p_i^* = \left[\frac{1}{\lambda} - \frac{\sigma_n^2}{(N - 2K - 1)\,\sigma_i^2}\right]^+ \tag{2.33}$$

where $\lambda = 2K \Big/ \left(\sum_{i=1}^{2K} \frac{\sigma_n^2}{(N-2K-1)\sigma_i^2} + \frac{P}{2}\right)$ is chosen to satisfy (2.28b).

Proof: Obviously, we can obtain the solution to (2.32) by the Lagrange multiplier approach associated with Karush-Kuhn-Tucker (KKT) conditions.

From (2.33), it can be concluded that the allocated powers are equal for ZFR/ZFT when $N - 2K \to \infty$.

2.7 Numerical Results

Simulations are conducted to validate the derived achievable rate expressions and examine the performance of the designed power allocation schemes, respectively. In the simulation study, we set the length of the coherent interval $T = 200$ (symbols), the number of user pairs $K = 10$, the training length $\tau = 2K$, and the noise variance is normalized to be $\sigma_n^2 = 1$. Furthermore, SNR $= P_{\mathrm{R}}$ is defined at the relay side.

2.7.1 Validation of Achievable Rate Results

Firstly, the effectiveness of the derived SCSI based achievable rate in (2.19) and (2.20) is evaluated by comparing the spectral efficiency with the Monte-Carlo simulation results. For simplicity, we assume that the large-scale fading factors are $\sigma_i^2 = 1$ for all $i = 1, 2, \cdots, 2K$ and equal power allocation for users is utilized with the total power $2KP_{\mathrm{S}} = P_{\mathrm{R}}$. In Fig. 2.2 with $p_{\mathrm{P}} = 10$ dB, the spectral efficiency

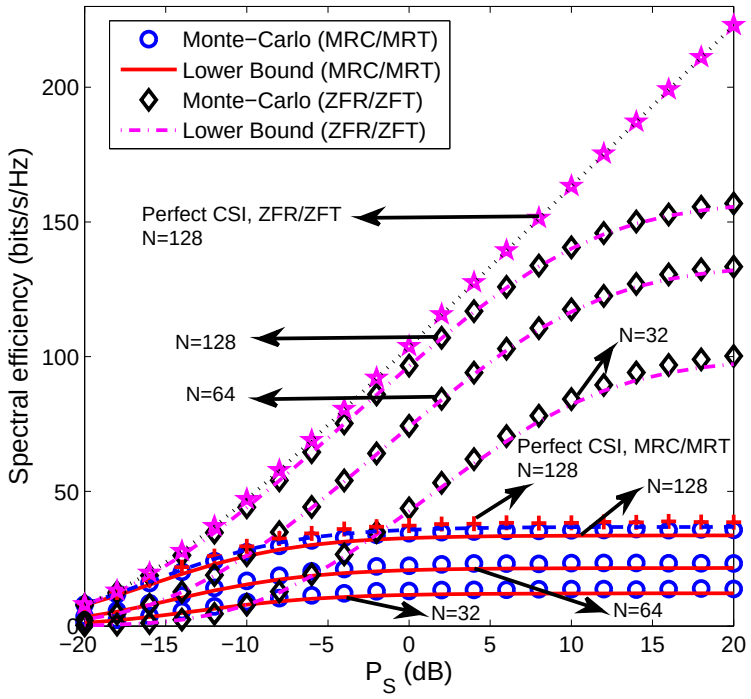

Figure 2.2: Spectral efficiency versus SNR for lower bounds and Monte-Carlo results ($p_{\mathrm{P}} = 10$ dB, EPA, $P_{\mathrm{R}} = 2KP_{\mathrm{S}}$).

curves versus P_{S} obtained from the analytical lower bounds (2.19) and (2.20), are compared with the ones given by the exact capacity expression (2.14) obtained through Monte-Carlo simulation. It is evident that the relative performance gap between the capacity lower bound (2.20) and the exact capacity (2.14) for ZFR/ZFT is even smaller than that for MRC/MRT, especially with larger number of antennas. Moreover, Fig. 2.2 shows that the spectral efficiency of ZFR/ZFT increases much faster than that of MRC/MRT as SNR increases. It is due to the fact that the effect of interference is much larger than that of the noise for higher SNR while ZFR/ZFT is able to null multi-user interference signals [23]. Hence, the effectiveness of the derived closed-form lower bounds for both MRC/MRT and ZFR/ZFT has been demonstrated.

Fig. 2.3 compares the spectral efficiency of MRC/MRT with that of ZFR/ZFT at different SNRs, which shows that ZFR/ZFT does not always outperform MRC/MRT in the massive MIMO two-way relaying systems. As K increases, i.e., the average SNR ($\frac{P_{\mathrm{R}}}{2K}$) of each user decreases, the noise effect exceeds the interference effect causing wore performance of ZFR/ZFT. Fig. 2.4 depicts the performance of ZFR/ZFT over MRC/MRT with different N/K ratios. Different from the results obtained by Fig. 2.3, Fig. 2.4 shows that under a fixed N/K, the gains brought by

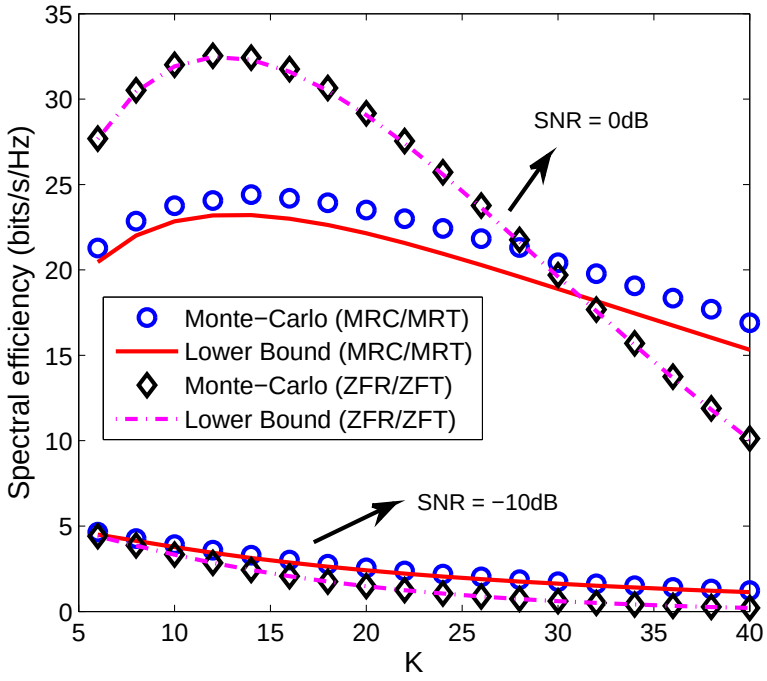

Figure 2.3: Spectral efficiency versus K for lower bounds and Monte-Carlo results ($p_{\mathrm{P}} = 10$ dB, EPA, SNR $= P_{\mathrm{R}} = 2KP_{\mathrm{S}}$, $N = 128$).

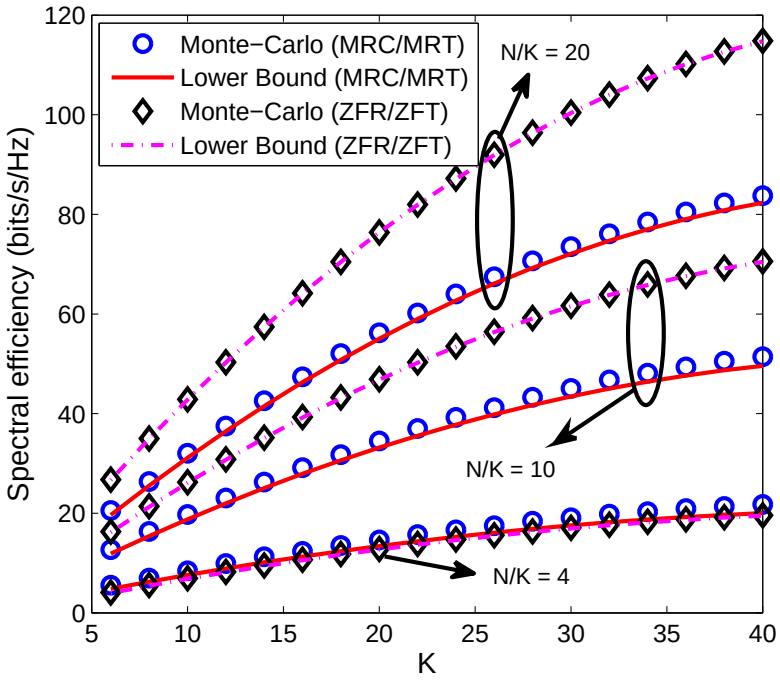

Figure 2.4: Spectral efficiency versus K for lower bounds and Monte-Carlo results ($p_{\mathrm{P}} = 10$ dB, EPA, SNR $= P_{\mathrm{R}} = 2KP_{\mathrm{S}} = 0$ dB).

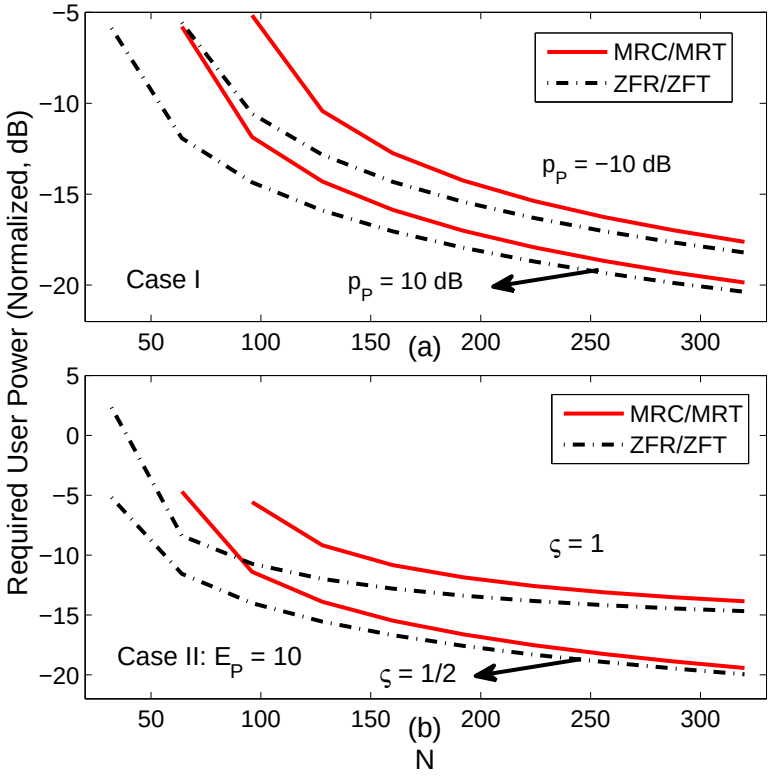

Figure 2.5: Required user power to achieve 1 bit/s/Hz per user for MRC/MRT and ZFR/ZFT (EPA, $P_{\mathrm{R}} = 2KP_{\mathrm{S}}$).

ZFT/ZFR grow even though the average SNR ($\frac{P_{\mathrm{R}}}{2K}$) of each user decreases as the number of user pairs K increases. These observations indicate that two asymptotic regimes, large-scale antenna arrays and high SNR, are equivalent [7]. Moreover, larger gains over MRC/MRT are achieved by the ZFR/ZFT processing at higher N/K.

Next the asymptotic analyses with massive arrays for the two cases in Propositions 2.1 and 2.2 are examined, supposing $P_{\mathrm{R}} = 2K * P_{\mathrm{S}}$ and EPA employed with the transmit power at each user satisfying $p_i = P_{\mathrm{S}}$ $(i = 1, 2, \cdots, 2K)$. Fig. 2.5 shows the required user transmit power P_{S} to achieve 1 bit/s/Hz per user. It is obvious from Fig. 2.5(a) that in case I where the pilot power p_{P} is fixed, the required user transmit power is significantly reduced as N increases, and that the required P_{S} with ZFR/ZFT is lower than that with MRC/MRT. Regarding the imperfect CSI effect, less user transmit power is required when p_{P} is high. On the other hand, when p_{P} is low and N is small, the required 1 bit/s/Hz achievable rate per user cannot be achieved even with infinite P_{S}, which means that the only way to reduce the imperfect CSI effect and thus achieve required spectral efficiency is to increase the number of antennas at the relay. For case II with $E_{\mathrm{P}} = 10$ and the pilot power

scaling down by $p_{\mathrm{P}} = \frac{E_{\mathrm{P}}}{N^{\varsigma}}$, Fig. 2.5(b) shows that higher ς leads to more slowly reduced P_{S}, because the imperfect CSI effect becomes much severer when the pilot power is reduced faster with the increase of N.

2.7.2 Power Allocation

In this subsection, the proposed power allocation schemes in Section 2.6 are examined in regard to the performance of the spectral efficiency.

In OPA, simulations are performed assuming that $P_0 = 10$ dB, $P_{\mathrm{R},0} = 23$ dB and $P = 23$ dB. First, we choose the large-scale fading matrix as follows

$$\begin{aligned}\mathbf{D} =& diag\{0.749\ 0.045\ 0.246\ 0.121\ 0.125\ 0.142\ 0.635\ 0.256\ 0.021\ 0.123\ 0.257\\ & 0.856\ 1.000\ 0.899\ 0.014\ 0.759\ 0.315\ 0.432\ 0.195\ 0.562\}\end{aligned}$$

which is a snapshot of the practical setup, indicating that all large-scale fading factors fall into the interval $[0.014, 1.000]$. Fig. 2.6 shows the spectral efficiency versus p_{P} with fixed $N = 32$, 64 and 128 under both OPA and EPA. The employed EPA here allocates equal power to each user where the sum power consumed by all users achieves its maximum value $P/2$ with $P_{\mathrm{R}} = P/2$, i.e., $p_i = \frac{P}{4K}$ for all $i = 1, 2, \cdots, 2K$. For the OPA, Algorithm 2.1 is utilized with the initial values chosen as follows: $\varepsilon = 0.01$, $L = 10$, $\beta = 1.1$, and $\chi_{k,1} = \frac{a_k P}{\sum_{i=1}^{2K}\left(b_{k,i}^{(1)}+2b_{k,i}^{(2)}P^{-1}\right)P+c_k P+\left(4Kd_k^{(1)}+8Kd_k^{(2)}P^{-1}\right)}$ $(k = 1, \cdots, 2K)$ are obtained by the EPA scheme. It can be observed from Fig. 2.6 that OPA outperforms EPA, especially when the number of relay antennas is high, which demonstrates the effectiveness of our proposed PA. Furthermore, the spectral efficiency improvement in OPA for MRC/MRT beamforming is always smaller than that for ZFR/ZFT under different p_{P}.

For AOPA, Fig. 2.7 validates the effectiveness of (2.29) and (2.33) at a large N and a high SNR in comparison to OPA and EPA. Notably, the OPA here considers fixed $P_{\mathrm{R}} = P/2$, i.e., only user power allocation is performed. Fig. 2.7(a) shows that when $P_{\mathrm{R}} = p_{\mathrm{P}} = 20$ dB, the obtained spectral efficiency in AOPA for MRC/MRT[5] is almost the same as that in OPA, reaching high gains over EPA. However, if the SNR is reduced to $P_{\mathrm{R}} = p_{\mathrm{P}} = 0$ dB, the gain achieved by AOPA over EPA becomes rather slight even when the number of antennas is significantly increased. While regarding the ZFR/ZFT, Fig. 2.7(b) illustrates that when the number of antennas is large and SNR is high, AOPA tends to EPA as predicted by (2.33). Furthermore,

[5]In the AOPA for MRC/MRT, we set $\sigma_{2i}^2 = 1/\sigma_{2i-1}^2$ for $i = 1, \cdots, K$ according to the assumption in Section 2.6.

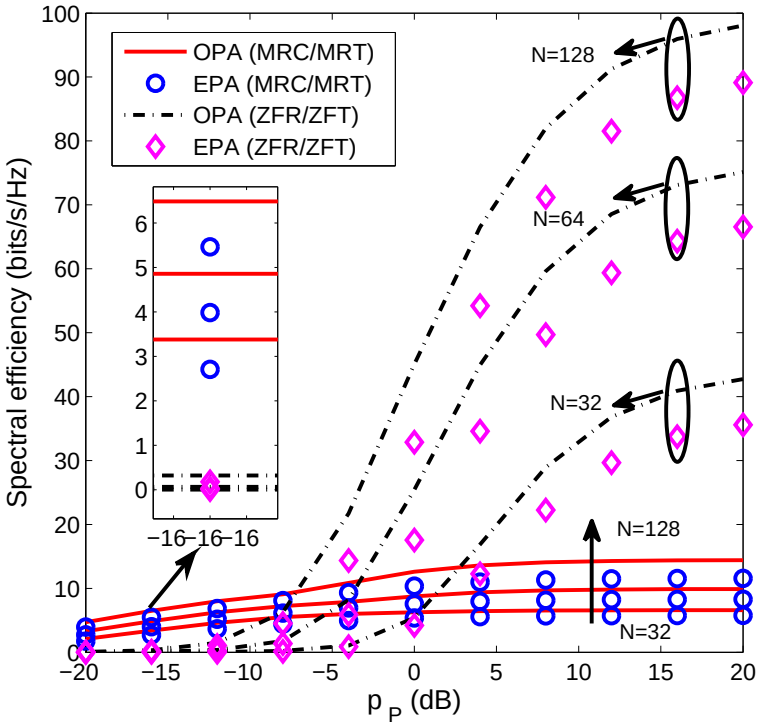

Figure 2.6: Spectral efficiency versus p_{P} ($P_0 = 10$ dB, $P_{\mathrm{R},0} = 23$ dB, $P = 23$ dB).

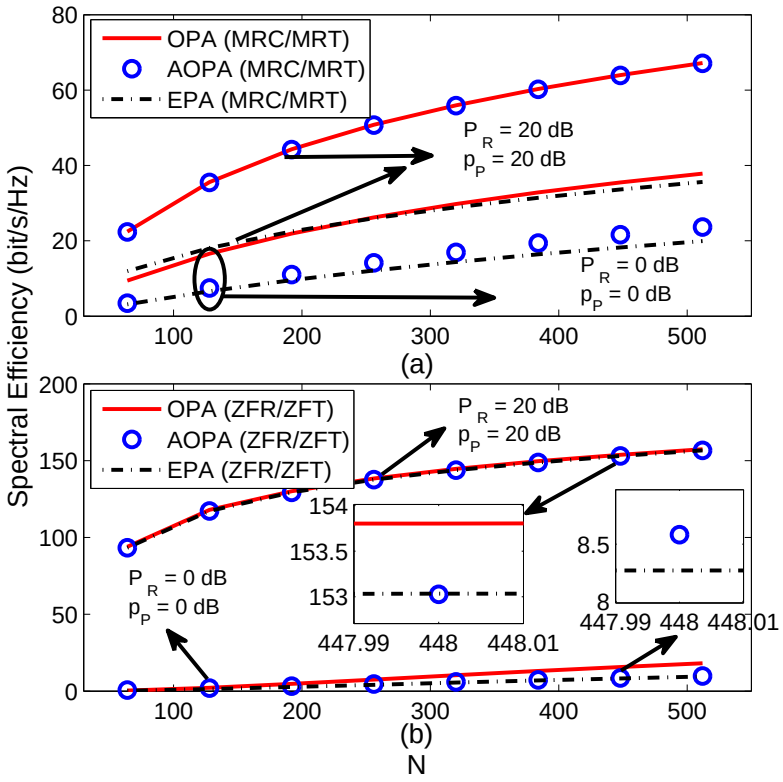

Figure 2.7: Spectral efficiency versus N for different power allocation schemes ($P_{\mathrm{R}} = P/2$).

Fig. 2.7(b) shows that AOPA outperforms EPA slightly when the SNR is reduced to $P_{\mathrm{R}} = p_{\mathrm{P}} = 0$ dB.

2.8 Conclusion

In this chapter, closed-form ergodic achievable rate expressions have been derived for a multi-pair massive MIMO two-way AF relaying system with imperfect CSI and linear processing. Optimal power allocations schemes based on the obtained rate expressions have been shown to outperform equal power allocation in various scenarios. It has been found that the asymptotically optimal power solutions for MRC/MRT and ZFR/ZFT achieve almost the same performance as OPA when the SNR is high and the number of antennas at the relay is large. Both AOPA and OPA outperform EPA on the spectral efficiency. Besides, the allocated power of each user in AOPA is inverse to the large-scale fading factor of the channel from the user to the relay, and proportional to the channel from the relay to its destination for MRC/MRT under the condition that the link end-to-end large-scale fading factors among all pairs are equal, and the AOPA for ZFR/ZFT tends to be EPA when N is large.

2.9 Appendices

2.9.1 Proof of Equation (2.10)

To prove (2.10), we start from the expectation $\mathrm{E}\left[\left\|\hat{\mathbf{G}}^{*}\mathbf{T}\hat{\mathbf{G}}^{H}\tilde{\mathbf{G}}\mathbf{x}\right\|^{2}\right]$, which is rewritten as

$$\begin{aligned}\mathrm{E}\left[\left\|\hat{\mathbf{G}}^{*}\mathbf{T}\hat{\mathbf{G}}^{H}\tilde{\mathbf{G}}\mathbf{x}\right\|^{2}\right] &\overset{(a)}{=} \mathrm{tr}\left\{\mathrm{E}\left[\hat{\mathbf{G}}^{*}\mathbf{T}\hat{\mathbf{G}}^{H}\left(\hat{\mathbf{G}}-\boldsymbol{\Xi}\right)\mathbf{P}\mathbf{P}^{H}\left(\hat{\mathbf{G}}-\boldsymbol{\Xi}\right)^{H}\hat{\mathbf{G}}\mathbf{T}^{H}\hat{\mathbf{G}}^{T}\right]\right\} \\ &\overset{(b)}{=} \mathrm{tr}\left\{\mathrm{E}\left[\hat{\mathbf{G}}\mathbf{P}\mathbf{P}^{H}\hat{\mathbf{G}}^{H}\hat{\mathbf{G}}\mathbf{T}^{H}\hat{\mathbf{G}}^{T}\hat{\mathbf{G}}^{*}\mathbf{T}\hat{\mathbf{G}}^{H}\right]\right\} + \mathrm{tr}\left\{\mathrm{E}\left[\boldsymbol{\Xi}\mathbf{P}\mathbf{P}^{H}\boldsymbol{\Xi}^{H}\hat{\mathbf{G}}\mathbf{T}^{H}\hat{\mathbf{G}}^{T}\hat{\mathbf{G}}^{*}\mathbf{T}\hat{\mathbf{G}}^{H}\right]\right\}\end{aligned} \tag{2.34}$$

where step (a) is obtained by $\mathrm{E}\left[\mathbf{x}\mathbf{x}^{H}\right] = \mathbf{I}_{2K}$ and substituting $\tilde{\mathbf{G}} = \mathbf{G}\mathbf{P}$ and (2.4) into the equation, and step (b) results from the independence between $\hat{\mathbf{G}}$ and $\boldsymbol{\Xi}$

and the property $\mathrm{tr}\{\mathbf{AB}\} = \mathrm{tr}\{\mathbf{BA}\}$. As for the first term in (2.34), we have

$$\begin{aligned}
&\mathrm{tr}\left\{\mathrm{E}\left[\hat{\mathbf{G}}\mathbf{P}\mathbf{P}^H\hat{\mathbf{G}}^H\hat{\mathbf{G}}\mathbf{T}^H\hat{\mathbf{G}}^T\hat{\mathbf{G}}^*\mathbf{T}\hat{\mathbf{G}}^H\right]\right\} \\
\overset{(a)}{=}&\mathrm{tr}\left\{\sum_{k=1}^{2K} p_k \mathrm{E}\left[\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\sum_{i=1}^{K}\left(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i}^T+\hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i-1}^T\right)\left(\hat{\mathbf{g}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H+\hat{\mathbf{g}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\right]\right\} \\
\overset{(b)}{=}&2N\left(N+1\right)\hat{\Psi}\hat{\Phi}+N(N+1)^2\sum_{i=1}^{K}\hat{\psi}_i\hat{\phi}_i
\end{aligned} \tag{2.35}$$

where $\hat{\Psi} = \sum_{i=1}^{K}\hat{\psi}_i$ with $\hat{\psi}_i = p_{2i-1}\hat{\sigma}_{2i-1}^2 + p_{2i}\hat{\sigma}_{2i}^2$ and $\hat{\Phi} = \sum_{i=1}^{K}\hat{\phi}_i$ with $\hat{\phi}_i = \hat{\sigma}_{2i-1}^2\hat{\sigma}_{2i}^2$. Step (a) in (2.35) results from substituting $\hat{\mathbf{G}}$ in (2.4), $\mathbf{P}$ and $\mathbf{T}$ into the equation and formula expansion based on the fact that the expectation of $\sum_{k=1}^{2K}\mathrm{E}\left[\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\left(\hat{\mathbf{g}}_{2j-1}\hat{\mathbf{g}}_{2j}^T+\hat{\mathbf{g}}_{2j}\hat{\mathbf{g}}_{2j-1}^T\right)\left(\hat{\mathbf{g}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H+\hat{\mathbf{g}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\right] = 0$ for any $i \neq j$. Step (b) is obtained by $\mathrm{tr}\{\mathbf{AB}\} = \mathrm{tr}\{\mathbf{BA}\}$, some results from Gaussian distributed estimated channel in (2.4)[6] [28], and [29, Lemma 2.9]. To elaborate in detail, for the items in (a) with any $k \neq 2i-1$ or $2i$, we have

$$\begin{aligned}
&\mathrm{tr}\left\{\mathrm{E}\left[\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i-1}^T\hat{\mathbf{g}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H\right]\right\} \\
=\ &\mathrm{tr}\left\{\mathrm{E}\left[\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\right]\mathrm{E}\left[\hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i-1}^H\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i}^H\right]\right\} = N^2\hat{\sigma}_{2i-1}^2\hat{\sigma}_{2i}^2\hat{\sigma}_k^2
\end{aligned} \tag{2.36}$$

$$\begin{aligned}
&\mathrm{tr}\left\{\mathrm{E}\left[\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i}^T\hat{\mathbf{g}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H\right]\right\} \\
=\ &\mathrm{tr}\left\{\mathrm{E}\left[\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\right]\mathrm{E}\left[\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i-1}^H\right]\mathrm{E}\left[\hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i}^H\right]\right\} = N\hat{\sigma}_{2i-1}^2\hat{\sigma}_{2i}^2\hat{\sigma}_k^2
\end{aligned} \tag{2.37}$$

where the properties $\mathbf{x}^*\mathbf{y}^T = \mathbf{y}\mathbf{x}^H$ and $\mathbf{x}^T\mathbf{y}^* = \mathbf{y}^H\mathbf{x}$ for arbitrary vectors $\mathbf{x}$ and $\mathbf{y}$ are utilized. While for the items with $k = 2i-1$ and $2i$, we have

$$\begin{aligned}
&\mathrm{tr}\left\{\mathrm{E}\left[\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i-1}^H\left(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i}^T+\hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i-1}^T\right)\left(\hat{\mathbf{g}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H+\hat{\mathbf{g}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\right]\right\} \\
=\ &2N\left(N+1\right)\hat{\sigma}_{2i}^2\hat{\sigma}_{2i-1}^4+N\left(N+1\right)^2\hat{\sigma}_{2i}^2\hat{\sigma}_{2i-1}^4
\end{aligned} \tag{2.38}$$

$$\begin{aligned}
&\mathrm{tr}\left\{\mathrm{E}\left[\hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i}^H\left(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i}^T+\hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i-1}^T\right)\left(\hat{\mathbf{g}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H+\hat{\mathbf{g}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\right]\right\} \\
=\ &2N\left(N+1\right)\hat{\sigma}_{2i-1}^2\hat{\sigma}_{2i}^4+N\left(N+1\right)^2\hat{\sigma}_{2i-1}^2\hat{\sigma}_{2i}^4.
\end{aligned} \tag{2.39}$$

[6]Due to the estimated channel model in (2.4), we have that $\hat{\mathbf{g}}_i$ and $\hat{\mathbf{g}}_j$ are mutually independent $N\times 1$ vectors with $\forall i \neq j$ whose elements are i.i.d. zero-mean Gaussian distributed with variances $\hat{\sigma}_i^2$ and $\hat{\sigma}_j^2$, respectively. Then, it can be concluded that $\mathrm{E}\left[\hat{\mathbf{g}}_i^H\hat{\mathbf{g}}_i\right] = N\hat{\sigma}_i^2$, $\mathrm{E}\{\hat{\mathbf{g}}_j^H\hat{\mathbf{g}}_j\} = N\hat{\sigma}_j^2$, and $\mathrm{E}\left[\hat{\mathbf{g}}_i^H\hat{\mathbf{g}}_j\right] = 0$. Also, we can obtain that $\mathrm{E}\left[|\hat{\mathbf{g}}_i^H\hat{\mathbf{g}}_j|^2\right] = N\hat{\sigma}_i^2\hat{\sigma}_j^2$.

On account of the independence between $\hat{\mathbf{G}}$ and $\mathbf{\Xi}$, the second term in (2.34) becomes

$$\begin{aligned} & \mathrm{tr}\left\{\mathrm{E}\left[\mathbf{\Xi}\mathbf{P}\mathbf{P}^H\mathbf{\Xi}^H\hat{\mathbf{G}}\mathbf{T}^H\hat{\mathbf{G}}^T\hat{\mathbf{G}}^*\mathbf{T}\hat{\mathbf{G}}^H\right]\right\} \\ = & \sum_{j=1}^{2K} p_j\sigma_{\xi_j}^2\,\mathrm{tr}\left\{\mathrm{E}\left[\sum_{i=1}^{K}\left(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i}^T+\hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i-1}^T\right)\left(\hat{\mathbf{g}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H+\hat{\mathbf{g}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\right]\right\} \\ = & 2N\left(N+1\right)\hat{\Phi}\sum_{j=1}^{2K}p_j\sigma_{\xi_j}^2. \end{aligned} \tag{2.40}$$

By substituting (2.35) and (2.40) into (2.34), we have

$$\mathrm{E}\left[\left\|\hat{\mathbf{G}}^*\mathbf{T}\hat{\mathbf{G}}^H\tilde{\mathbf{G}}\mathbf{x}\right\|^2\right]=N\left(N+1\right)\left[2\Psi\hat{\Phi}+\left(N+1\right)\sum_{i=1}^{K}\hat{\psi}_i\hat{\phi}_i\right] \tag{2.41}$$

where $\Psi=\sum_{i=1}^{K}\psi_i$ with $\psi_i=p_{2i-1}\sigma_{2i-1}^2+p_{2i}\sigma_{2i}^2$.

To proceed, we need to calculate the expectation $\mathrm{E}\left[\left\|\hat{\mathbf{G}}^*\mathbf{T}\hat{\mathbf{G}}^H\mathbf{n}_{\mathrm{r}}\right\|^2\right]$ in (2.10), which is elaborated as follows

$$\begin{aligned} & \mathrm{E}\left[\left\|\hat{\mathbf{G}}^*\mathbf{T}\hat{\mathbf{G}}^H\mathbf{n}_{\mathrm{r}}\right\|^2\right]\overset{(a)}{=}\mathrm{tr}\left\{\sigma_n^2\mathrm{E}\left[\hat{\mathbf{G}}^*\mathbf{T}\hat{\mathbf{G}}^H\hat{\mathbf{G}}\mathbf{T}^H\hat{\mathbf{G}}^T\right]\right\} \\ \overset{(b)}{=} & \mathrm{tr}\left\{\sigma_n^2\mathrm{E}\left[\sum_{i=1}^{K}\left(\hat{\mathbf{g}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H+\hat{\mathbf{g}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\left(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i}^T+\hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i-1}^T\right)\right]\right\} \\ \overset{(c)}{=} & 2N\left(N+1\right)\sigma_n^2\hat{\Phi} \end{aligned} \tag{2.42}$$

where step (a) is obtained by $\mathrm{E}\left[\mathbf{n}_{\mathrm{r}}\mathbf{n}_{\mathrm{r}}^H\right]=\sigma_n^2\mathbf{I}_N$ and $\mathrm{tr}\left\{\mathbf{AB}\right\}=\mathrm{tr}\left\{\mathbf{BA}\right\}$, (b) results from substituting $\hat{\mathbf{G}}$ in (2.4), $\mathbf{P}$ and $\mathbf{T}$ into the equation and the fact that the expectation of $\mathrm{E}\left[\left(\hat{\mathbf{g}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H+\hat{\mathbf{g}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\left(\hat{\mathbf{g}}_{2j-1}\hat{\mathbf{g}}_{2j}^T+\hat{\mathbf{g}}_{2j}\hat{\mathbf{g}}_{2j-1}^T\right)\right]=0$ for any $i\neq j$, and step (c) results from the properties of $\mathrm{tr}\left\{\mathbf{AB}\right\}=\mathrm{tr}\left\{\mathbf{BA}\right\}$, $\mathbf{x}^*\mathbf{y}^T=\mathbf{y}\mathbf{x}^H$ and $\mathbf{x}^T\mathbf{y}^*=\mathbf{y}^H\mathbf{x}$ for arbitrary vectors $\mathbf{x}$ and $\mathbf{y}$, and the properties of Gaussian distributed vectors, respectively, the detailed derivation of which is

$$\begin{aligned} & \mathrm{tr}\left\{\mathrm{E}\left[\left(\hat{\mathbf{g}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H+\hat{\mathbf{g}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\left(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i}^T+\hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i-1}^T\right)\right]\right\} \\ = & 2\mathrm{E}\left[\hat{\mathbf{g}}_{2i}^H\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i-1}^H\hat{\mathbf{g}}_{2i}\right]+2\mathrm{E}\left[\hat{\mathbf{g}}_{2i}^H\hat{\mathbf{g}}_{2i}\right]\mathrm{E}\left[\hat{\mathbf{g}}_{2i-1}^H\hat{\mathbf{g}}_{2i-1}\right] \\ = & 2N\left(N+1\right)\hat{\sigma}_{2i-1}^2\hat{\sigma}_{2i}^2. \end{aligned} \tag{2.43}$$

Hence, by substituting (2.41) and (2.42) into the step (a) in (2.10), the proof of (2.10) is completed.

2.9.2 Proof of Equation (2.13)

To prove (2.13), likewise, we start from $\mathrm{E}\left[\left\|\hat{\tilde{\mathbf{G}}}^{*}\mathbf{T}\hat{\tilde{\mathbf{G}}}^{H}\tilde{\mathbf{G}}\mathbf{x}\right\|^{2}\right]$, which can be rewritten as

$$\begin{aligned}
&\mathrm{E}\left[\left\|\hat{\tilde{\mathbf{G}}}^{*}\mathbf{T}\hat{\tilde{\mathbf{G}}}^{H}\tilde{\mathbf{G}}\mathbf{x}\right\|^{2}\right] = \mathrm{tr}\left\{\mathrm{E}\left[\hat{\tilde{\mathbf{G}}}^{*}\mathbf{T}\hat{\tilde{\mathbf{G}}}^{H}\left(\hat{\mathbf{G}}-\boldsymbol{\Xi}\right)\mathbf{P}\mathbf{P}^{H}\left(\hat{\mathbf{G}}-\boldsymbol{\Xi}\right)^{H}\hat{\tilde{\mathbf{G}}}\mathbf{T}\hat{\tilde{\mathbf{G}}}^{T}\right]\right\} \\
&\overset{(a)}{=} \mathrm{tr}\left\{\mathrm{E}\left[\hat{\tilde{\mathbf{G}}}^{*}\mathbf{T}\hat{\tilde{\mathbf{G}}}^{H}\hat{\mathbf{G}}\mathbf{P}\mathbf{P}^{H}\hat{\mathbf{G}}^{H}\hat{\tilde{\mathbf{G}}}\mathbf{T}\hat{\tilde{\mathbf{G}}}^{T}\right]\right\} + \mathrm{tr}\left\{\mathrm{E}\left[\boldsymbol{\Xi}\mathbf{P}\mathbf{P}^{H}\boldsymbol{\Xi}^{H}\hat{\tilde{\mathbf{G}}}\mathbf{T}\hat{\tilde{\mathbf{G}}}^{T}\hat{\tilde{\mathbf{G}}}^{*}\mathbf{T}\hat{\tilde{\mathbf{G}}}^{H}\right]\right\} \\
&\overset{(b)}{=} \mathrm{tr}\left\{\mathrm{E}\left[\hat{\tilde{\mathbf{G}}}^{*}\mathbf{P}_{T}\hat{\tilde{\mathbf{G}}}^{T}\right]\right\} + \sum_{i=1}^{2K} p_i\sigma_{\xi_i}^{2}\,\mathrm{tr}\left\{\mathrm{E}\left[\hat{\boldsymbol{\Omega}}^{*}\mathbf{T}\hat{\boldsymbol{\Omega}}\mathbf{T}\right]\right\} \\
&\overset{(c)}{=} \sum_{i=1}^{2K} p_{i'}\mathrm{E}\left[\hat{\tilde{\mathbf{g}}}_i^{H}\hat{\tilde{\mathbf{g}}}_i\right] + \sum_{i=1}^{2K} p_i\sigma_{\xi_i}^{2}\sum_{j=1}^{2K}\left(\mathrm{E}\left[\hat{\omega}_{j,j}^{*}\hat{\omega}_{j',j'}\right] + \mathrm{E}\left[|\hat{\omega}_{j,j'}|^{2}\right]\right) \\
&\overset{(d)}{=} \sum_{i=1}^{2K}\frac{p_{i'}}{(N-2K-1)\,\hat{\sigma}_i^{2}} + \hat{\eta}\sum_{i=1}^{2K} p_i\sigma_{\xi_i}^{2}
\end{aligned} \tag{2.44}$$

where $\hat{\eta} = \sum_{j=1}^{2K}\frac{1}{(N-2K)(N-2K-3)\hat{\sigma}_j^{2}\hat{\sigma}_{j'}^{2}}$, step (a) is based on the fact that the estimation error matrix $\boldsymbol{\Xi}$ is independent of $\hat{\mathbf{G}}$, and step (b) is obtained by an intuitive property of $\hat{\tilde{\mathbf{G}}}^{H}\hat{\mathbf{G}} = \hat{\mathbf{G}}^{H}\hat{\tilde{\mathbf{G}}} = \mathbf{I}_{2K}$, the definition of $\mathbf{P}_T = \mathbf{T}\mathbf{P}\mathbf{P}^{H}\mathbf{T} = \mathrm{diag}\left\{p_2, p_1, \ldots p_{2K}, p_{2K-1}\right\}$ and $\mathrm{E}\left[\boldsymbol{\Xi}\mathbf{P}\mathbf{P}^{H}\boldsymbol{\Xi}^{H}\right] = \sum_{i=1}^{2K} p_i\sigma_{\xi_i}^{2}\mathbf{I}_N$, which is derived from the distribution of $\boldsymbol{\Xi}$. Here, $\hat{\boldsymbol{\Omega}} \triangleq \left(\hat{\mathbf{G}}^{H}\hat{\mathbf{G}}\right)^{-1}$ is defined with $\hat{\omega}_{i,j} = \left(\hat{\boldsymbol{\Omega}}\right)_{i,j}$ for $\forall i,j \in \{1,2,\cdots,2K\}$, following an inverse Wishart distribution of $\mathcal{W}_{2K}^{-1}(N+2K+1, \hat{\mathbf{D}}^{-1})$. Furthermore, step (c) is based on $\mathrm{tr}\left\{\mathbf{A}\mathbf{B}\right\} = \mathrm{tr}\left\{\mathbf{B}\mathbf{A}\right\}$. As to the detailed derivation of (d), we use the identity [30, 31]

$$\mathrm{E}\left[\mathbf{W}^{-1}\right] = \frac{\boldsymbol{\Sigma}^{-1}}{n-m-1} \tag{2.45}$$

where $\mathbf{W} \sim \mathcal{W}_m(n, \boldsymbol{\Sigma})$ is an $m \times m$ central complex Wishart matrix with n $(n > m)$ degrees of freedom and the distribution of $\mathbf{W}^{-1}$ is called an inverted Wishart distribution, following $\mathcal{W}_m^{-1}(n+m+1, \boldsymbol{\Sigma}^{-1})$. It can be easily concluded that $\hat{\boldsymbol{\Omega}} \triangleq \left(\hat{\mathbf{G}}^{H}\hat{\mathbf{G}}\right)^{-1} \sim \mathcal{W}_{2K}^{-1}(N+2K+1, \hat{\mathbf{D}}^{-1})$ with $\hat{\omega}_{i,j} = \left(\hat{\boldsymbol{\Omega}}\right)_{i,j}$ for $\forall i,j \in \{1,2,\cdots,2K\}$, hence

$$\mathrm{E}\left[\hat{\tilde{\mathbf{G}}}^{H}\hat{\tilde{\mathbf{G}}}\right] = \mathrm{E}\left[\left(\hat{\mathbf{G}}^{H}\hat{\mathbf{G}}\right)^{-1}\right] = \frac{\hat{\mathbf{D}}^{-1}}{N-2K-1} \tag{2.46}$$

where $N > 2K$. In this way, we have $\mathrm{E}\left[\hat{\mathbf{g}}_{\mathbf{k}}^{\mathbf{H}}\hat{\mathbf{g}}_{\mathbf{k}}\right] = \mathrm{E}\left[\hat{\omega}_{k,k}\right] = \frac{1}{(N-2K-1)\hat{\sigma}_k^2}$ for $k = 1, 2, \cdots, 2K$, and

$$\begin{aligned}
&\mathrm{E}\left[\hat{\omega}_{j,j}^{*}\hat{\omega}_{j',j'}\right] = \mathrm{cov}\left[\hat{\omega}_{j,j}^{*}\hat{\omega}_{j',j'}\right] + \mathrm{E}\left[\hat{\omega}_{j,j}^{*}\right]\mathrm{E}\left[\hat{\omega}_{j',j'}\right] \\
= &\frac{2}{(N-2K)(N-2K-1)^2(N-2K-3)\hat{\sigma}_j^2\hat{\sigma}_{j'}^2} + \frac{1}{(N-2K-1)^2\hat{\sigma}_j^2\hat{\sigma}_{j'}^2}
\end{aligned} \tag{2.47}$$

$$\mathrm{E}\left[|\hat{\omega}_{i,j}|^2\right] = \mathrm{var}\left[\hat{\omega}_{i,j}\right] + \mathrm{E}^2\left[\hat{\omega}_{i,j}\right] = \frac{2}{(N-2K-1)^2(N-2K-3)\hat{\sigma}_i^2\hat{\sigma}_j^2} \tag{2.48}$$

where $\mathrm{cov}\left[\hat{\omega}_{k,k}^{*}\hat{\omega}_{k',k'}\right]$ $(k, k' \in \{1, 2, \cdots, 2K\})$, $\mathrm{E}\left[\hat{\omega}_{i,j}\right]$ and $\mathrm{var}\left[\hat{\omega}_{i,j}\right]$ $(i, j \in \{1, 2, \cdots, 2K\})$ are calculated based on the properties of the inverse Wishart matrix $\hat{\mathbf{\Omega}}$ [30, 31].

Then, the calculation of $\mathrm{E}\left[\left\|\hat{\tilde{\mathbf{G}}}^{*}\mathbf{T}\hat{\tilde{\mathbf{G}}}^{H}\mathbf{n}_{\mathrm{r}}\right\|^2\right]$ in (2.13) can be elaborated as

$$\begin{aligned}
\mathrm{E}\left[\left\|\hat{\tilde{\mathbf{G}}}^{*}\mathbf{T}\hat{\tilde{\mathbf{G}}}^{H}\mathbf{n}_{\mathrm{r}}\right\|^2\right] &= \mathrm{tr}\left\{\sigma_n^2\mathrm{E}\left[\hat{\mathbf{\Omega}}^{*}\mathbf{T}\hat{\mathbf{\Omega}}\mathbf{T}\right]\right\} \\
= \sigma_n^2\sum_{j=1}^{2K}\left(\mathrm{E}\left[\hat{\omega}_{j,j}^{*}\hat{\omega}_{j',j'}\right] + \mathrm{E}\left[|\hat{\omega}_{i,j}|^2\right]\right) &= \hat{\eta}\sigma_n^2.
\end{aligned} \tag{2.49}$$

Hence, by substituting (2.44) and (2.49) into the step (a) in (2.13), the proof of (2.13) is completed.

2.9.3 Proof of Theorem 2.1

To derive the closed-form expression of the achievable rate in (2.17), we start from the expectation $\mathrm{E}\left[\mathbf{g}_{k'}^{T}\mathbf{F}\mathbf{g}_k\right]$ in the numerator based on MRC/MRT in (2.9), given by

$$\begin{aligned}
\mathrm{E}\left[\mathbf{g}_{k'}^{T}\mathbf{F}\mathbf{g}_k\right] &\overset{(a)}{=} \mathrm{E}\left[\hat{\mathbf{g}}_{k'}^{T}\mathbf{F}\hat{\mathbf{g}}_k\right] + \mathrm{E}\left[\xi_{k'}^{T}\mathbf{F}\xi_k\right] \overset{(b)}{=} \alpha_1\mathrm{E}\left[\hat{\mathbf{g}}_{k'}^{T}\sum_{i=1}^{K}\left(\hat{\mathbf{g}}_{2i-1}^{*}\hat{\mathbf{g}}_{2i}^{H} + \hat{\mathbf{g}}_{2i}^{*}\hat{\mathbf{g}}_{2i-1}^{H}\right)\hat{\mathbf{g}}_k\right] \\
&\overset{(c)}{=} \alpha_1\mathrm{E}\left[\hat{\mathbf{g}}_{k'}^{T}\hat{\mathbf{g}}_k^{*}\hat{\mathbf{g}}_{k'}^{H}\hat{\mathbf{g}}_k + \hat{\mathbf{g}}_{k'}^{T}\hat{\mathbf{g}}_{k'}^{*}\hat{\mathbf{g}}_k^{H}\hat{\mathbf{g}}_k\right] \overset{(d)}{=} \alpha_1 N(N+1)\hat{\phi}_{\lceil\frac{k'}{2}\rceil}
\end{aligned} \tag{2.50}$$

where step (a) is obtained by $\mathbf{g}_{k'} = \hat{\mathbf{g}}_{k'} - \xi_{k'}$ and the independence between $\hat{\mathbf{g}}_{k'}$ and $\xi_{k'}$, step (b) results from substituting $\mathbf{F}$ in (2.9), $\hat{\mathbf{G}}$ in (2.4) and $\mathbf{T}$ into the equation and the fact $\mathrm{E}\left[\xi_{k'}^{T}\mathbf{F}\xi_k\right] = 0$, (c) is obtained by formula expansion based on the fact that the expectation of $\mathrm{E}\left[\hat{\mathbf{g}}_{k'}^{T}\left(\hat{\mathbf{g}}_{2i-1}^{*}\hat{\mathbf{g}}_{2i}^{H} + \hat{\mathbf{g}}_{2i}^{*}\hat{\mathbf{g}}_{2i-1}^{H}\right)\hat{\mathbf{g}}_k\right] = 0$ for any $i \neq \lceil\frac{k}{2}\rceil$ and $\lceil\frac{k'}{2}\rceil$, and step (d) results from the property $\mathrm{tr}\{\mathbf{AB}\} = \mathrm{tr}\{\mathbf{BA}\}$ and the properties of Gaussian distributed vectors.

Then, the variance of $\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k$ in the denominator of (2.17) is

$$\begin{aligned}
&\mathrm{Var}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right] \overset{(a)}{=} \mathrm{E}\left[\left|\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right|^2\right] - \left|\mathrm{E}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right]\right|^2 \\
\overset{(b)}{=}&\mathrm{E}\left[\hat{\mathbf{g}}_{k'}^T\mathbf{F}\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\mathbf{F}^H\hat{\mathbf{g}}_{k'}^*\right] + \mathrm{E}\left[\hat{\mathbf{g}}_{k'}^T\mathbf{F}\xi_k\xi_k^H\mathbf{F}^H\hat{\mathbf{g}}_{k'}^*\right] + \mathrm{E}\left[\xi_{k'}^T\mathbf{F}\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\mathbf{F}^H\xi_{k'}^*\right] \\
&+ \mathrm{E}\left[\xi_{k'}^T\mathbf{F}\xi_k\xi_k^H\mathbf{F}^H\xi_{k'}^*\right] - \alpha_1^2N^2(N+1)^2\hat{\phi}^2_{\lceil\frac{k'}{2}\rceil}
\end{aligned} \tag{2.51}$$

where step (a) indicates the definition of the variance and (b) results from $\mathbf{g}_i = \hat{\mathbf{g}}_i - \xi_i$ and the independence between $\hat{\mathbf{g}}_i$ and ξ_i ($\forall i \in \{1, \cdots, 2K\}$). For the first term of step (b) in (2.51), we have

$$\begin{aligned}
&\mathrm{E}\left[\hat{\mathbf{g}}_{k'}^T\mathbf{F}\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\mathbf{F}^H\hat{\mathbf{g}}_{k'}^*\right] \overset{(a)}{=} \alpha_1^2\sum_{i=1}^{K}\mathrm{E}\left[\hat{\mathbf{g}}_{k'}^T\left(\hat{\mathbf{g}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H + \hat{\mathbf{g}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\left(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i}^T + \hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i-1}^T\right)\hat{\mathbf{g}}_{k'}^*\right] \\
&\overset{(b)}{=} 2N\left(N+1\right)^2\alpha_1^2\hat{\phi}^2_{\lceil\frac{k'}{2}\rceil} + 2N\left(N+1\right)\alpha_1^2\hat{\phi}_{\lceil\frac{k'}{2}\rceil}\hat{\Phi} + N^2\left(N+1\right)^2\alpha_1^2\hat{\phi}^2_{\lceil\frac{k'}{2}\rceil}
\end{aligned} \tag{2.52}$$

where step (b) is obtained simply by the property $\mathrm{tr}\{\mathbf{AB}\} = \mathrm{tr}\{\mathbf{BA}\}$, the properties of Gaussian distributed vectors, and [29, Lemma 2.9]. To elaborate in detail, for the items of (a) in (2.52) with $i = \lceil\frac{k'}{2}\rceil$, we have

$$\begin{aligned}
&\mathrm{E}\left[\hat{\mathbf{g}}_{k'}^T\left(\hat{\mathbf{g}}_k^*\hat{\mathbf{g}}_{k'}^H + \hat{\mathbf{g}}_{k'}^*\hat{\mathbf{g}}_k^H\right)\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\left(\hat{\mathbf{g}}_k\hat{\mathbf{g}}_{k'}^T + \hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_k^T\right)\hat{\mathbf{g}}_{k'}^*\right] \\
=&\ 2\mathrm{tr}\left\{\mathrm{E}\left[\hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_{k'}^H\hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_{k'}^H\right]\mathrm{E}\left[\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\right]\right\} + \mathrm{E}\left[\left|\hat{\mathbf{g}}_k^H\hat{\mathbf{g}}_{k'}\right|^4\right] + \mathrm{E}\left[\left\|\hat{\mathbf{g}}_k\right\|^4\right]\mathrm{E}\left[\left\|\hat{\mathbf{g}}_{k'}\right\|^4\right] \\
=&\ 2N\left(N+1\right)^2\hat{\phi}^2_{\lceil\frac{k'}{2}\rceil} + 2N\left(N+1\right)\hat{\phi}^2_{\lceil\frac{k'}{2}\rceil} + N^2\left(N+1\right)^2\hat{\phi}^2_{\lceil\frac{k'}{2}\rceil}
\end{aligned} \tag{2.53}$$

where the properties of $\mathrm{E}\left[\hat{\mathbf{g}}_i\hat{\mathbf{g}}_i^H\hat{\mathbf{g}}_i\hat{\mathbf{g}}_i^H\right] = (N+1)\sigma_i^4\mathbf{I}_N$ $(i = 1, 2, \cdots, 2K)$ resulting from the fact that vectors $\hat{\mathbf{g}}_i$ contains the i.i.d. $\mathcal{CN}(0, \sigma_i^2)$ elements, $\mathrm{E}\left[|\theta|^4\right] = 2\sigma_\theta^4$ for arbitrary complex value $\theta \sim \mathcal{CN}(0, \sigma_\theta^2)$, and [29, Lemma 2.9] are utilized, respectively. While for the items with $i \neq \lceil\frac{k'}{2}\rceil$, we have

$$\begin{aligned}
&\mathrm{E}\left[\hat{\mathbf{g}}_{k'}^T\left(\hat{\mathbf{g}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H + \hat{\mathbf{g}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\left(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i}^T + \hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i-1}^T\right)\hat{\mathbf{g}}_{k'}^*\right] = 2\mathrm{E}\left[\hat{\mathbf{g}}_{2i-1}^H\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\hat{\mathbf{g}}_{2i-1}\right] \\
&\times \mathrm{E}\left[\hat{\mathbf{g}}_{2i}^H\hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_{k'}^H\hat{\mathbf{g}}_{2i}\right] + 2\mathrm{tr}\left\{\mathrm{E}\left[\hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_{k'}^H\right]\mathrm{E}\left[\hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i}^H\right]\mathrm{E}\left[\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\right]\mathrm{E}\left[\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i-1}^H\right]\right\} \\
=&\ 2N\left(N+1\right)\hat{\phi}_i\hat{\phi}_{\lceil\frac{k'}{2}\rceil}
\end{aligned} \tag{2.54}$$

where $\mathrm{tr}\{\mathbf{AB}\} = \mathrm{tr}\{\mathbf{BA}\}$ and the properties of Gaussian distributed vectors are utilized, respectively. Subsequently, the left three terms of step (b) in (2.51) are

calculated and expressed as

$$\mathrm{E}\left[\hat{\mathbf{g}}_{k'}^T\mathbf{F}\xi_k\xi_k^H\mathbf{F}^H\hat{\mathbf{g}}_{k'}^*\right] = N\left(N+1\right)^2\alpha_1^2\sigma_{\xi_k}^2\hat{\sigma}_{k'}^2\hat{\phi}_{\lceil\frac{k'}{2}\rceil} + 2N\left(N+1\right)\alpha_1^2\sigma_{\xi_k}^2\hat{\sigma}_{k'}^2\hat{\Phi} \quad (2.55a)$$

$$\mathrm{E}\left[\xi_{k'}^T\mathbf{F}\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\mathbf{F}^H\xi_{k'}^*\right] = N\left(N+1\right)^2\alpha_1^2\sigma_{\xi_{k'}}^2\hat{\sigma}_k^2\hat{\phi}_{\lceil\frac{k'}{2}\rceil} + 2N\left(N+1\right)\alpha_1^2\sigma_{\xi_{k'}}^2\hat{\sigma}_k^2\hat{\Phi} \quad (2.55b)$$

$$\mathrm{E}\left[\xi_{k'}^T\mathbf{F}\xi_k\xi_k^H\mathbf{F}^H\xi_{k'}^*\right] = 2N\left(N+1\right)\alpha_1^2\sigma_{\xi_k}^2\sigma_{\xi_{k'}}^2\hat{\Phi}. \quad (2.55c)$$

Substituting (2.52) and (2.55) into (2.51) leads to

$$\mathrm{Var}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right] = N\left(N+1\right)\alpha_1^2\left[\left(N+1\right)\hat{\phi}_{\lceil\frac{k'}{2}\rceil}\left(\sigma_{k'}^2\hat{\sigma}_k^2+\sigma_k^2\hat{\sigma}_{k'}^2\right)+2\phi_{\lceil\frac{k'}{2}\rceil}\hat{\Phi}\right] \quad (2.56)$$

where we define $\Phi = \sum_{i=1}^{K}\phi_i$ with $\phi_i = \sigma_{2i-1}^2\sigma_{2i}^2$.

Similarly, we obtain

$$\mathrm{SI}_{k'} = p_{k'}4N\left(N+1\right)\alpha_1^2\sigma_{\xi_{k'}}^2\left[\left(N+1\right)\hat{\sigma}_k^2\hat{\sigma}_{k'}^4+\left(\sigma_{k'}^2+\hat{\sigma}_{k'}^2\right)\hat{\Phi}\right] \quad (2.57a)$$

$$\mathrm{IP}_{k'} = \sum_{i\neq k,k'}^{2K} p_i N\left(N+1\right)\alpha_1^2\left[\left(N+1\right)\left(\sigma_i^2\hat{\sigma}_{k'}^4\hat{\sigma}_k^2+\sigma_{k'}^2\hat{\sigma}_i^4\hat{\sigma}_{i'}^2\right)+2\sigma_i^2\sigma_{k'}^2\hat{\Phi}\right] \quad (2.57b)$$

$$\mathrm{NR}_{k'} = N\left(N+1\right)\alpha_1^2\sigma_n^2\left[\left(N+1\right)\hat{\sigma}_k^2\hat{\sigma}_{k'}^4+2\sigma_{k'}^2\hat{\Phi}\right],\ \mathrm{NU}_{k'} = \sigma_n^2. \quad (2.57c)$$

Substituting (2.50), (2.56) and (2.57) into (2.17), we have

$$\begin{aligned}&\frac{p_k\left|\mathrm{E}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right]\right|^2}{p_k\mathrm{Var}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right]+\mathrm{SI}_{k'}+\mathrm{IP}_{k'}+\mathrm{NR}_{k'}+\mathrm{NU}_{k'}}\\ \overset{(a)}{=}&\frac{p_kN\left(N+1\right)\hat{\sigma}_k^4\hat{\sigma}_{k'}^4}{\sum_{i=1}^{2K}p_i\left\{\varsigma_{k',i}+\frac{\sigma_n^2\left[2\hat{\Phi}\sigma_i^2\sigma_{i'}^2+(N+1)\hat{\sigma}_i^4\hat{\sigma}_{i'}^2\right]}{P_{\mathrm{R}}}\right\}+c_{k'}p_{k'}+\sigma_n^2\left[\left(N+1\right)\hat{\sigma}_k^2\hat{\sigma}_{k'}^4+2\sigma_{k'}^2\hat{\Phi}\right]+\frac{2\sigma_n^4\hat{\Phi}}{P_{\mathrm{R}}}}\end{aligned} \quad (2.58)$$

where $\varsigma_{k',i}\triangleq\left(N+1\right)\left(\sigma_i^2\hat{\sigma}_{k'}^4\hat{\sigma}_k^2+\sigma_{k'}^2\hat{\sigma}_i^4\hat{\sigma}_{i'}^2\right)+2\sigma_i^2\sigma_{k'}^2\hat{\Phi}$, $c_{k'}=2\left[\left(N+1\right)\left(\sigma_{k'}^2-2\hat{\sigma}_{k'}^2\right)\hat{\sigma}_k^2\hat{\sigma}_{k'}^4\right.$ $\left.+\left(\sigma_{k'}^4-2\hat{\sigma}_{k'}^4\right)\hat{\Phi}\right]$, and step (a) is obtained by substituting (2.10) into (2.58). In this way, (2.19) is obtained and thus Theorem 2.1 is demonstrated.

2.9.4 Proof of Theorem 2.2

In this appendix, we focus on the proof of the closed-form expression in (2.17) with imperfect CSI based ZFR/ZFT processing. First, we start from the expecta-

tion $\mathrm{E}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right]$ in the numerator, given by

$$\begin{aligned}\mathrm{E}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right] &= \alpha_2\mathrm{E}\left[\left(\hat{\mathbf{g}}_{k'}-\xi_{k'}\right)^T\hat{\bar{\mathbf{G}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\left(\hat{\mathbf{g}}_k-\xi_k\right)\right]\\ &\overset{(a)}{=}\alpha_2\mathrm{E}\left[\hat{\mathbf{g}}_{k'}^T\hat{\bar{\mathbf{G}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\hat{\mathbf{g}}_k\right]+\alpha_2\mathrm{E}\left[\xi_{k'}^T\hat{\bar{\mathbf{G}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\xi_k\right]\overset{(b)}{=}\alpha_2\end{aligned} \tag{2.59}$$

where step (a) is based on the fact that the estimation error matrix $\mathbf{\Xi}$ is independent of $\hat{\mathbf{G}}$, i.e., ξ_i is independent of $\hat{\mathbf{g}}_j$ for $\forall i,j\in\{1,2,\cdots,2K\}$, and step (b) is obtained by $\hat{\mathbf{g}}_{i'}^T\mathbf{F}\hat{\mathbf{g}}_j=\alpha_2\delta_{ij}$ on account of $\hat{\mathbf{G}}^T\mathbf{F}\hat{\mathbf{G}}=\alpha_2\mathbf{I}_{2K}$.

Then, based on the imperfect CSI based ZFR/ZFT processing in (2.13), the variance of $\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k$ in the denominator of (2.17) is

$$\begin{aligned}&\mathrm{Var}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right]=\mathrm{E}\left[\left|\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right|^2\right]-\left|\mathrm{E}\left[\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_k\right]\right|^2=\mathrm{E}\left[\hat{\mathbf{g}}_{k'}^T\mathbf{F}\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\mathbf{F}^H\hat{\mathbf{g}}_{k'}^*\right]-{\alpha_2}^2\\ &\quad+\mathrm{E}\left[\hat{\mathbf{g}}_{k'}^T\mathbf{F}\xi_k\xi_k^H\mathbf{F}^H\hat{\mathbf{g}}_{k'}^*\right]+\mathrm{E}\left[\xi_{k'}^T\mathbf{F}\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\mathbf{F}^H\xi_{k'}^*\right]+\mathrm{E}\left[\xi_{k'}^T\mathbf{F}\xi_k\xi_k^H\mathbf{F}^H\xi_{k'}^*\right]\\ &\overset{(a)}{=}\sigma_{\xi_k}^2\mathrm{E}\left[\hat{\mathbf{g}}_{k'}^T\mathbf{F}\mathbf{F}^H\hat{\mathbf{g}}_{k'}^*\right]+\sigma_{\xi_{k'}}^2\mathrm{E}\left[\hat{\mathbf{g}}_k^H\mathbf{F}^H\mathbf{F}\hat{\mathbf{g}}_k\right]+\sigma_{\xi_k}^2\sigma_{\xi_{k'}}^2\mathrm{tr}\left\{\mathrm{E}\left[\mathbf{F}\mathbf{F}^H\right]\right\}\\ &\overset{(b)}{=}{\alpha_2}^2\sigma_{\xi_k}^2\mathrm{E}\left[\mathbf{e}_{k'}^T\mathbf{T}\hat{\mathbf{\Omega}}\mathbf{T}\mathbf{e}_{k'}\right]+{\alpha_2}^2\sigma_{\xi_{k'}}^2\mathrm{E}\left[\mathbf{e}_k^T\mathbf{T}\hat{\mathbf{\Omega}}^*\mathbf{T}\mathbf{e}_k\right]+{\alpha_2}^2\sigma_{\xi_k}^2\sigma_{\xi_{k'}}^2\mathrm{tr}\left\{\mathrm{E}\left[\hat{\mathbf{\Omega}}^*\mathbf{T}\hat{\mathbf{\Omega}}\mathbf{T}\right]\right\}\\ &\overset{(c)}{=}{\alpha_2}^2\sigma_{\xi_k}^2\mathrm{E}\left[\hat{\omega}_{k,k}\right]+{\alpha_2}^2\sigma_{\xi_{k'}}^2\mathrm{E}\left[\hat{\omega}_{k',k'}^*\right]+{\alpha_2}^2\sigma_{\xi_k}^2\sigma_{\xi_{k'}}^2\sum_{j=1}^{2K}\left(\mathrm{E}\left[\hat{\omega}_{j,j}^*\hat{\omega}_{j',j'}\right]+\mathrm{E}\left[\left|\hat{\omega}_{j,j'}\right|^2\right]\right)\\ &\overset{(d)}{=}{\alpha_2}^2\theta_{k',k}\end{aligned} \tag{2.60}$$

where $\theta_{i,j}=\frac{\sigma_{\xi_j}^2}{(N-2K-1)\hat{\sigma}_{i'}^2}+\frac{\sigma_{\xi_i}^2}{(N-2K-1)\hat{\sigma}_{j'}^2}+\sigma_{\xi_i}^2\sigma_{\xi_j}^2\hat{\eta}$ with $i,j\in\{1,2,\cdots,2K\}$, step (a) results from the property $\mathrm{tr}\{\mathbf{AB}\}=\mathrm{tr}\{\mathbf{BA}\}$, $\hat{\mathbf{g}}_{k'}^T\mathbf{F}\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\mathbf{F}^H\hat{\mathbf{g}}_{k'}^*={\alpha_2}^2\delta_{k,k}\delta_{k',k'}$ and $\mathrm{E}\left[\xi_i\xi_i^H\right]=\sigma_{\xi_i}^2\mathbf{I}_N$ for $\forall i\in\{1,2,\cdots,2K\}$, step (b) is obtained by the definition of $\hat{\mathbf{\Omega}}\triangleq\left(\hat{\mathbf{G}}^H\hat{\mathbf{G}}\right)^{-1}$ with $\hat{\omega}_{i,j}=\left(\hat{\mathbf{\Omega}}\right)_{i,j}$ for $\forall i,j\in\{1,2,\cdots,2K\}$ and the fact of $\mathbf{g}_{k'}^T\mathbf{G}^*\left(\mathbf{G}^T\mathbf{G}^*\right)^{-1}=\mathbf{e}_{k'}^T$ and $\left(\mathbf{G}^T\mathbf{G}^*\right)^{-1}\mathbf{G}^T\mathbf{g}_{k'}^*=\mathbf{e}_{k'}$, (c) is just an intuitive transformation, and (d) is derived according to the properties of the inverse Wishart matrix in (2.47) and (2.48).

Considering that no SIC is performed as $\hat{\mathbf{g}}_{k'}^T\mathbf{F}\hat{\mathbf{g}}_{k'}=0$, the self-interference term can be rewritten as

$$\mathrm{SI}_{k'}=p_{k'}\mathrm{E}\left[\left|\mathbf{g}_{k'}^T\mathbf{F}\mathbf{g}_{k'}\right|^2\right]=p_{k'}\alpha_2^2\left(\theta_{k',k'}+\sigma_{\xi_{k'}}^4\hat{\eta}\right). \tag{2.61}$$

Similarly, we obtain

$$\mathrm{IP}_{k'} = \sum_{i \neq k,k'}^{2K} p_i {\alpha_2}^2 \theta_{k',i}, \ \mathrm{NU}_{k'} = \sigma_n^2 \tag{2.62a}$$

$$\mathrm{NR}_{k'} = \frac{{\alpha_2}^2 \sigma_n^2}{(N-2K-1)\,\hat{\sigma}_k^2} + {\alpha_2}^2 \sigma_n^2 \sigma_{\xi_{k'}}^2 \hat{\eta}. \tag{2.62b}$$

Substituting (2.59), (2.60), (2.61) and (2.62) into (2.17), we have

$$\begin{aligned} &\frac{p_k \left| \mathrm{E}\left[\mathbf{g}_{k'}^T \mathbf{F} \mathbf{g}_k\right] \right|^2}{p_k \mathrm{Var}\left[\mathbf{g}_{k'}^T \mathbf{F} \mathbf{g}_k\right] + \mathrm{SI}_{k'} + \mathrm{IP}_{k'} + \mathrm{NR}_{k'} + \mathrm{NU}_{k'}} \\ &\overset{(a)}{=} \frac{p_k}{\sum_{i=1}^{2K} p_i \left[\theta_{k',i} + \frac{\sigma_n^2 \left(\frac{1}{(N-2K-1)\hat{\sigma}_{i'}^2} + \sigma_{\xi_i}^2 \hat{\eta} \right)}{P_{\mathrm{R}}} \right] + p_{k'} \sigma_{\xi_{k'}}^4 \hat{\eta} + \frac{\sigma_n^2}{(N-2K-1)\hat{\sigma}_k^2} + \sigma_n^2 \sigma_{\xi_{k'}}^2 \hat{\eta} + \frac{\sigma_n^4 \hat{\eta}}{P_{\mathrm{R}}}} \end{aligned} \tag{2.63}$$

where (a) results form substituting (2.13) into (2.63). Thus (2.20) is obtained and Theorem 2.2 is demonstrated.

Chapter 3

MSE-based Precoding for MIMO Downlinks in Heterogeneous Networks

In the above chapter, we have investigated massive multiple-input multiple-output (MIMO) combined with cooperative relaying for peer to peer transmissions utilizing linear processing. Considering the increasing data traffic demand in cellular networks today, heterogeneous networks (HetNets) has been investigated by the 3rd Generation Partnership Project (3GPP) for further improvements in network capacity. Due to the large number of potential interfering nodes in the network, properly mitigating both the inter-cell and intra-cell multiuser interference is a crucial issue facing HetNet. In this chapter, the mean square error based precoding design to be employed by the macro base station and the SC nodes will be studied for MIMO downlinks.

3.1 Introduction

It is widely acknowledged that further improvements in network capacity are only possible by increasing the node deployment density [32, 33]. On the other hand, deploying more macro tiers in already dense networks may be prohibitively expensive and result in significantly reduced cell splitting gains due to severe inter-cell interference [34]. Heterogeneous networks (HetNets) that embed a large number of low-power nodes into an existing macro network with the aim of offloading traffic from the macro cell to small cells has emerged as a viable and cost-effective way to increase network capacity [4, 32–34].

In a typical HetNet consisting of a macro cell (MC) and several small cells (SCs),

the MC serves its user equipments (UEs) in a large region by a high-power base station (BS), while each SC serves its UEs in its own coverage region by a low-power SC node if there is no cooperative transmission between the BSs and SCs. Due to the large number of potential interfering nodes in the network, properly mitigating both the inter-cell and intra-cell multiuser interference is a crucial issue facing HetNet. Interference control (IC) for the interference networks recently has been intensively studied and applied in HetNet [23, 35–38], and the coordinated multi-point (CoMP) transmission is demonstrated to be an effective approach in [35], including joint processing (JP) and coordinated beamforming (CB). When the backhaul among the coordinated tiers is able to share both user data and channel state information (CSI), the CoMP-JP transmission is shown to provide high spectral efficiency [23, 38]. However, JP also introduces limitations for practical implementation due to its needs for high signaling overhead. On the other hand, with the BSs and SCs cooperated in the beamformer or precoder level, CB strategies only require the share of CSI in order to mitigate the cross-tier interference between the macro cell and co-channel deployed SCs. Reference [36] has implemented the cross-tier IC with CB based on a prioritized user selection scheme. Later, a joint selection based IC is presented to achieve more balanced performances between the macro cell UEs and the SC UEs [37]. Nevertheless, these schemes with closed-form expressions are only available in certain cases, such as a two-user MIMO interference channel.

3.2 Related Work

In practical systems, the design of specific interference control schemes is subject to various criteria and constraints. Typically, interference control is formulated as problems that optimize certain system utility functions, which are directly associated with the UE rates or mean square error. Since the signal-to-noise ratio (SNR) is not so high in the practical wireless systems, especially at the cell edge, imperative performance improvement in the low and intermediate SNR region becomes a motivation in the IC scheme design. In [39], new MSE-based transceiver schemes are designed through efficient iterative algorithms for the peer to peer multiple-input multiple-output (MIMO) interference channel. In addition, source and relay precoding designs based on the mean square error (MSE) criterion in MIMO two-way relay systems are investigated in [40, 41]. Unfortunately, due to their differences in network architecture, they may not be employed directly into the HetNet systems, where there are hierarchical nodes including BS and SCs and each of them can transmit to multiple users.

3.3 Contributions

To the best knowledge of the authors, there are no MSE-based precoding schemes for HetNet in the literature. In this chapter, we develop three new MSE-based precoding schemes for MIMO downlinks in HetNet systems consisting of a macro tier overlaid with a second tier of SCs. Collectively, the proposed precoding schemes form a design toolbox that is expected to cover a wide spectrum of system needs ranging from superior precoding performance for systems with sufficient computing power to non-iterative precoder for systems without the need to exchange CSI among cells. First, the design of transmit precoding matrices and vectors is tackled by jointly minimizing a sum-MSE of all users subject to individual transmit power constraints at each cell. Based on this formulation, two alternating optimization algorithms named relaxed-constraints based alternating optimization (RAO) and unconstrained alternating optimization with normalization (UAON) are presented, where the RAO relaxes the non-convex constraints involved to convex ones first and then employs an alternating optimization technique to produce the solution, while the UAON is performed by embedding the constraints into the optimization process via a normalization step. Motivated by the techniques aimed at multi-cell time division duplex (TDD) systems [42, 43], next we develop a low complexity precoding scheme for HetNet where the precoder in each cell is designed separately without the need to exchange user data or CSI over the backhaul. By employing block diagonalization (BD) techniques at the node side [44], we derive a two-level precoder by a non-iterative algorithm where different interference thresholds are utilized to control the relative weights associated with the interferences for performance enhancement. Moreover, robust precoding schemes are presented correspondingly with imperfect CSI known at each node. Finally, we present results from numerical experiments for the proposed precoding strategies under different system configurations as well as a comparison study on performance in terms of MSE and bit error rate (BER).

The rest of the chapter is organized as follows. The system model for the MIMO downlinks in HetNet systems is described in Section 3.4. In Section 3.5, a new sum-MSE based precoding scheme for HetNet is proposed and two implementation algorithms are elaborated. In Section 3.6, a separate MSE based precoding algorithm is developed for the BS and SCs, respectively, and two-level precoders are derived. Then, robust precoders are designed based on the estimated channel knowledge in Section 3.7. Simulation results for several different system configurations are presented in Section 3.8 to demonstrate the performance of the proposed precoding techniques. Finally, we draw our conclusions in Section 3.9.

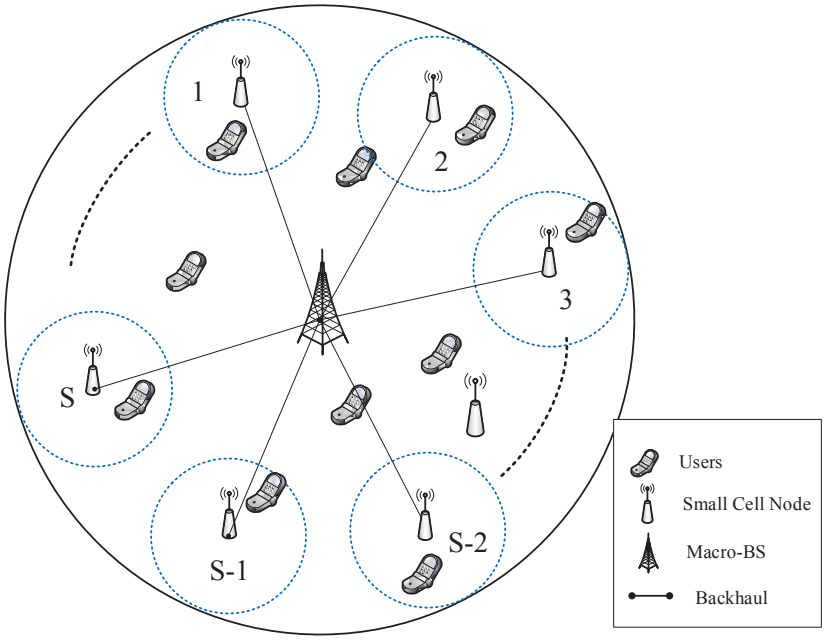

Figure 3.1: System model for HetNet with SCs deployment.

3.4 System Model

We consider a two-tier network architecture with one cell consisting of one macro BS, which is overlaid with a dense tier of S uniformly distributed SCs as shown in Fig. 3.1. Assume that the BS and SCs are respectively equipped with N_{BS} and N_{SC} antennas, while each user is dropped uniformly in the cell area and processes N_{UE} antennas. Based on the maximum reference signal received power (RSRP) [32], the users served by the macro BS are assigned to a macro UE (MUE) set, and those served by the SCs are assigned to a small cell UE (SUE) set. Suppose the macro BS serves K MUEs with $K \leq N_{\mathrm{BS}}$ while s-th SC ($s \in \Omega = \{1, 2, \ldots, S\}$) serves $L_s \leq N_{\mathrm{SC}}$ SUEs, thus the MUE and s-th SUE sets can be denoted by $I = \{1, 2, \ldots, K\}$ and $J_s = \{1, 2, \ldots, L_s\}$, respectively.

If the BS and SCs apply linear precoding to serve their UEs during the downlink transmissions, then the received signals at the i-th ($i \in I$) MUE and j-th ($j \in J_s$) SUE in the s-th SC are given by

$$\begin{aligned}\mathbf{y}_{\mathrm{BS}}^{(i)} &= \sum_{k=1}^{K} \sqrt{P_{\mathrm{BS}}} \left(\mathbf{H}_{\mathrm{B-M}}^{(i)}\right)^{H} \mathbf{W}_{\mathrm{BS}}^{(k)} \mathbf{x}_{\mathrm{BS}}^{(k)} + \sum_{s=1}^{S} \sum_{l=1}^{L_s} \sqrt{P_{\mathrm{SC}}} \left(\mathbf{H}_{\mathrm{S-M}}^{(s,i)}\right)^{H} \mathbf{W}_{\mathrm{SC}}^{(s,l)} \mathbf{x}_{\mathrm{SC}}^{(s,l)} + \mathbf{n}_{\mathrm{BS}}^{(i)} \\ &= \left(\mathbf{G}_{\mathrm{B-M}}^{(i)}\right)^{H} \mathbf{W}_{\mathrm{BS}} \mathbf{x}_{\mathrm{BS}} + \sum_{s=1}^{S} \left(\mathbf{G}_{\mathrm{S-M}}^{(s,i)}\right)^{H} \mathbf{W}_{\mathrm{SC}}^{(s)} \mathbf{x}_{\mathrm{SC}}^{(s)} + \mathbf{n}_{\mathrm{BS}}^{(i)}\end{aligned} \tag{3.1}$$

$$\begin{aligned}\mathbf{y}_{\mathrm{SC}}^{(s,j)} &= \sum_{k=1}^{K}\sqrt{P_{\mathrm{BS}}}\left(\mathbf{H}_{\mathrm{B-S}}^{(s,j)}\right)^{H}\mathbf{W}_{\mathrm{BS}}^{(k)}\mathbf{x}_{\mathrm{BS}}^{(k)} + \sum_{t=1}^{S}\sum_{l=1}^{L_t}\sqrt{P_{\mathrm{SC}}}\left(\mathbf{H}_{\mathrm{S-S}}^{(t,s,j)}\right)^{H}\mathbf{W}_{\mathrm{SC}}^{(t,l)}\mathbf{x}_{\mathrm{SC}}^{(t,l)} + \mathbf{n}_{\mathrm{SC}}^{(s,j)} \\ &= \left(\mathbf{G}_{\mathrm{B-S}}^{(s,j)}\right)^{H}\mathbf{W}_{\mathrm{BS}}\mathbf{x}_{\mathrm{BS}} + \sum_{t=1}^{S}\left(\mathbf{G}_{\mathrm{S-S}}^{(t,s,j)}\right)^{H}\mathbf{W}_{\mathrm{SC}}^{(t)}\mathbf{x}_{\mathrm{SC}}^{(t)} + \mathbf{n}_{\mathrm{SC}}^{(s,j)}\end{aligned} \tag{3.2}$$

respectively, where P_{BS} and P_{SC} represent the average power at the macro BS and SCs; $\mathbf{H}_{\mathrm{B-M}}^{(i)}$ and $\mathbf{H}_{\mathrm{B-S}}^{(j)}$ denote the $N_{\mathrm{BS}} \times N_{\mathrm{UE}}$ channel vectors from the BS to the i-th MUE and j-th SUE, respectively; $\mathbf{H}_{\mathrm{S-M}}^{(s,i)}$ and $\mathbf{H}_{\mathrm{S-S}}^{(s,j)}$ denote the $N_{\mathrm{SC}} \times N_{\mathrm{UE}}$ channel vectors from s-th SC to the i-th MUE and j-th SUE, respectively; $\mathbf{x}_{\mathrm{BS}}^{(k)} \in \mathbb{C}^{N_{\mathrm{S}}\times 1}$ and $\mathbf{x}_{\mathrm{SC}}^{(s,j)} \in \mathbb{C}^{N_{\mathrm{S}}\times 1}$ are the complex-valued Gaussian N_{S} transmitted symbol streams from BS to its k-th MUE and from s-th SC to its own SUE; $\mathbf{W}_{\mathrm{BS}}^{(k)}$ and $\mathbf{W}_{\mathrm{SC}}^{(s,j)}$ are the $N_{\mathrm{BS}} \times N_{\mathrm{S}}$ and $N_{\mathrm{SC}} \times N_{\mathrm{S}}$ precoding matrices, respectively; and $\mathbf{n}_{\mathrm{BS}}^{(i)}$ and $\mathbf{n}_{\mathrm{SC}}^{(s,j)}$ are the additive white Gaussian noise vectors with each element of variance N_0. Besides, $\mathbf{W}_{\mathrm{BS}} = \left[\mathbf{W}_{\mathrm{BS}}^{(1)}, \mathbf{W}_{\mathrm{BS}}^{(2)}, \ldots, \mathbf{W}_{\mathrm{BS}}^{(K)}\right]$, $\mathbf{W}_{\mathrm{SC}}^{(t)} = \left[\mathbf{W}_{\mathrm{SC}}^{(t,1)}, \mathbf{W}_{\mathrm{SC}}^{(t,2)}, \ldots, \mathbf{W}_{\mathrm{SC}}^{(t,L_t)}\right]$, $\mathbf{x}_{\mathrm{BS}} = \left[\mathbf{x}_{\mathrm{BS}}^{(1)}; \mathbf{x}_{\mathrm{BS}}^{(2)}; \ldots; \mathbf{x}_{\mathrm{BS}}^{(K)}\right]$, $\mathbf{x}_{\mathrm{SC}}^{(t)} = \left[\mathbf{x}_{\mathrm{SC}}^{(t,1)}; \mathbf{x}_{\mathrm{SC}}^{(t,2)}; \ldots; \mathbf{x}_{\mathrm{SC}}^{(t,L_t)}\right]$, $\mathbf{G}_{\mathrm{B-M}}^{(i)} = \sqrt{P_{\mathrm{BS}}}\mathbf{H}_{\mathrm{B-M}}^{(i)}$, $\mathbf{G}_{\mathrm{B-S}}^{(j)} = \sqrt{P_{\mathrm{BS}}}\mathbf{H}_{\mathrm{B-S}}^{(j)}$, $\mathbf{G}_{\mathrm{S-M}}^{(s,i)} = \sqrt{P_{\mathrm{SC}}}\mathbf{H}_{\mathrm{S-M}}^{(s,i)}$ and $\mathbf{G}_{\mathrm{S-S}}^{(t,s,j)} = \sqrt{P_{\mathrm{SC}}}\mathbf{H}_{\mathrm{S-S}}^{(t,s,j)}$ are defined for analysis simplicity. Moreover, the propagation factor here is defined as the product of a fast fading factor and an amplitude factor that accounts for geometric attenuation and shadow fading. For example, $h_{\mathrm{B-M}}^{(m_1,n_1,i)}$ (the (m_1, n_1)-th element of $\mathbf{H}_{\mathrm{B-M}}^{(i)}$) and $h_{\mathrm{S-M}}^{(m_2,n_2,s,i)}$ (the (m_2, n_2)-th element of $\mathbf{H}_{\mathrm{S-M}}^{(s,i)}$) in (3.1) assume the form

$$h_{\mathrm{B-M}}^{(m_1,n_1,i)} = \sqrt{\beta_{\mathrm{B-M}}^{(i)}}\upsilon_{\mathrm{B-M}}^{(m_1,n_1,i)},\ h_{\mathrm{S-M}}^{(m_2,n_2,s,i)} = \sqrt{\beta_{\mathrm{S-M}}^{(s,i)}}\upsilon_{\mathrm{S-M}}^{(m_2,n_2,s,i)} \tag{3.3}$$

where $m_1 \in \{1, 2, \ldots, N_{\mathrm{BS}}\}$, $m_2 \in \{1, 2, \ldots, N_{\mathrm{SC}}\}$, $n_1,\ n_2 \in \{1, 2, \ldots, N_{\mathrm{UE}}\}$; $\upsilon_{\mathrm{B-M}}^{(m_1,n_1,i)} \sim \mathcal{CN}(0, 1)$ and $\upsilon_{\mathrm{S-M}}^{(m_2,n_2,s,i)} \sim \mathcal{CN}(0, 1)$ denote the fast fading coefficients; and $\beta_{\mathrm{B-M}}^{(i)}$ and $\beta_{\mathrm{S-M}}^{(s,i)}$ are the amplitude factors. Because the geometric and shadow fading change slowly over space, $\beta_{\mathrm{B-M}}^{(i)}$ and $\beta_{\mathrm{S-M}}^{(s,i)}$ are treated as constants with respect to the index of the base station antenna, and we can write

$$\beta_{\mathrm{B-M}}^{(i)} = \zeta_{\mathrm{BS}}\theta_{\mathrm{BS}}\left(d_{\mathrm{B-M}}^{(i)}\right),\ \beta_{\mathrm{S-M}}^{(s,i)} = \zeta_{\mathrm{SC}}\theta_{\mathrm{SC}}\left(d_{\mathrm{S-M}}^{(s,i)}\right) \tag{3.4}$$

where ζ_{BS} and ζ_{SC} denote the corresponding penetration loss that are independent over all the indices [45], and functions $\theta_{\mathrm{BS}}\left(d_{\mathrm{B-M}}^{(i)}\right)$ and $\theta_{\mathrm{SC}}\left(d_{\mathrm{S-M}}^{(s,i)}\right)$ represent the pathloss model at the BS and the SCs, respectively, where the arguments $d_{\mathrm{B-M}}^{(i)}$ and $d_{\mathrm{S-M}}^{(s,i)}$ are the distance between the BS and the i-th MUE and the distance between the s-th SC and the i-th MUE, respectively. Similar expressions for the propagation factors $\mathbf{H}_{\mathrm{B-S}}^{(s,j)}$ and $\mathbf{H}_{\mathrm{S-S}}^{(t,s,j)}$ in (3.2) can be obtained. We assume that time division

duplex is adopted with channel reciprocity satisfied, i.e., the propagation factor is the same for both forward and reverse links and block fading remains constant for a duration symbols. Hence, exact CSI for the downlinks can be obtained for both BS and SCs.

From (3.1), the signal received at MUEs can be expressed as

$$\mathbf{y}_{\mathrm{BS}} = \mathbf{G}_{\mathrm{B-M}}^{H}\mathbf{W}_{\mathrm{BS}}\mathbf{x}_{\mathrm{BS}} + \sum_{s=1}^{S}\left(\mathbf{G}_{\mathrm{S-M}}^{(s)}\right)^{H}\mathbf{W}_{\mathrm{SC}}^{(s)}\mathbf{x}_{\mathrm{SC}}^{(s)} + \mathbf{n}_{\mathrm{BS}} \tag{3.5}$$

where $\mathbf{G}_{\mathrm{B-M}} = \left[\mathbf{G}_{\mathrm{B-M}}^{(1)}, \mathbf{G}_{\mathrm{B-M}}^{(2)}, \ldots, \mathbf{G}_{\mathrm{B-M}}^{(K)}\right]$, $\mathbf{G}_{\mathrm{S-M}}^{(s)} = \left[\mathbf{G}_{\mathrm{S-M}}^{(s,1)}, \mathbf{G}_{\mathrm{S-M}}^{(s,2)}, \ldots, \mathbf{G}_{\mathrm{S-M}}^{(s,K)}\right]$, $\mathbf{n}_{\mathrm{BS}} = \left[\mathbf{n}_{\mathrm{BS}}^{(1)}; \mathbf{n}_{\mathrm{BS}}^{(2)}; \ldots; \mathbf{n}_{\mathrm{BS}}^{(K)}\right]$ and $\mathbf{y}_{\mathrm{BS}} = \left[\mathbf{y}_{\mathrm{BS}}^{(1)}; \mathbf{y}_{\mathrm{BS}}^{(2)}; \ldots; \mathbf{y}_{\mathrm{BS}}^{(K)}\right]$. Similarly, from (3.2) the signal received at SUEs of the s-th SC is

$$\mathbf{y}_{\mathrm{SC}}^{(s)} = \left(\mathbf{G}_{\mathrm{B-S}}^{(s)}\right)^{H}\mathbf{W}_{\mathrm{BS}}\mathbf{x}_{\mathrm{BS}} + \sum_{t=1}^{S}\left(\mathbf{G}_{\mathrm{S-S}}^{(t,s)}\right)^{H}\mathbf{W}_{\mathrm{SC}}^{(t)}\mathbf{x}_{\mathrm{SC}}^{(t)} + \mathbf{n}_{\mathrm{SC}}^{(s)} \tag{3.6}$$

where $\mathbf{G}_{\mathrm{B-S}}^{(s)} = \left[\mathbf{G}_{\mathrm{B-S}}^{(s,1)}, \mathbf{G}_{\mathrm{B-S}}^{(s,2)}, \ldots, \mathbf{G}_{\mathrm{B-S}}^{(s,L_s)}\right]$, $\mathbf{G}_{\mathrm{S-S}}^{(t,s)} = \left[\mathbf{G}_{\mathrm{S-S}}^{(t,s,1)}, \mathbf{G}_{\mathrm{S-S}}^{(t,s,2)}, \ldots, \mathbf{G}_{\mathrm{S-S}}^{(t,s,L_s)}\right]$, $\mathbf{n}_{\mathrm{SC}}^{(s)} = \left[\mathbf{n}_{\mathrm{SC}}^{(s,1)}; \mathbf{n}_{\mathrm{SC}}^{(s,2)}; \ldots; \mathbf{n}_{\mathrm{SC}}^{(s,L_s)}\right]$ and $\mathbf{y}_{\mathrm{SC}}^{(s)} = \left[\mathbf{y}_{\mathrm{SC}}^{(s,1)}; \mathbf{y}_{\mathrm{SC}}^{(s,2)}; \ldots; \mathbf{y}_{\mathrm{SC}}^{(s,L_s)}\right]$.

Assume that the linear receiver is applied at each user, then

$$\hat{\mathbf{x}}_{\mathrm{BS}}^{(i)} = \mathbf{R}_{\mathrm{BS}}^{(i)}\mathbf{y}_{\mathrm{BS}}^{(i)},\ \hat{\mathbf{x}}_{\mathrm{SC}}^{(s,j)} = \mathbf{R}_{\mathrm{SC}}^{(s,j)}\mathbf{y}_{\mathrm{SC}}^{(s,j)},\ s \in \Omega \tag{3.7}$$

where $\mathbf{R}_{\mathrm{BS}}^{(i)} \in \mathbb{C}^{N_{\mathrm{S}} \times N_{\mathrm{UE}}}$ and $\mathbf{R}_{\mathrm{SC}}^{(s,j)} \in \mathbb{C}^{N_{\mathrm{S}} \times N_{\mathrm{UE}}}$ are the receiving filter matrices of MUE i and SUE j in the s-th SC, respectively. For simplicity, in the rest of the paper (3.7) is rewritten as

$$\hat{\mathbf{x}}_{\mathrm{BS}} = \mathbf{R}_{\mathrm{BS}}\mathbf{y}_{\mathrm{BS}},\ \hat{\mathbf{x}}_{\mathrm{SC}}^{(s)} = \mathbf{R}_{\mathrm{SC}}^{(s)}\mathbf{y}_{\mathrm{SC}}^{(s)},\ s \in \Omega \tag{3.8}$$

where $\mathbf{R}_{\mathrm{BS}} = \mathrm{bd}\left\{\mathbf{R}_{\mathrm{BS}}^{(1)}, \ldots, \mathbf{R}_{\mathrm{BS}}^{(K)}\right\}$ and $\mathbf{R}_{\mathrm{SC}}^{(s)} = \mathrm{bd}\left\{\mathbf{R}_{\mathrm{SC}}^{(s,1)}, \ldots, \mathbf{R}_{\mathrm{SC}}^{(s,L_s)}\right\}$.

3.5 Sum-MSE Minimization Based Precoding in HetNet

In this section, the design of precoding matrices $\mathbf{W}_{\mathrm{BS}}$ and $\mathbf{W}_{\mathrm{SC}}^{(s)}$ ($s \in \Omega$) is addressed by minimizing the total MSE (we call it sum-MSE) where each squared error term involves its corresponding receiver matrix $\mathbf{R}_{\mathrm{BS}}$ or $\mathbf{R}_{\mathrm{SC}}^{(s)}$ that can be performed by the user. This minimization is carried out subject to average power constraints on $\mathbf{W}_{\mathrm{BS}}$ and $\mathbf{W}_{\mathrm{SC}}^{(s)}$ for $s \in \Omega$. Under these circumstances, the precoding design

problem can be cast as a constrained optimization problem

$$\min_{\substack{\mathbf{W}_{\mathrm{BS}},\,\mathbf{W}_{\mathrm{SC}}^{(t)}\\ \mathbf{R}_{\mathrm{BS}},\,\mathbf{R}_{\mathrm{SC}}^{(t)},\,t\in\Omega}} \quad \mathrm{E}\left[\left\|\hat{\mathbf{x}}_{\mathrm{BS}}-\mathbf{x}_{\mathrm{BS}}\right\|^2+\sum_{s=1}^{S}\left\|\hat{\mathbf{x}}_{\mathrm{SC}}^{(s)}-\mathbf{x}_{\mathrm{SC}}^{(s)}\right\|^2\right] \tag{3.9a}$$

$$\text{subject to} \quad \mathrm{tr}\left\{\mathbf{W}_{\mathrm{BS}}^H\mathbf{W}_{\mathrm{BS}}\right\}=1 \tag{3.9b}$$

$$\mathrm{tr}\left\{\left(\mathbf{W}_{\mathrm{SC}}^{(t)}\right)^H\mathbf{W}_{\mathrm{SC}}^{(t)}\right\}=1,\ \text{for } t\in\Omega \tag{3.9c}$$

$$\mathbf{R}_{\mathrm{BS}}=\mathrm{bd}\left\{\mathbf{R}_{\mathrm{BS}}^{(1)},\ldots,\mathbf{R}_{\mathrm{BS}}^{(K)}\right\} \tag{3.9d}$$

$$\mathbf{R}_{\mathrm{SC}}^{(t)}=\mathrm{bd}\left\{\mathbf{R}_{\mathrm{SC}}^{(t,1)},\ldots,\mathbf{R}_{\mathrm{SC}}^{(t,L_t)}\right\},\ \text{for } t\in\Omega. \tag{3.9e}$$

Let $\mathbf{W}_{\mathrm{SC}}=\left[\mathbf{W}_{\mathrm{SC}}^{(1)};\ \mathbf{W}_{\mathrm{SC}}^{(2)};\ \ldots;\ \mathbf{W}_{\mathrm{SC}}^{(S)}\right]$ and $\mathbf{R}_{\mathrm{SC}}=\left[\mathbf{R}_{\mathrm{SC}}^{(1)},\mathbf{R}_{\mathrm{SC}}^{(2)},\ldots,\mathbf{R}_{\mathrm{SC}}^{(S)}\right]$, and note that the transmission symbols satisfy $\mathrm{E}\{\mathbf{x}\}=0$, $\mathrm{E}\left\{\mathbf{x}\mathbf{x}^H\right\}=\mathbf{I}$, and $\|\mathbf{x}\|^2=\mathrm{tr}\{\mathbf{x}\mathbf{x}^H\}$, the objective function in (3.9a) can be rewritten to make its dependence on $\mathbf{W}_{\mathrm{BS}}$ and $\mathbf{W}_{\mathrm{SC}}^{(s)}$ explicit as

$$f\left(\mathbf{W}_{\mathrm{BS}},\mathbf{W}_{\mathrm{SC}},\mathbf{R}_{\mathrm{BS}},\mathbf{R}_{\mathrm{SC}}\right)=\mathrm{MSE}_{\mathrm{BS}}+\mathrm{MSE}_{\mathrm{SC}} \tag{3.10}$$

where

$$\mathrm{MSE}_{\mathrm{BS}}\triangleq\mathrm{E}\left[\left\|\hat{\mathbf{x}}_{\mathrm{BS}}-\mathbf{x}_{\mathrm{BS}}\right\|^2\right]=\mathrm{tr}\left\{\mathbf{R}_{\mathrm{BS}}\left[\mathbf{G}_{\mathrm{B-M}}^H\mathbf{W}_{\mathrm{BS}}\mathbf{W}_{\mathrm{BS}}^H\mathbf{G}_{\mathrm{B-M}}+\sum_{s=1}^{S}\left(\mathbf{G}_{\mathrm{S-M}}^{(s)}\right)^H\right.\right.$$
$$\left.\left.\times\mathbf{W}_{\mathrm{SC}}^{(s)}\left(\mathbf{W}_{\mathrm{SC}}^{(s)}\right)^H\mathbf{G}_{\mathrm{S-M}}^{(s)}\right]\mathbf{R}_{\mathrm{BS}}^H-2\mathbf{R}_{\mathrm{BS}}\mathbf{G}_{\mathrm{B-M}}^H\mathbf{W}_{\mathrm{BS}}+\mathbf{I}_{KN_{\mathrm{S}}}+\sigma_0^2\mathbf{R}_{\mathrm{BS}}\mathbf{R}_{\mathrm{BS}}^H\right\} \tag{3.11}$$

and

$$\mathrm{MSE}_{\mathrm{SC}}\triangleq\mathrm{E}\left[\sum_{s=1}^{S}\left\|\hat{\mathbf{x}}_{\mathrm{SC}}^{(s)}-\mathbf{x}_{\mathrm{SC}}^{(s)}\right\|^2\right]=\sum_{s=1}^{S}\mathrm{tr}\left\{\mathbf{R}_{\mathrm{SC}}^{(s)}\left[\left(\mathbf{G}_{\mathrm{B-S}}^{(s)}\right)^H\mathbf{W}_{\mathrm{BS}}\mathbf{W}_{\mathrm{BS}}^H\mathbf{G}_{\mathrm{B-S}}^{(s)}\right.\right.$$
$$+\sum_{t=1}^{S}\left(\mathbf{G}_{\mathrm{S-S}}^{(t,s)}\right)^H\mathbf{W}_{\mathrm{SC}}^{(t)}\left(\mathbf{W}_{\mathrm{SC}}^{(t)}\right)^H\left.\mathbf{G}_{\mathrm{S-S}}^{(t,s)}\right]\left(\mathbf{R}_{\mathrm{SC}}^{(s)}\right)^H-2\mathbf{R}_{\mathrm{SC}}^{(s)}\left(\mathbf{G}_{\mathrm{S-S}}^{(s,s)}\right)^H\mathbf{W}_{\mathrm{SC}}^{(s)}$$
$$\left.+\mathbf{I}_{L_sN_{\mathrm{S}}}+\sigma_0^2\mathbf{R}_{\mathrm{SC}}^{(s)}\left(\mathbf{R}_{\mathrm{SC}}^{(s)}\right)^H\right\}. \tag{3.12}$$

From (3.11) and (3.12) it follows that the sum-MSE is convex w.r.t $\mathbf{W}_{\mathrm{BS}}$ and $\mathbf{W}_{\mathrm{SC}}^{(t)}$; and that it is also convex w.r.t. $\mathbf{R}_{\mathrm{BS}}$ and the matrices in $\mathbf{R}_{\mathrm{SC}}$. An essential techni-

cal difficulty in dealing with problem (3.9) is that both its objective function and the constraints on average power are nonconvex. In what follows, we propose an alternating convex optimization (ACO) technique which turns out to be well suited for the precoding design problem at hand. Specifically, a significant advantage of using ACO-based techniques is that all sub-problems involved are convex, and fast algorithms for their solutions and reliable software code for implementations are available [20, 21]. In what follows we present two alternating-optimization based techniques. The first technique partitions the design variables into two subsets such that the objective becomes convex with respect to each subset of variables, and this variable partitioning is done while the constraints on average power are relaxed to their convex counterparts. The second technique carries out unconstrained alternating optimization with respect to the above-mentioned two subsets of design variables alternatively, followed by a simple norm normalization step to satisfy the requirement on average power.

3.5.1 Relaxed-constraints based Alternating Optimization (RAO)

Here we consider a variant of problem (3.9) by a natural convex relaxation of the nonconvex constraints in (3.9b) and (3.9c), namely,

$$\begin{aligned}
\min_{\substack{\mathbf{W}_{\mathrm{BS}},\,\mathbf{W}_{\mathrm{SC}}^{(t)} \\ \mathbf{R}_{\mathrm{BS}},\,\mathbf{R}_{\mathrm{SC}}^{(t)},\,t\in\Omega}} \quad & f\left(\mathbf{W}_{\mathrm{BS}},\mathbf{W}_{\mathrm{SC}},\mathbf{R}_{\mathrm{BS}},\mathbf{R}_{\mathrm{SC}}\right) && \text{(3.13a)} \\
\text{subject to} \quad & \mathrm{tr}\left\{\mathbf{W}_{\mathrm{BS}}^{H}\mathbf{W}_{\mathrm{BS}}\right\} \le 1 && \text{(3.13b)} \\
& \mathrm{tr}\left\{\left(\mathbf{W}_{\mathrm{SC}}^{(t)}\right)^{H}\mathbf{W}_{\mathrm{SC}}^{(t)}\right\} \le 1, \ \text{for } t\in\Omega && \text{(3.13c)} \\
& (3.9d),\ (3.9e) && \text{(3.13d)}
\end{aligned}$$

As (3.9b) and (3.9c) impose conditions on the average power at the BS and SCs, its convex relaxation as seen in (3.13b) and (3.13c) are well justified as it limits the average power at the BS and SCs to be within the given values. As will become transparent shortly, the convex relaxation removes the only obstacle that would otherwise prevent from applying an ACO-based technique to the precoding problem.

To solve problem (3.13), we begin by partitioning the design variables into two sets, namely $\mathbf{X}_1 = \{\mathbf{W}_{\mathrm{BS}},\mathbf{W}_{\mathrm{SC}}\}$ and $\mathbf{X}_2 = \{\mathbf{R}_{\mathrm{BS}},\mathbf{R}_{\mathrm{SC}}\}$. Note that $f\left(\mathbf{W}_{\mathrm{BS}},\mathbf{W}_{\mathrm{SC}},\mathbf{R}_{\mathrm{BS}},\mathbf{R}_{\mathrm{SC}}\right)$ in (3.13a) is convex w.r.t. variable set $\mathbf{X}_1$ while variable set $\mathbf{X}_2$ is fixed, and that it is also convex w.r.t. $\mathbf{X}_2$ while $\mathbf{X}_1$ is fixed. There-

fore, it is natural to apply an ACO approach for the solution of (3.13), which is outlined as follows. With variables in $\mathbf{X}_1$ fixed, one minimizes convex objective function $f\left(\mathbf{W}_{\rm BS}, \mathbf{W}_{\rm SC}, \mathbf{R}_{\rm BS}, \mathbf{R}_{\rm SC}\right)$ w.r.t. variables $\{\mathbf{R}_{\rm BS}, \mathbf{R}_{\rm SC}\}$. Clearly this is an unconstrained convex problem because variables $\{\mathbf{R}_{\rm BS}, \mathbf{R}_{\rm SC}\}$ are not involved in (3.13b) and (3.13c) and constraints in (3.13d) can be removed by substituting it into the objective function. The solution of the above problem, denoted by $\mathbf{X}_2^* = \{\mathbf{R}_{\rm BS}^*, \mathbf{R}_{\rm SC}^*\}$, are then fixed and one minimizes the convex objective function $f\left(\mathbf{W}_{\rm BS}, \mathbf{W}_{\rm SC}, \mathbf{R}_{\rm BS}^*, \mathbf{R}_{\rm SC}^*\right)$ w.r.t. $\{\mathbf{W}_{\rm BS}, \mathbf{W}_{\rm SC}\}$ subject to constraints (3.13b) and (3.13c). Obviously this is a constrained convex problem that can be solved efficiently. Having obtained its solution $\{\mathbf{W}_{\rm BS}^*, \mathbf{W}_{\rm SC}^*\}$, the next round of ACO starts, and the procedure continues until a norm of the variations in both variable sets obtained from the two current consecutive rounds is less than a prescribed tolerance and the most current $\{\mathbf{W}_{\rm BS}^*, \mathbf{W}_{\rm SC}^*, \mathbf{R}_{\rm BS}^*, \mathbf{R}_{\rm SC}^*\}$ is taken as the solution of the problem. The technical details of solving the two convex sub-problems now follow.

With $\mathbf{X}_1$ fixed:

In this case, $\mathbf{W}_{\rm BS}$ and $\mathbf{W}_{\rm SC}$ are given and the optimization problem in (3.13) assumes the form

$$\min_{\mathbf{R}_{\rm BS}, \mathbf{R}_{\rm SC}} f_1\left(\mathbf{R}_{\rm BS}, \mathbf{R}_{\rm SC}\right) \tag{3.14a}$$

$$\text{subject to} \quad (3.9d), \ (3.9e) \tag{3.14b}$$

Substituting constraints (3.9d) and (3.9e) into eq. (3.10), it follows that

$$\begin{aligned} f_1\left(\mathbf{R}_{\rm BS}, \mathbf{R}_{\rm SC}\right) = \sum_{i=1}^{K} \mathrm{tr}\Bigg\{ & \mathbf{R}_{\rm BS}^{(i)} \mathbf{\Psi}_{\rm BS}^{(i)} \left(\mathbf{R}_{\rm BS}^{(i)}\right)^H - 2\mathbf{R}_{\rm BS}^{(i)} \left(\mathbf{G}_{\rm B-M}^{(i)}\right)^H \mathbf{W}_{\rm BS}^{(i)} + \mathbf{I}_{N_{\rm S}} \\ & + \sigma_0^2 \mathbf{R}_{\rm BS}^{(i)} \left(\mathbf{R}_{\rm BS}^{(i)}\right)^H \Bigg\} + \sum_{s=1}^{S} \sum_{j=1}^{l_s} \mathrm{tr}\Bigg\{ \mathbf{R}_{\rm SC}^{(s,j)} \mathbf{\Psi}_{\rm SC}^{(s,j)} \left(\mathbf{R}_{\rm SC}^{(s,j)}\right)^H \\ & - 2\mathbf{R}_{\rm SC}^{(s,j)} \left(\mathbf{G}_{\rm S-S}^{(s,s,j)}\right)^H \mathbf{W}_{\rm SC}^{(s,j)} + \mathbf{I}_{N_{\rm S}} + \sigma_0^2 \mathbf{R}_{\rm SC}^{(s,j)} \left(\mathbf{R}_{\rm SC}^{(s,j)}\right)^H \Bigg\} \end{aligned} \tag{3.15}$$

where

$$\mathbf{\Psi}_{\rm BS}^{(i)} = \left(\mathbf{G}_{\rm B-M}^{(i)}\right)^H \mathbf{W}_{\rm BS} \mathbf{W}_{\rm BS}^H \mathbf{G}_{\rm B-M}^{(i)} + \sum_{s=1}^{S} \left(\mathbf{G}_{\rm S-M}^{(s,i)}\right)^H \mathbf{W}_{\rm SC}^{(s)} \left(\mathbf{W}_{\rm SC}^{(s)}\right)^H \mathbf{G}_{\rm S-M}^{(s,i)} \tag{3.16a}$$

$$\mathbf{\Psi}_{\rm SC}^{(s,j)} = \left(\mathbf{G}_{\rm B-S}^{(s,j)}\right)^H \mathbf{W}_{\rm BS} \mathbf{W}_{\rm BS}^H \mathbf{G}_{\rm B-S}^{(s,j)} + \sum_{t=1}^{S} \left(\mathbf{G}_{\rm S-S}^{(t,s,j)}\right)^H \mathbf{W}_{\rm SC}^{(t)} \left(\mathbf{W}_{\rm SC}^{(t)}\right)^H \mathbf{G}_{\rm S-S}^{(t,s,j)} \tag{3.16b}$$

Hence, the global minimizer $\mathbf{R}_{\mathrm{BS}}{}^{*}$ and $\mathbf{R}_{\mathrm{SC}}^{*}$ can be found by solving

$$\frac{\partial f_1\left(\mathbf{R}_{\mathrm{BS}}, \mathbf{R}_{\mathrm{SC}}\right)}{\partial \mathbf{R}_{\mathrm{BS}}^{(i)}} = 0, \ \frac{\partial f_1\left(\mathbf{R}_{\mathrm{BS}}, \mathbf{R}_{\mathrm{SC}}\right)}{\partial \mathbf{R}_{\mathrm{SC}}^{(s,j)}} = 0, \ \forall i \in I, j \in J_s, s \in \Omega \tag{3.17}$$

which gives

$$\mathbf{R}_{\mathrm{BS}}^{(i)*} = \left(\mathbf{W}_{\mathrm{BS}}^{(i)}\right)^H \mathbf{G}_{\mathrm{B-M}}^{(i)} \left(\mathbf{\Psi}_{\mathrm{BS}}^{(i)} + \sigma_0^2 \mathbf{I}_{N_{\mathrm{UE}}}\right)^{-1}, \ i \in I \tag{3.18a}$$

$$\mathbf{R}_{\mathrm{SC}}^{(s,j)*} = \left(\mathbf{W}_{\mathrm{SC}}^{(s,j)}\right)^H \mathbf{G}_{\mathrm{S-S}}^{(s,s,j)} \left(\mathbf{\Psi}_{\mathrm{SC}}^{(s,j)} + \sigma_0^2 \mathbf{I}_{N_{\mathrm{UE}}}\right)^{-1}, \ j \in J_s, s \in \Omega. \tag{3.18b}$$

As we can see, the optimal linear receivers $\mathbf{R}_{\mathrm{BS}}^{*}$ and $\mathbf{R}_{\mathrm{SC}}^{(s,j)*}$ $(s \in \Omega)$ depend on the optimal transmit precoding matrices $\mathbf{W}_{\mathrm{BS}}$ and $\mathbf{W}_{\mathrm{SC}}^{(s,j)}$. In this way, the optimal solution $\mathbf{R}_{\mathrm{BS}}^{*}$ and $\mathbf{R}_{\mathrm{SC}}^{(s)*}$ $(s \in \Omega)$ for problem (3.14) can be easily obtained by (3.18) based on the assumption of $\mathbf{X}_1$ being fixed.

With $\mathbf{X}_2$ fixed:

In this case, $\mathbf{R}_{\mathrm{BS}}$ and $\mathbf{R}_{\mathrm{SC}}$ are fixed, and the optimization problem in (3.9) assumes the form

$$\min_{\mathbf{W}_{\mathrm{BS}}, \mathbf{W}_{\mathrm{SC}}} f_2\left(\mathbf{W}_{\mathrm{BS}}, \mathbf{W}_{\mathrm{SC}}\right) \tag{3.19a}$$

$$\text{subject to} \ \ (3.13b), \ (3.13c). \tag{3.19b}$$

With λ_0 and λ_s $(s \in \Omega)$ as the Lagrange multipliers associated with the power constraints, the Lagrangian of problem (3.19) is given by [20]

$$\begin{aligned} L\left(\mathbf{W}_{\mathrm{BS}}, \mathbf{W}_{\mathrm{SC}}, \lambda\right) =& f_2\left(\mathbf{W}_{\mathrm{BS}}, \mathbf{W}_{\mathrm{SC}}\right) + \lambda_0 \left[\mathrm{tr}\left\{\mathbf{W}_{\mathrm{BS}}^H \mathbf{W}_{\mathrm{BS}}\right\} - 1\right] \\ &+ \sum_{s=1}^{S} \lambda_s \left[\mathrm{tr}\left\{\left(\mathbf{W}_{\mathrm{SC}}^{(s)}\right)^H \mathbf{W}_{\mathrm{SC}}^{(s)}\right\} - 1\right] \end{aligned} \tag{3.20}$$

where for notation simplicity we have defined $\lambda = [\lambda_0, \lambda_1, \ldots, \lambda_S]^T$. It is clear that, given $\mathbf{R}_{\mathrm{BS}}$, $\mathbf{R}_{\mathrm{SC}}$, λ_0 and λ_s $(s \in \Omega)$, the Lagrangian in (3.20) is minimized if and only if

$$\frac{\partial L\left(\mathbf{W}_{\mathrm{BS}}, \mathbf{W}_{\mathrm{SC}}, \lambda\right)}{\partial \mathbf{W}_{\mathrm{BS}}} = 0, \ \frac{\partial L\left(\mathbf{W}_{\mathrm{BS}}, \mathbf{W}_{\mathrm{SC}}, \lambda\right)}{\partial \mathbf{W}_{\mathrm{SC}}^{(s)}} = 0, \ \forall s \in \Omega \tag{3.21}$$

i.e.,

$$\mathbf{W}_{\mathrm{BS}}^{*} = \left(\mathbf{\Phi}_{\mathrm{BS}} + \lambda_0 \mathbf{I}_{N_{\mathrm{BS}}}\right)^{-1} \mathbf{G}_{\mathrm{B-M}} \mathbf{R}_{\mathrm{BS}}^{H} \tag{3.22a}$$

$$\mathbf{W}_{\mathrm{SC}}^{(s)*} = \left(\mathbf{\Phi}_{\mathrm{SC}}^{(s)} + \lambda_s \mathbf{I}_{N_{\mathrm{SC}}}\right)^{-1} \mathbf{G}_{\mathrm{S-S}}^{(s,s)} \left(\mathbf{R}_{\mathrm{SC}}^{(s)}\right)^{H}, \ s \in \Omega. \tag{3.22b}$$

where

$$\mathbf{\Phi}_{\mathrm{BS}} = \mathbf{G}_{\mathrm{B-M}} \mathbf{R}_{\mathrm{BS}}^{H} \mathbf{R}_{\mathrm{BS}} \mathbf{G}_{\mathrm{B-M}}^{H} + \sum_{s=1}^{S} \mathbf{G}_{\mathrm{B-S}}^{(s)} \left(\mathbf{R}_{\mathrm{SC}}^{(s)}\right)^{H} \mathbf{R}_{\mathrm{SC}}^{(s)} \left(\mathbf{G}_{\mathrm{B-S}}^{(s)}\right)^{H} \tag{3.23a}$$

$$\mathbf{\Phi}_{\mathrm{SC}}^{(s)} = \mathbf{G}_{\mathrm{S-M}}^{(s)} \mathbf{R}_{\mathrm{BS}}^{H} \mathbf{R}_{\mathrm{BS}} \left(\mathbf{G}_{\mathrm{S-M}}^{(s)}\right)^{H} + \sum_{t=1}^{S} \mathbf{G}_{\mathrm{S-S}}^{(s,t)} \left(\mathbf{R}_{\mathrm{SC}}^{(t)}\right)^{H} \mathbf{R}_{\mathrm{SC}}^{(t)} \left(\mathbf{G}_{\mathrm{S-S}}^{(s,t)}\right)^{H}, \ s \in \Omega. \tag{3.23b}$$

To obtain non-negative multipliers λ_0 and λ_s ($s \in \Omega$) in the above equations, we substitute (3.22) into (3.20) and write $L(\lambda) = L(\mathbf{W}_{\mathrm{BS}}^{*}, \mathbf{W}_{\mathrm{SC}}^{*}, \lambda)$. From the complementarity equalities in the Karush-Kuhn-Tucker (KKT) conditions for (3.19), namely

$$\lambda_0 \left(\mathrm{tr}\left\{\mathbf{W}_{\mathrm{BS}}^{H} \mathbf{W}_{\mathrm{BS}}\right\} - 1\right) = 0 \tag{3.24a}$$

$$\lambda_s \left[\mathrm{tr}\left\{\left(\mathbf{W}_{\mathrm{SC}}^{(s)}\right)^{H} \mathbf{W}_{\mathrm{SC}}^{(s)}\right\} - 1\right] = 0, \ s \in \Omega \tag{3.24b}$$

we see that the optimal Lagrange multipliers are either positive such that the equality constraints in (3.9b) hold or zeros such that the constraints in (3.13b) hold strictly. Recalling that the equality constraints in (3.9b) are relaxed to the convex inequalities, we first assume that all the multipliers are greater than zero so that the equalities in constraints (3.13b) hold. This is the same as stating that taking partial derivative of L w.r.t. λ_0 and λ_s ($s \in \Omega$) yields zero values.

Given that $\mathbf{\Phi}_{\mathrm{BS}}$ can be factorized in the form $\mathbf{S}_{\mathrm{BS}}^{H} \mathbf{D}_{\mathrm{BS}} \mathbf{S}_{\mathrm{BS}}$ where $\mathbf{S}_{\mathrm{BS}}^{H} \mathbf{S}_{\mathrm{BS}} = \mathbf{I}_{N_{\mathrm{BS}}}$ and $\mathbf{D}_{\mathrm{BS}} = \mathrm{diag}\left\{d_{\mathrm{BS}}^{(1)}, d_{\mathrm{BS}}^{(2)}, \ldots, d_{\mathrm{BS}}^{(N_{\mathrm{BS}})}\right\}$, and that each $\mathbf{\Phi}_{\mathrm{SC}}^{(s)}$ can be expressed as $\left(\mathbf{S}_{\mathrm{SC}}^{(s)}\right)^{H} \mathbf{D}_{\mathrm{SC}}^{(s)} \mathbf{S}_{\mathrm{SC}}^{(s)}$, with $\left(\mathbf{S}_{\mathrm{SC}}^{(s)}\right)^{H} \mathbf{S}_{\mathrm{SC}}^{(s)} = \mathbf{I}_{N_{\mathrm{SC}}}$ and $\mathbf{D}_{\mathrm{SC}}^{(s)} = \mathrm{diag}\left\{d_{\mathrm{SC}}^{(s,1)}, d_{\mathrm{SC}}^{(s,2)}, \ldots, d_{\mathrm{SC}}^{(s,N_{\mathrm{SC}})}\right\}$, the Lagrangian $L(\lambda)$ can be simplified to an explicit expression in terms of $\lambda_0, \lambda_1, \ldots, \lambda_S$, see (3.51) in Appendix 3.10.1. Differentiating $L(\lambda)$ in (3.51)

w.r.t. λ_0 and λ_s $(s \in \Omega)$ and setting the results to zero yield

$$\frac{\partial L(\lambda)}{\partial \lambda_0} = \sum_{n=1}^{N_{\mathrm{BS}}} \frac{a_{\mathrm{BS}}^{(n)}}{\left(d_{\mathrm{BS}}^{(n)} + \lambda_0\right)^2} - 1 \triangleq \chi_0(\lambda_0) = 0 \tag{3.25a}$$

$$\frac{\partial L(\lambda)}{\partial \lambda_s} = \sum_{n=1}^{N_{\mathrm{SC}}} \frac{a_{\mathrm{SC}}^{(s,n)}}{\left(d_{\mathrm{SC}}^{(s,n)} + \lambda_s\right)^2} - 1 \triangleq \chi_s(\lambda_s) = 0, \ \ s \in \Omega \tag{3.25b}$$

where $\mathbf{A}_{\mathrm{BS}} = \mathbf{S}_{\mathrm{BS}}^H \mathbf{G}_{\mathrm{B-M}} \mathbf{R}_{\mathrm{BS}}^H \mathbf{R}_{\mathrm{BS}} \mathbf{G}_{\mathrm{B-M}}^H \mathbf{S}_{\mathrm{BS}}$ is defined with its (n,n)-th entry denoted as $a_{\mathrm{BS}}^{(n)}$, and $\mathbf{A}_{\mathrm{SC}}^{(s)} = \left(\mathbf{S}_{\mathrm{SC}}^{(s)}\right)^H \mathbf{G}_{\mathrm{S-S}}^{(s,s)} \left(\mathbf{R}_{\mathrm{SC}}^{(s)}\right)^H \mathbf{R}_{\mathrm{SC}}^{(s)} \left(\mathbf{G}_{\mathrm{S-S}}^{(s,s)}\right)^H \mathbf{S}_{\mathrm{SC}}^{(s)}$ is defined with its (n,n)-th entry denoted as $a_{\mathrm{SC}}^{(s,n)}$. Based on the equations in (3.25), we propose a bisection search algorithm to compute the numerical values of the optimal Lagrange multipliers λ_s $(s \in \{0, \Omega\})$. The reader is referred to Algorithm 3.1 for a step-by-step description of the search method.

Algorithm 3.1: Bisection search algorithm

Decide search region: Calculate $\chi_s(0)$ and decide the search region. If $\chi_s(0) > 0$, find a $\tilde{\lambda}_s$ satisfying $\chi_s\left(\tilde{\lambda}_s\right) \geq 0$ and then go to the initialization step. Otherwise, output $\lambda_s^* = 0$ as the solution.
Initialize: Set $\lambda_{s,\min} = 0$, $\lambda_{s,\max} = \tilde{\lambda}_s$ and a tolerance ε.
Repeat:

1) Set $\lambda_s = (\lambda_{s,\min} + \lambda_{s,\max})/2$;
2) Calculate $\chi_s(\lambda_s)$;
3) Update the search region: If $\chi_s(\lambda_s) \geq 0$, set lower bound to $\lambda_{s,\min} = \lambda_s$. If $\chi_s(\lambda_s) < 0$, set upper bound to $\lambda_{s,\max} = \lambda_s$.

Until: $\chi_s(\lambda_{s,\min}) - \chi_s(\lambda_{s,\max}) < \varepsilon$ (search error is less than tolerance).
Output: Output $\lambda_s^* = (\lambda_{s,\min} + \lambda_{s,\max})/2$ as the solution.

By substituting the optimal λ_s^* $(s \in \{0, \Omega\})$ obtained into (3.22), the optimal $\mathbf{W}_{\mathrm{BS}}^*, \mathbf{W}_{\mathrm{SC}}^{(s)*}$ $(s \in \Omega)$ can be calculated according to (3.22), where $\mathbf{X}_2$ is assumed to be fixed. As the alternating convex minimization continues, the objective function in (3.13a) monotonically decreases that ensures the algorithms convergence because the objective function is nonnegative hence it is bounded from below. In practice, the alternating minimization is run sufficient number of times so as to reach a steady-state hence practically optimal design. The reader is referred to Algorithm 3.2 for a step-by-step summary of the proposed method.

Algorithm 3.2: RAO

Initialize: Input initial $\mathbf{R}_{\text{BS}}^{(0)}$, $\mathbf{R}_{\text{SC}}^{(s)(0)}$ $(s \in \Omega)$ and a maximum number of iterations N_{iter}. Set $k = 1$.
Repeat:

1) Calculate optimal λ_s^* $(s \in \{0, \Omega\})$ in (3.25) by Algorithm 1;

2) Calculate $\mathbf{W}_{\text{BS}}^{(k)}$ and $\mathbf{W}_{\text{SC}}^{(s)(k)}$ $(s \in \Omega)$ using (3.22);

3) Calculate optimal $\mathbf{R}_{\text{BS}}^{(k)}$ and $\mathbf{R}_{\text{SC}}^{(s)(k)}$ $(s \in \Omega)$ by substituting the $\mathbf{W}_{\text{BS}}^{(k)}$ and $\mathbf{W}_{\text{SC}}^{(s)(k)}$ $(s \in \Omega)$ obtained in step 2) into (3.18);

4) Set $k = k + 1$.

Until: $k = N_{\text{iter}}$.
Output: Output $\mathbf{R}_{\text{BS}}^{(N_{\text{iter}})}$, $\mathbf{R}_{\text{SC}}^{(s)(N_{\text{iter}})}$ $(s \in \Omega)$, $\mathbf{W}_{\text{BS}}^{(N_{\text{iter}})}$ and $\mathbf{W}_{\text{SC}}^{(s)(N_{\text{iter}})}$ $(s \in \Omega)$ as the solution.

3.5.2 Unconstrained Alternating Optimization with Normalization (UAON)

As will be demonstrated later in Section 3.8, the RAO algorithm described above offers superior performance, but at the cost of considerable complexity. Below we present an alternative solution for the sum-MSE problem based on unconstrained alternating convex optimization combined with a simple normalization step. More precisely, by relaxing the equality constraints in (3.9b) and (3.9c) to constraints on average power which are in turn satisfied by normalizing the $\bar{\mathbf{W}}_{\text{BS}}$ and $\bar{\mathbf{W}}_{\text{SC}}^{(s)}$ $(s \in \Omega)$ obtained by minimizing the objective function without constraints, optimal precoding can be achieved quickly with reduced complexity relative to that of the RAO algorithm. The technical details that materialize this approach are given as follows.

With $\mathbf{X}_1$ fixed:

The optimal $\mathbf{R}_{\text{BS}}^*$ and $\mathbf{R}_{\text{SC}}^{(s)*}$ $(s \in \Omega)$ can be acquired in the same way as the constrained alternating optimization, which results in (3.18).

With $\mathbf{X}_2$ fixed:

Given $\mathbf{R}_{\text{BS}}$ and $\mathbf{R}_{\text{SC}}^{(s)}$ $(s \in \Omega)$, the optimization problem becomes

$$\min_{\mathbf{W}_{\text{BS}}, \mathbf{W}_{\text{SC}}} f_2\left(\mathbf{W}_{\text{BS}}, \mathbf{W}_{\text{SC}}\right) \tag{3.26}$$

where no constraints are imposed. Consequently, the global minimizer $\mathbf{W}_{\mathrm{BS}}^{*}, \mathbf{W}_{\mathrm{SC}}^{(s)*}$ $(s \in \Omega)$ are obtained by solving [20]

$$\frac{\partial f_2\left(\mathbf{W}_{\mathrm{BS}}, \mathbf{W}_{\mathrm{SC}}\right)}{\partial \mathbf{W}_{\mathrm{BS}}} = 0, \ \frac{\partial f_2\left(\mathbf{W}_{\mathrm{BS}}, \mathbf{W}_{\mathrm{SC}}\right)}{\partial \mathbf{W}_{\mathrm{SC}}^{(s)}} = 0, \ s \in \Omega \tag{3.27}$$

which yield

$$\bar{\mathbf{W}}_{\mathrm{BS}} = \mathbf{\Phi}_{\mathrm{BS}}^{-1} \mathbf{G}_{\mathrm{B-M}} \mathbf{R}_{\mathrm{BS}}^{H} \tag{3.28a}$$

$$\bar{\mathbf{W}}_{\mathrm{SC}}^{(s)} = \left(\mathbf{\Phi}_{\mathrm{SC}}^{(s)}\right)^{-1} \mathbf{G}_{\mathrm{S-S}}^{(s,s)} \left(\mathbf{R}_{\mathrm{SC}}^{(s)}\right)^{H}, \ s \in \Omega. \tag{3.28b}$$

Then, the normalized optimal solutions are expressed as

$$\mathbf{W}_{\mathrm{BS}}^{*} = \frac{\bar{\mathbf{W}}_{\mathrm{BS}}}{\sqrt{\mathrm{tr}\left\{\bar{\mathbf{W}}_{\mathrm{BS}}^{H} \bar{\mathbf{W}}_{\mathrm{BS}}\right\}}} \tag{3.29a}$$

$$\mathbf{W}_{\mathrm{SC}}^{(s)*} = \frac{\bar{\mathbf{W}}_{\mathrm{SC}}^{(s)}}{\sqrt{\left(\bar{\mathbf{W}}_{\mathrm{SC}}^{(s)}\right)^{H} \bar{\mathbf{W}}_{\mathrm{SC}}^{(s)}}}, \ s \in \Omega. \tag{3.29b}$$

Algorithm 3.3: UAON

Initialize: Set initial $\mathbf{R}_{\mathrm{BS}}^{(0)}$, $\mathbf{R}_{\mathrm{SC}}^{(s)(0)}$ $(s \in \Omega)$ and a maximum number of iterations N_{iter}. Set $k = 1$.
Repeat:

1) Calculate $\bar{\mathbf{W}}_{\mathrm{BS}}^{(k)}$ and $\bar{\mathbf{W}}_{\mathrm{SC}}^{(s)(k)}$ $(s \in \Omega)$ using (3.28), and normalize them using (3.29) to obtain $\mathbf{W}_{\mathrm{BS}}^{(k)}, \mathbf{W}_{\mathrm{SC}}^{(s)(k)}$ $(s \in \Omega)$;

2) Calculate optimal $\mathbf{R}_{\mathrm{BS}}^{(k)}$ and $\mathbf{R}_{s}^{\mathrm{SC}(k)}$ $(s \in \Omega)$ by substituting the $\mathbf{W}_{\mathrm{BS}}^{(k)}$ and $\mathbf{W}_{\mathrm{SC}}^{(s)(k)}$ $(s \in \Omega)$ obtained in step 1) into (3.18);

3) Set $k = k + 1$.

Until: $k = N_{\mathrm{iter}}$.
Output: Output $\mathbf{R}_{\mathrm{BS}}^{(N_{\mathrm{iter}})}$, $\mathbf{R}_{\mathrm{SC}}^{(s)(N_{\mathrm{iter}})}$ $(s \in \Omega)$, $\mathbf{W}_{\mathrm{BS}}^{(N_{\mathrm{iter}})}$ and $\mathbf{W}_{\mathrm{SC}}^{(s)(N_{\mathrm{iter}})}$ $(s \in \Omega)$ as the solution.

We remark that although UAON is much simpler than RAO, the optimal precoder based on UAON still requires the knowledge about the channels from the nodes to both MUEs and SUEs.

3.6 Separate MSE Minimization Based Two-level Precoding in HetNet

In Section 3.5, the precoders at the BS and all the SCs are jointly designed by minimizing the sum-MSE. However, due to the non-convexity of the objective functions and constraints, no non-iterative algorithms are available for the precoder designs and intensive computation is required. In this section, a simplified solution procedure is derived based on separate MSE minimization where block diagonalization techniques act in the first-level and the second-level precoders at each node are designed separately. As shown in what follows, the separate treatment of individual precoders leads to a non-iterative algorithm.

3.6.1 MSE Minimization at the BS

In order to determine the precoding matrix $\mathbf{W}_{\mathrm{BS}}$ at the BS, the signal and interference associated with the BS are taken into account in a way similar to [42]. This leads to

$$\min_{\mathbf{W}_{\mathrm{BS}},\mathbf{R}_{\mathrm{BS}}} \mathrm{E}\left[\left\|\hat{\mathbf{x}}_{\mathrm{BS}}-\mathbf{x}_{\mathrm{BS}}\right\|^2\right] \tag{3.30a}$$

$$\text{subject to } \left\|\mathbf{G}_{\mathrm{B-S}}^H\mathbf{W}_{\mathrm{BS}}\mathbf{x}_{\mathrm{BS}}\right\|^2 \le \gamma_{\mathrm{BS}} \tag{3.30b}$$

$$\mathrm{tr}\left\{\mathbf{W}_{\mathrm{BS}}^H\mathbf{W}_{\mathrm{BS}}\right\} \le 1 \tag{3.30c}$$

where $\gamma_{\mathrm{BS}} > 0$ is a threshold parameter set to control the relative interference involved. The item in the objective in (3.30a) is the sum of squares of errors seen by the MUEs assuming no interferences included, i.e., $\mathbf{y}_{\mathrm{BS}} = \mathbf{G}_{\mathrm{B-M}}^H\mathbf{W}_{\mathrm{BS}}\mathbf{x}_{\mathrm{BS}} + \mathbf{n}_{\mathrm{BS}}$; and the item in (3.30b) is the sum of squares of the interference seen by the SUEs. By tuning γ_{BS}, the BS trades off the beamforming gains for its target MUEs against interference reduction to the neighboring SUEs. We stress that the objective function is not jointly convex with respect to all design variables, but that the induced interference constraint in (3.30b) and the average power constraint in (3.30c) are convex. Certainly, similar to the RAO proposed in Section 3.5, an iterative algorithm could provide an optimal solution for (3.30). To obtain a non-iterative algorithm, we further simplify (3.30) by employing BD technique at the BS side as a first-level precoder. Thereby, all the inter-MUE interferences are eliminated and each MUE perceives an interference-free MIMO channel, which means that

problem in (3.30) can be divided into K independent sub-problems of the form

$$\min_{\mathbf{W}_{\mathrm{BS}}^{(i)},\mathbf{R}_{\mathrm{BS}}^{(i)}} \ \mathrm{E}\left[\left\|\hat{\mathbf{x}}_{\mathrm{BS}}^{(i)}-\mathbf{x}_{\mathrm{BS}}^{(i)}\right\|^2\right] \tag{3.31a}$$

$$\text{subject to} \ \left\|\mathbf{G}_{\mathrm{B-S}}^{H}\mathbf{W}_{\mathrm{BS}}^{(i)}\mathbf{x}_{\mathrm{BS}}^{(i)}\right\|^2 \le \gamma_{\mathrm{BS}}^{(i)} \tag{3.31b}$$

$$\mathrm{tr}\left\{\left(\mathbf{W}_{\mathrm{BS}}^{(i)}\right)^{H}\mathbf{W}_{\mathrm{BS}}^{(i)}\right\} \le \alpha_{\mathrm{BS}}^{(i)} \tag{3.31c}$$

$$\left(\bar{\mathbf{G}}_{\mathrm{B-M}}^{(i)}\right)^{H}\mathbf{W}_{\mathrm{BS}}^{(i)} = \mathbf{0} \tag{3.31d}$$

where $i \in I$, $\bar{\mathbf{G}}_{\mathrm{B-M}}^{(i)} = \left[\mathbf{G}_{\mathrm{B-M}}^{(1)},\ldots,\mathbf{G}_{\mathrm{B-M}}^{(i-1)},\mathbf{G}_{\mathrm{B-M}}^{(i+1)},\ldots\mathbf{G}_{\mathrm{B-M}}^{(K)}\right]$, $\sum_{i=1}^{K}\gamma_{\mathrm{BS}}^{(i)} = \gamma_{\mathrm{BS}}$ and $\sum_{i=1}^{K}\alpha_{\mathrm{BS}}^{(i)} = 1$ with $\gamma_{\mathrm{BS}}^{(i)} > 0$ and $\alpha_{\mathrm{BS}}^{(i)} > 0$. Here, the BD constraint of (3.31d) is imposed to eliminate all inter-MUE interferences. By applying the SVD, we have $\bar{\mathbf{G}}_{\mathrm{B-M}}^{(i)} = \mathbf{U}_{\mathrm{BS}}^{(i)}\mathbf{Z}_{\mathrm{BS}}^{(i)}\left[\mathbf{V}_{\mathrm{BS}}^{(i,1)}\ \mathbf{V}_{\mathrm{BS}}^{(i,0)}\right]^{H}$, where $\mathbf{Z}_{\mathrm{BS}}^{(i)}$ is the diagonal matrix with non-negative singular values as its diagonal elements, $\mathbf{V}_{\mathrm{BS}}^{(i,1)}$ contains the singular vectors corresponding to the nonzero singular values and $\mathbf{V}_{\mathrm{BS}}^{(i,0)}$ consists of vectors corresponding to the zero singular values. Hence, $\mathbf{V}_{\mathrm{BS}}^{(i,0)}$ is an orthogonal basis for the null space of $\bar{\mathbf{G}}_{\mathrm{B-M}}^{(i)}$. For simplicity, we suppose that $\mathbf{W}_{\mathrm{BS}}^{(i)} = \mathbf{W}_{\mathrm{BS},1}^{(i)}\mathbf{W}_{\mathrm{BS},2}^{(i)}$ for $\forall i \in I$ with $\mathbf{W}_{\mathrm{BS},1}^{(i)} = \mathbf{V}_{\mathrm{BS}}^{(i,0)}$ to satisfy the BD constraint of (3.31d). In this way, we transform our focus from the design of $\mathbf{W}_{\mathrm{BS}}^{(i)}$ to that of $\mathbf{W}_{\mathrm{BS},2}^{(i)}$.

Similar to the RAO algorithm, suppose that $\mathbf{W}_{\mathrm{BS}}^{(i)}$ are fixed, then the optimal $\mathbf{R}_{\mathrm{BS}}^{(i)}$ $(i \in I)$ can be expressed as

$$\mathbf{R}_{\mathrm{BS}}^{(i)*} = \left(\mathbf{W}_{\mathrm{BS}}^{(i)}\right)^{H}\mathbf{G}_{\mathrm{B-M}}^{(i)}\left[\left(\mathbf{G}_{\mathrm{B-M}}^{(i)}\right)^{H}\mathbf{W}_{\mathrm{BS}}^{(i)}\left(\mathbf{W}_{\mathrm{BS}}^{(i)}\right)^{H}\mathbf{G}_{\mathrm{B-M}}^{(i)} + \sigma_0^2\mathbf{I}_{N_{\mathrm{UE}}}\right]^{-1}. \tag{3.32}$$

Substituting (3.32) into the objective function of (3.31a), we obtain

$$\mathrm{MSE}_{\mathrm{BS}}^{(i)} = N_{\mathrm{S}} - \mathrm{tr}\left\{\left[\sigma_0^2\left(\left(\mathbf{G}_{\mathrm{B-M}}^{(i)}\right)^{H}\mathbf{W}_{\mathrm{BS}}^{(i)}\left(\mathbf{W}_{\mathrm{BS}}^{(i)}\right)^{H}\mathbf{G}_{\mathrm{B-M}}^{(i)}\right)^{-1} + \mathbf{I}_{N_{\mathrm{UE}}}\right]^{-1}\right\} \tag{3.33}$$

which means that minimizing MSE is equivalent to maximizing the term of $\left\|\left(\mathbf{G}_{\mathrm{B-M}}^{(i)}\right)^{H}\mathbf{W}_{\mathrm{BS}}^{(i)}\right\|_{\mathrm{F}}^2$. Since transmission symbols satisfy $\mathrm{E}\{\mathbf{x}\} = 0$ and $\mathrm{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$, the left-hand in (3.31c) equals to $\mathrm{tr}\left\{\mathbf{Q}_{\mathrm{BS}}^{(i)}\right\}$, where $\mathbf{Q}_{\mathrm{BS}}^{(i)} = \mathbf{B}_{\mathrm{BS}}^{(i)}\mathbf{W}_{\mathrm{BS},2}^{(i)}\left(\mathbf{W}_{\mathrm{BS},2}^{(i)}\right)^{H}\mathbf{B}_{\mathrm{BS}}^{(i)}$ with $\mathbf{B}_{\mathrm{BS}}^{(i)} = \left[\tilde{\mathbf{G}}_{\mathrm{B-S}}^{(i)}\left(\tilde{\mathbf{G}}_{\mathrm{B-S}}^{(i)}\right)^{H}\right]^{\frac{1}{2}}$, where $\tilde{\mathbf{G}}_{\mathrm{B-S}}^{(i)} \triangleq \left(\mathbf{W}_{\mathrm{BS},1}^{(i)}\right)^{H}\mathbf{G}_{\mathrm{B-S}}$ denotes the equivalent channel matrix. Similarly, suppose that $\tilde{\mathbf{G}}_{\mathrm{B-M}}^{(i)} \triangleq \left(\mathbf{W}_{\mathrm{BS},1}^{(i)}\right)^{H}\mathbf{G}_{\mathrm{B-M}}^{(i)}$, then the

objective function becomes

$$\begin{aligned}&\left\|\left(\mathbf{G}_{\mathrm{B-M}}^{(i)}\right)^{H}\mathbf{W}_{\mathrm{BS}}^{(i)}\right\|_{\mathrm{F}}^{2}=\mathrm{tr}\left\{\left(\tilde{\mathbf{G}}_{\mathrm{B-M}}^{(i)}\right)^{H}\left(\mathbf{B}_{\mathrm{BS}}^{(i)}\right)^{-1}\mathbf{Q}_{\mathrm{BS}}^{(i)}\left(\mathbf{B}_{\mathrm{BS}}^{(i)}\right)^{-1}\tilde{\mathbf{G}}_{\mathrm{B-M}}^{(i)}\right\}\\ =\ &\mathrm{tr}\left\{\mathrm{P}_{\mathrm{BS}}^{(i)}\mathbf{\Sigma}_{\mathrm{BS}}^{(i)}\left(\mathbf{T}_{\mathrm{BS}}^{(i)}\right)^{H}\mathbf{Q}_{\mathrm{BS}}^{(i)}\mathbf{T}_{\mathrm{BS}}^{(i)}\left(\mathbf{\Sigma}_{\mathrm{BS}}^{(i)}\right)^{H}\left(\mathrm{P}_{\mathrm{BS}}^{(i)}\right)^{H}\right\}\end{aligned}\tag{3.34}$$

where $\left(\tilde{\mathbf{G}}_{\mathrm{B-M}}^{(i)}\right)^{H}\left(\mathbf{B}_{\mathrm{BS}}^{(i)}\right)^{-1}=\mathrm{P}_{\mathrm{BS}}^{(i)}\mathbf{\Sigma}_{\mathrm{BS}}^{(i)}\left(\mathbf{T}_{\mathrm{BS}}^{(i)}\right)^{H}$ with $\mathbf{\Sigma}_{\mathrm{BS}}^{(i)}=\mathrm{diag}\left\{\sigma_{\mathrm{BS}}^{(i,1)},\ldots,\sigma_{\mathrm{BS}}^{(i,N_{\mathrm{UE}})}\right\}$ is obtained by SVD in order to further simplify the problem. Using the Hadamards inequality (see, e.g., [25]), the optimal solution for maximizing (3.34) is obtained as $\mathbf{Q}_{\mathrm{BS}}^{(i)*}=\mathbf{T}_{\mathrm{BS}}^{(i)}\mathbf{\Lambda}_{\mathrm{BS}}^{(i)}\left(\mathbf{T}_{\mathrm{BS}}^{(i)}\right)^{H}$, where $\mathbf{\Lambda}_{\mathrm{BS}}^{(i)}=\mathrm{diag}\left\{\lambda_{\mathrm{BS}}^{(i,1)},\ldots,\lambda_{\mathrm{BS}}^{(i,N_{\mathrm{UE}})}\right\}$ with $\lambda_{\mathrm{BS}}^{(i,n)}$ $(n=1,\ldots,N_{\mathrm{UE}})$ being the only parameters to be determined. Thus, the objective function in (3.31) can be transformed into $\sum_{n=1}^{N_{\mathrm{UE}}}\left(\sigma_{\mathrm{BS}}^{(i,n)}\right)^{2}\lambda_{\mathrm{BS}}^{(i,n)}$, the constraint in (3.31b) is equivalent to $\sum_{n=1}^{N_{\mathrm{UE}}}\lambda_{\mathrm{BS}}^{(i,n)}\leq\gamma_{\mathrm{BS}}^{(i)}$, and (3.31c) is equivalent to $\mathrm{tr}\left\{\mathbf{\Lambda}_{\mathrm{BS}}^{(i)}\mathbf{X}_{\mathrm{BS}}^{(i)}\right\}=\sum_{n=1}^{N_{\mathrm{UE}}}x_{\mathrm{BS}}^{(i,n)}\lambda_{\mathrm{BS}}^{(i,n)}\leq\alpha_{\mathrm{BS}}^{(i)}$ with $\mathbf{X}_{\mathrm{BS}}^{(i)}\triangleq\left(\mathbf{T}_{\mathrm{BS}}^{(i)}\right)^{H}\left(\mathbf{B}_{\mathrm{BS}}^{(i)}\right)^{-2}\mathbf{T}_{\mathrm{BS}}^{(i)}$, where $x_{\mathrm{BS}}^{(i,n)}$ denotes the (n,n)-th element of $\mathbf{X}_{\mathrm{BS}}^{(i)}$. In this way, the optimization problem (3.31) can be formulated as

$$\min_{\lambda_{\mathrm{BS}}^{(i)}}\ \left(\mathbf{c}_{\mathrm{BS}}^{(i)}\right)^{T}\lambda_{\mathrm{BS}}^{(i)}\tag{3.35a}$$

$$\text{subject to}\ \ \mathbf{e}^{T}\lambda_{\mathrm{BS}}^{(i)}\leq\gamma_{\mathrm{BS}}^{(i)}\tag{3.35b}$$

$$\left(\mathbf{x}_{\mathrm{BS}}^{(i)}\right)^{T}\lambda_{\mathrm{BS}}^{(i)}\leq\alpha_{\mathrm{BS}}^{(i)}\tag{3.35c}$$

$$-\lambda_{\mathrm{BS}}^{(i)}\leq\mathbf{0}\tag{3.35d}$$

where $\lambda_{\mathrm{BS}}^{(i)}=\left[\lambda_{\mathrm{BS}}^{(i,1)},\ldots,\lambda_{\mathrm{BS}}^{(i,N_{\mathrm{UE}})}\right]^{T}$, $\mathbf{c}_{\mathrm{BS}}^{(i)}=\left[-\left(\sigma_{\mathrm{BS}}^{(i,n)}\right)^{2},\ldots,-\left(\sigma_{\mathrm{BS}}^{(i,N_{\mathrm{UE}})}\right)^{2}\right]^{T}$ and $\mathbf{x}_{\mathrm{BS}}^{(i)}=\left[x_{\mathrm{BS}}^{(i,1)},\ \ldots,x_{\mathrm{BS}}^{(i,N_{\mathrm{UE}})}\right]^{T}$. Notably, (3.35) is a standard linear programming (LP) problem and can easily be solved by CVX. Upon obtaining the optimal $\lambda_{\mathrm{BS}}^{(i)*}$, the optimal precoder at the BS is found to be

$$\mathbf{W}_{\mathrm{BS}}\mathbf{W}_{\mathrm{BS}}^{H}=\mathbf{V}_{\mathrm{BS}}^{(i,0)}\left(\mathbf{B}_{\mathrm{BS}}^{(i)}\right)^{-1}\mathbf{T}_{\mathrm{BS}}^{(i)}\mathbf{\Lambda}_{\mathrm{BS}}^{(i)*}\left(\mathbf{T}_{\mathrm{BS}}^{(i)}\right)^{H}\left(\mathbf{B}_{\mathrm{BS}}^{(i)}\right)^{-1}\left(\mathbf{V}_{\mathrm{BS}}^{(i,0)}\right)^{H}\tag{3.36}$$

from which the optimal $\mathbf{W}_{\mathrm{BS}}$ can be obtained by SVD.

From the above solution procedure, it is clear that constructing an optimal precoder at the BS only requires the knowledge about the channels from BS to both MUEs and SUEs, a less stringent requirement relative to the sum-MUE minimiza-

tion based precoding scheme.

3.6.2 MSE Minimization at each SC

Similarly, the design of precoding vector $\mathbf{W}_{\mathrm{SC}}^{(s)}$ ($s \in \Omega$) can be handled by solving the LP problem for each SUE, given by

$$\min_{\lambda_{\mathrm{SC}}^{(s,j)}} \left(\mathbf{c}_{\mathrm{SC}}^{(s,j)}\right)^T \lambda_{\mathrm{SC}}^{(s,j)} \tag{3.37a}$$

$$\text{subject to } \mathbf{e}^T \lambda_{\mathrm{SC}}^{(s,j)} \leq \gamma_{\mathrm{SC}}^{(s,j)} \tag{3.37b}$$

$$\left(\mathbf{x}_{\mathrm{SC}}^{(s,j)}\right)^T \lambda_{\mathrm{SC}}^{(s,j)} \leq \alpha_{\mathrm{SC}}^{(s,j)} \tag{3.37c}$$

$$-\lambda_{\mathrm{SC}}^{(s,j)} \leq \mathbf{0} \tag{3.37d}$$

where $j \in J_s$, $\lambda_{\mathrm{SC}}^{(s,j)} = \left[\lambda_{\mathrm{SC}}^{(s,j,1)}, \ldots, \lambda_{\mathrm{SC}}^{(s,j,N_{\mathrm{UE}})}\right]^T$, $\mathbf{c}_{\mathrm{SC}}^{(s,j)} = \left[-\left(\sigma_{\mathrm{SC}}^{(s,j,n)}\right)^2, \ldots, -\left(\sigma_{\mathrm{SC}}^{(s,j,N_{\mathrm{UE}})}\right)^2\right]^T$ and $\mathbf{x}_{\mathrm{SC}}^{(s,j)} = \left[x_{\mathrm{SC}}^{(s,j,1)}, \ldots, x_{\mathrm{SC}}^{(s,j,N_{\mathrm{UE}})}\right]^T$. Here, the vector elements are calculated accordingly based on the definitions and derivations in Subsection 3.6.1.

Like the precoder at the BS, only the knowledge about channels from s-th SC to both MUEs and SUEs are required to construct the optimal precoder at s-th SC.

3.7 Robust Precoding Design With Imperfect CSI in HetNet

Since perfect CSI is required in the above precoding design, it is often not practical due to channel estimation error, feedback error and quantization error. In this section, we propose more practical precoders for the HetNet with imperfect CSI known at each node.

Assume that the CSI errors of all links are stochastic and modeled as $\hat{\mathbf{G}}_* = \mathbf{G}_* + \mathbf{\Xi}_*$, where $* \in \{\mathrm{B-M}, \mathrm{B-S}, \mathrm{S-M}, \mathrm{S-S}\}$, $\hat{\mathbf{G}}_*$ is the estimated channel matrix, and $\mathbf{\Xi}_*$ denotes the channel estimation error matrix which is assumed to be Gaussian distributed with $\mathrm{E}[\mathbf{\Xi}] = \mathbf{0}$ and $\mathrm{E}\left[\mathrm{vec}(\mathbf{\Xi}_*)\,\mathrm{vec}(\mathbf{\Xi}_*)^H\right] = \sigma_h^2\mathbf{I}$.

3.7.1 Robust RAO With Imperfect CSI

With imperfect CSI known at each node, the RAO problem becomes

$$\min_{\substack{\mathbf{W}_{\mathrm{BS}},\mathbf{W}_{\mathrm{SC}}^{(t)}\\ \mathbf{R}_{\mathrm{BS}},\mathbf{R}_{\mathrm{SC}}^{(t)},t\in\Omega}} \quad f\left(\mathbf{W}_{\mathrm{BS}},\mathbf{W}_{\mathrm{SC}},\mathbf{R}_{\mathrm{BS}},\mathbf{R}_{\mathrm{SC}}\right)\Big|\hat{\mathbf{G}}_* \tag{3.38a}$$

$$\text{subject to} \quad (3.13b),\ (3.13c),\ (3.13d) \tag{3.38b}$$

where

$$f\left(\mathbf{W}_{\mathrm{BS}},\mathbf{W}_{\mathrm{SC}},\mathbf{R}_{\mathrm{BS}},\mathbf{R}_{\mathrm{SC}}\right)\Big|\hat{\mathbf{G}}_* = \hat{\mathrm{MSE}}_{\mathrm{BS}} + \hat{\mathrm{MSE}}_{\mathrm{SC}} \tag{3.39}$$

with

$$\begin{aligned}
&\hat{\mathrm{MSE}}_{\mathrm{BS}} \triangleq \mathrm{E}\left[\left\|\hat{\mathbf{x}}_{\mathrm{BS}}-\mathbf{x}_{\mathrm{BS}}\right\|^2\Big|\hat{\mathbf{G}}_*\right]\\
=&\mathrm{tr}\left\{\mathbf{R}_{\mathrm{BS}}\left[\hat{\mathbf{G}}_{\mathrm{B-M}}^H\mathbf{W}_{\mathrm{BS}}\mathbf{W}_{\mathrm{BS}}^H\hat{\mathbf{G}}_{\mathrm{B-M}}+\sum_{s=1}^{S}\left(\hat{\mathbf{G}}_{\mathrm{S-M}}^{(s)}\right)^H\mathbf{W}_{\mathrm{SC}}^{(s)}\left(\mathbf{W}_{\mathrm{SC}}^{(s)}\right)^H\hat{\mathbf{G}}_{\mathrm{S-M}}^{(s)}\right]\mathbf{R}_{\mathrm{BS}}^H\right.\\
&\left.-2\mathbf{R}_{\mathrm{BS}}\hat{\mathbf{G}}_{\mathrm{B-M}}^H\mathbf{W}_{\mathrm{BS}}+\mathbf{I}_{KN_{\mathrm{S}}}+\sigma_0^2\mathbf{R}_{\mathrm{BS}}\mathbf{R}_{\mathrm{BS}}^H\right\}+\sigma_h^2\bar{\omega}\mathrm{tr}\left\{\mathbf{R}_{\mathrm{BS}}^H\mathbf{R}_{\mathrm{BS}}\right\}
\end{aligned} \tag{3.40}$$

$$\begin{aligned}
&\hat{\mathrm{MSE}}_{\mathrm{SC}} \triangleq \mathrm{E}\left[\sum_{s=1}^{S}\left\|\hat{\mathbf{x}}_{\mathrm{SC}}^{(s)}-\mathbf{x}_{\mathrm{SC}}^{(s)}\right\|^2\Big|\hat{\mathbf{G}}_*\right]\\
=&\sum_{s=1}^{S}\mathrm{tr}\left\{\mathbf{R}_{\mathrm{SC}}^{(s)}\left[\left(\hat{\mathbf{G}}_{\mathrm{B-S}}^{(s)}\right)^H\mathbf{W}_{\mathrm{BS}}\mathbf{W}_{\mathrm{BS}}^H\hat{\mathbf{G}}_{\mathrm{B-S}}^{(s)}+\sum_{t=1}^{S}\left(\hat{\mathbf{G}}_{\mathrm{S-S}}^{(t,s)}\right)^H\mathbf{W}_{\mathrm{SC}}^{(t)}\left(\mathbf{W}_{\mathrm{SC}}^{(t)}\right)^H\hat{\mathbf{G}}_{\mathrm{S-S}}^{(t,s)}\right]\left(\mathbf{R}_{\mathrm{SC}}^{(s)}\right)^H\right.\\
&\left.-2\mathbf{R}_{\mathrm{SC}}^{(s)}\left(\hat{\mathbf{G}}_{\mathrm{S-S}}^{(s,s)}\right)^H\mathbf{W}_{\mathrm{SC}}^{(s)}+\mathbf{I}_{L_sN_{\mathrm{S}}}+\sigma_0^2\mathbf{R}_{\mathrm{SC}}^{(s)}\left(\mathbf{R}_{\mathrm{SC}}^{(s)}\right)^H\right\}+\sigma_h^2\bar{\omega}\mathrm{tr}\left\{\left(\mathbf{R}_{\mathrm{SC}}^{(s)}\right)^H\mathbf{R}_{\mathrm{SC}}^{(s)}\right\}
\end{aligned} \tag{3.41}$$

and $\bar{\omega}=\mathrm{tr}\left\{\mathbf{W}_{\mathrm{BS}}^H\mathbf{W}_{\mathrm{BS}}\right\}+\sum_{t=1}^{S}\mathrm{tr}\left\{\left(\mathbf{W}_{\mathrm{SC}}^{(t)}\right)^H\mathbf{W}_{\mathrm{SC}}^{(t)}\right\}$. Thus, following the step of RAO, key equations can be derived from the KKT conditions for problem (3.38) as

$$\mathbf{R}_{\mathrm{BS}}^{(i)*}=\left(\mathbf{W}_{\mathrm{BS}}^{(i)}\right)^H\hat{\mathbf{G}}_{\mathrm{B-M}}^{(i)}\left[\hat{\mathbf{\Psi}}_{\mathrm{BS}}^{(i)}+\left(\sigma_0^2+\sigma_h^2\bar{\omega}\right)\mathbf{I}_{N_{\mathrm{UE}}}\right]^{-1},\ i\in I \tag{3.42a}$$

$$\mathbf{R}_{\mathrm{SC}}^{(s,j)*}=\left(\mathbf{W}_{\mathrm{SC}}^{(s,j)}\right)^H\hat{\mathbf{G}}_{\mathrm{S-S}}^{(s,s,j)}\left[\hat{\mathbf{\Psi}}_{\mathrm{SC}}^{(s,j)}+\left(\sigma_0^2+\sigma_h^2\bar{\omega}\right)\mathbf{I}_{N_{\mathrm{UE}}}\right]^{-1},\ j\in J_s, s\in\Omega. \tag{3.42b}$$

$$\mathbf{W}_{\mathrm{BS}}^{*}=\left[\hat{\mathbf{\Phi}}_{\mathrm{BS}}+\left(\hat{\lambda}_0+\sigma_h^2\bar{r}\right)\mathbf{I}_{N_{\mathrm{BS}}}\right]^{-1}\hat{\mathbf{G}}_{\mathrm{B-M}}\mathbf{R}_{\mathrm{BS}}^H \tag{3.42c}$$

$$\mathbf{W}_{\mathrm{SC}}^{(s)*}=\left[\hat{\mathbf{\Phi}}_{\mathrm{SC}}^{(s)}+\left(\hat{\lambda}_s+\sigma_h^2\bar{r}\right)\mathbf{I}_{N_{\mathrm{SC}}}\right]^{-1}\hat{\mathbf{G}}_{\mathrm{S-S}}^{(s,s)}\left(\mathbf{R}_{\mathrm{SC}}^{(s)}\right)^H,\ s\in\Omega. \tag{3.42d}$$

where $\hat{\mathbf{\Psi}}_{\rm BS}^{(i)} = \left(\hat{\mathbf{G}}_{\rm B-M}^{(i)}\right)^H \mathbf{W}_{\rm BS}\mathbf{W}_{\rm BS}^H\hat{\mathbf{G}}_{\rm B-M}^{(i)} + \sum_{s=1}^{S}\left(\hat{\mathbf{G}}_{\rm S-M}^{(s,i)}\right)^H \mathbf{W}_{\rm SC}^{(s)}\left(\mathbf{W}_{\rm SC}^{(s)}\right)^H\hat{\mathbf{G}}_{\rm S-M}^{(s,i)}$, $\hat{\mathbf{\Psi}}_{\rm SC}^{(s,j)} = \left(\hat{\mathbf{G}}_{\rm B-S}^{(s,j)}\right)^H \mathbf{W}_{\rm BS}\mathbf{W}_{\rm BS}^H\hat{\mathbf{G}}_{\rm B-S}^{(s,j)} + \sum_{t=1}^{S}\left(\hat{\mathbf{G}}_{\rm S-S}^{(t,s,j)}\right)^H \mathbf{W}_{\rm SC}^{(t)}\left(\mathbf{W}_{\rm SC}^{(t)}\right)^H\hat{\mathbf{G}}_{\rm S-S}^{(t,s,j)}$, $\hat{\mathbf{\Phi}}_{\rm BS} = \hat{\mathbf{G}}_{\rm B-M}\mathbf{R}_{\rm BS}^H\mathbf{R}_{\rm BS}\hat{\mathbf{G}}_{\rm B-M}^H + \sum_{s=1}^{S}\hat{\mathbf{G}}_{\rm B-S}^{(s)}\left(\mathbf{R}_{\rm SC}^{(s)}\right)^H\mathbf{R}_{\rm SC}^{(s)}\left(\hat{\mathbf{G}}_{\rm B-S}^{(s)}\right)^H$, $\hat{\mathbf{\Phi}}_{\rm SC}^{(s)} = \hat{\mathbf{G}}_{\rm S-M}^{(s)}\mathbf{R}_{\rm BS}^H\mathbf{R}_{\rm BS}\left(\hat{\mathbf{G}}_{\rm S-M}^{(s)}\right)^H + \sum_{t=1}^{S}\hat{\mathbf{G}}_{\rm S-S}^{(s,t)}\left(\mathbf{R}_{\rm SC}^{(t)}\right)^H\mathbf{R}_{\rm SC}^{(t)}\left(\hat{\mathbf{G}}_{\rm S-S}^{(s,t)}\right)^H$, and $\bar{r} = \mathrm{tr}\left\{\mathbf{R}_{\rm BS}^H\mathbf{R}_{\rm BS}\right\} + \sum_{t=1}^{S}\mathrm{tr}\left\{\left(\mathbf{R}_{\rm SC}^{(t)}\right)^H\mathbf{R}_{\rm SC}^{(t)}\right\}$.

Based on (3.42), a robust RAO algorithm can be constructed in a way similar to that in Subsection 3.5.1 where the RAO algorithm was developed. The optimal Lagrange in the present case satisfy

$$\frac{\partial L\left(\hat{\lambda}\right)}{\partial\hat{\lambda}_0} = \sum_{n=1}^{N_{\rm BS}}\frac{\hat{a}_{\rm BS}^{(n)}}{\left(\hat{d}_{\rm BS}^{(n)} + \hat{\lambda}_0 + \sigma_h^2\bar{r}\right)^2} - 1 \triangleq \hat{\chi}_0\left(\hat{\lambda}_0\right) = 0 \tag{3.43a}$$

$$\frac{\partial L\left(\hat{\lambda}\right)}{\partial\hat{\lambda}_s} = \sum_{n=1}^{N_{\rm SC}}\frac{\hat{a}_{\rm SC}^{(s,n)}}{\left(\hat{d}_{\rm SC}^{(s,n)} + \hat{\lambda}_s + \sigma_h^2\bar{r}\right)^2} - 1 \triangleq \hat{\chi}_s\left(\hat{\lambda}_s\right) = 0,\ s\in\Omega \tag{3.43b}$$

where $\hat{a}_{\rm BS}^{(n)}$, $\hat{d}_{\rm BS}^{(n)}$, $\hat{a}_{\rm SC}^{(s,n)}$ and $\hat{d}_{\rm SC}^{(s,n)}$ are defined in an entirely similar way to their counterparts in Subsection 3.5.1. Evidently, a bisection search is applicable to (3.43) to identify the optimal Lagrange multipliers.

3.7.2 Robust UAON With Imperfect CSI

As expected, the design of robust UAON with imperfect CSI can be carried out by steps in parallel to those of Algorithm 3. Specifically, the optimal $\mathbf{R}_{\rm BS}^*$ and $\mathbf{R}_{\rm SC}^{(s)*}$ ($s\in\Omega$) have the same expressions as (3.42a) and (3.42b). Consequently, the global minimizers $\mathbf{W}_{\rm BS}^*$ and $\mathbf{W}_{\rm SC}^{(s)*}$ ($s\in\Omega$) for robust UAON can be obtained by first computing

$$\bar{\mathbf{W}}_{\rm BS} = \left(\hat{\mathbf{\Phi}}_{\rm BS} + \sigma_h^2\bar{r}\mathbf{I}_{N_{\rm BS}}\right)^{-1}\hat{\mathbf{G}}_{\rm B-M}\mathbf{R}_{\rm BS}^H \tag{3.44a}$$

$$\bar{\mathbf{W}}_{\rm SC}^{(s)} = \left(\hat{\mathbf{\Phi}}_{\rm SC}^{(s)} + \sigma_h^2\bar{r}\mathbf{I}_{N_{\rm SC}}\right)^{-1}\hat{\mathbf{G}}_{\rm S-S}^{(s,s)}\left(\mathbf{R}_{\rm SC}^{(s)}\right)^H,\ s\in\Omega. \tag{3.44b}$$

followed by a norm normalization step as in (3.29).

3.7.3 Robust Non-iterative Algorithm With Imperfect CSI

With imperfect CSI, there is a robust counterpart of the non-iterative precoding developed in Subsection 3.6 based on separate MSE. To see this, note that the

optimal $\mathbf{R}_{\mathrm{BS}}^{(i)}$ $(i \in I)$ with fixed $\mathbf{W}_{\mathrm{BS}}^{(i)}$ can be expressed as

$$\mathbf{R}_{\mathrm{BS}}^{(i)*} = \left(\mathbf{W}_{\mathrm{BS}}^{(i)}\right)^H \hat{\mathbf{G}}_{\mathrm{B-M}}^{(i)} \left[\left(\hat{\mathbf{G}}_{\mathrm{B-M}}^{(i)}\right)^H \mathbf{W}_{\mathrm{BS}}^{(i)} \left(\mathbf{W}_{\mathrm{BS}}^{(i)}\right)^H \hat{\mathbf{G}}_{\mathrm{B-M}}^{(i)} + \left(\sigma_0^2 + \sigma_h^2 \bar{\omega}_{\mathrm{BS}}^{(i)}\right) \mathbf{I}_{N_{\mathrm{UE}}}\right]^{-1} \tag{3.45}$$

where $\bar{\omega}_{\mathrm{BS}}^{(i)} = \mathrm{tr}\left\{\mathbf{W}_{\mathrm{BS}}^{(i)} \left(\mathbf{W}_{\mathrm{BS}}^{(i)}\right)^H\right\}$. By substituting (3.45) into the imperfect CSI based objective function, it is evident that minimizing MSE can be transformed into maximizing the term of $\left\| \left(\hat{\mathbf{G}}_{\mathrm{B-M}}^{(i)}\right)^H \mathbf{W}_{\mathrm{BS}}^{(i)} \right\|_{\mathrm{F}}^2$. In this way, the optimization problem after certain transformations can be rewritten as

$$\min_{\hat{\lambda}_{\mathrm{BS}}^{(i)}} \quad \left(\hat{\mathbf{c}}_{\mathrm{BS}}^{(i)}\right)^T \hat{\lambda}_{\mathrm{BS}}^{(i)} \tag{3.46a}$$

$$\text{subject to} \quad \mathbf{e}^T \hat{\lambda}_{\mathrm{BS}}^{(i)} \leq \gamma_{\mathrm{BS}}^{(i)} \tag{3.46b}$$

$$\left(\hat{\mathbf{x}}_{\mathrm{BS}}^{(i)}\right)^T \hat{\lambda}_{\mathrm{BS}}^{(i)} \leq \alpha_{\mathrm{BS}}^{(i)} \tag{3.46c}$$

where $\hat{\lambda}_{\mathrm{BS}}^{(i)} = \left[\hat{\lambda}_{\mathrm{BS}}^{(i,1)}, \ldots, \hat{\lambda}_{\mathrm{BS}}^{(i,N_{\mathrm{UE}})}\right]^T$, $\hat{\mathbf{c}}_{\mathrm{BS}}^{(i)} = \left[-\left(\hat{\sigma}_{\mathrm{BS}}^{(i,n)}\right)^2, \ldots, -\left(\hat{\sigma}_{\mathrm{BS}}^{(i,N_{\mathrm{UE}})}\right)^2\right]^T$ and $\hat{\mathbf{x}}_{\mathrm{BS}}^{(i)} = \left[\hat{x}_{\mathrm{BS}}^{(i,1)}, \ldots, \hat{x}_{\mathrm{BS}}^{(i,N_{\mathrm{UE}})}\right]^T$. Notably, (3.46) is a standard linear programming (LP) problem and can easily be solved by CVX. Upon obtaining the optimal $\hat{\lambda}_{\mathrm{BS}}^{(i)*}$, the optimal precoder at the BS is found to be

$$\mathbf{W}_{\mathrm{BS}}^{(i)} \left(\mathbf{W}_{\mathrm{BS}}^{(i)}\right)^H = \hat{\mathbf{V}}_{\mathrm{BS}}^{(i,0)} \left(\hat{\mathbf{B}}_{\mathrm{BS}}^{(i)}\right)^{-1} \hat{\mathbf{T}}_{\mathrm{BS}}^{(i)} \hat{\mathbf{\Lambda}}_{\mathrm{BS}}^{(i)*} \left(\hat{\mathbf{T}}_{\mathrm{BS}}^{(i)}\right)^H \left(\hat{\mathbf{B}}_{\mathrm{BS}}^{(i)}\right)^{-1} \left(\hat{\mathbf{V}}_{\mathrm{BS}}^{(i,0)}\right)^H \tag{3.47}$$

where $\hat{\mathbf{B}}_{\mathrm{BS}}^{(i)}$, $\hat{\mathbf{T}}_{\mathrm{BS}}^{(i)}$ and $\hat{\mathbf{V}}_{\mathrm{BS}}^{(i,0)}$ are obtained by replacing all involved $\mathbf{G}_*$ with $\hat{\mathbf{G}}_*$. Thus, the optimal imperfect CSI based $\mathbf{W}_{\mathrm{BS}}$ can be obtained by applying SVD to eq. (3.47). Clearly, constructing an optimal precoder at the BS only requires estimated knowledge about the channels from BS to both MUEs and SUEs, a less stringent requirement relative to the sum-MUE minimization based precoding scheme.

Similarly, the design of precoding vector $\mathbf{W}_{\mathrm{SC}}^{(s)}$ $(s \in \Omega)$ can be handled by solving the LP problem for each SUE, given by

$$\min_{\hat{\lambda}_{\mathrm{SC}}^{(s,j)}} \quad \left(\hat{\mathbf{c}}_{\mathrm{SC}}^{(s,j)}\right)^T \hat{\lambda}_{\mathrm{SC}}^{(s,j)} \tag{3.48a}$$

$$\text{subject to} \quad \mathbf{e}^T \hat{\lambda}_{\mathrm{SC}}^{(s,j)} \leq \gamma_{\mathrm{SC}}^{(s,j)} \tag{3.48b}$$

$$\left(\hat{\mathbf{x}}_{\mathrm{SC}}^{(s,j)}\right)^T \hat{\lambda}_{\mathrm{SC}}^{(s,j)} \leq \alpha_{\mathrm{SC}}^{(s,j)} \tag{3.48c}$$

where $j \in J_s$, $\hat{\lambda}_{\mathrm{SC}}^{(s,j)} = \left[\hat{\lambda}_{\mathrm{SC}}^{(s,j,1)}, \ldots, \hat{\lambda}_{\mathrm{SC}}^{(s,j,N_{\mathrm{UE}})}\right]^T$, $\hat{\mathbf{c}}_{\mathrm{SC}}^{(s,j)} = \left[-\left(\hat{\sigma}_{\mathrm{SC}}^{(s,j,n)}\right)^2, \ldots, -\left(\hat{\sigma}_{\mathrm{SC}}^{(s,j,N_{\mathrm{UE}})}\right)^2\right]^T$ and $\hat{\mathbf{x}}_{\mathrm{SC}}^{(s,j)} = \left[\hat{x}_{\mathrm{SC}}^{(s,j,1)}, \ldots, \hat{x}_{\mathrm{SC}}^{(s,j,N_{\mathrm{UE}})}\right]^T$. The components of the above vectors are calculated in a way entirely similar to that performed in Subsection 3.6.1. Here we omit the details due to limited space. Like the precoder at the BS, only the estimated knowledge about channels from s-th SC to both MUEs and SUEs are required to construct the optimal precoder at s-th SC.

3.8 Simulation Results

Simulations were performed for the three MSE-based precoding strategies in the MIMO HetNet systems to demonstrate the efficiency and performance of the proposed precoder design schemes. In the simulations, the bandwidth was 20 MHz, the cell radiuses for macro-cell and small cell were set to 800 m and 100 m, respectively, and the inter site distance between MC and SC was set to 700 m, see Table 3.1 for simulation parameters and assumption details. Throughout the simulations, a total of 1000 sets of channel realizations were utilized with each set consisting of $(K + L \times S)$ BS-to-UE channels of size $N_{\mathrm{BS}} \times N_{\mathrm{UE}}$ and $S \times (K + L \times S)$ SC-to-UE channels of size $N_{\mathrm{SC}} \times N_{\mathrm{UE}}$, and $10,000$ quadrature-phase-shift keying (QPSK) symbols were transmitted from the BS and each SC node under each channel realization to obtain the BER performance. In all comparisons, unless specified otherwise, the normalized channel estimation error defined by $\bar{\sigma}_h^2 \triangleq \frac{\sigma_h^2}{\sigma_0^2}$ was set to be 1.

Table 3.1: Simulation Parameters

Parameters	Setting
Bandwidth	20 MHz
Cell radius	MC: 800 m, SC: 100 m
Inter site distance	700 m
Transmit power	BS: $46 \sim 56$ dBm, SC: 24 dBm
Noise power density	-174 dBm/Hz
Number of Antennas	$N_{\mathrm{BS}} = 36$, $N_{\mathrm{SC}} = 8$
Number of UEs	$K = 8 \sim 18$, $L = 4$
Pathloss model (BS)	$\theta_{\mathrm{BS}}(d) = 128.1{+}37.6\log_{10}(d)$, d (km) [45]
Pathloss model (SC)	$\theta_{\mathrm{SC}}(d) = 140.7{+}36.7\log_{10}(d)$, d (km) [45]
Penetration loss	$\zeta_{\mathrm{BS}} = \zeta_{\mathrm{SC}} = 20$ dB [45]

Using the proposed sum-MSE based precoding schemes with perfect and imperfect CSI respectively, Fig. 3.2 plots the average MSE learning curves over 100 runs via alternating optimization methods. For comparison purpose, the red lines

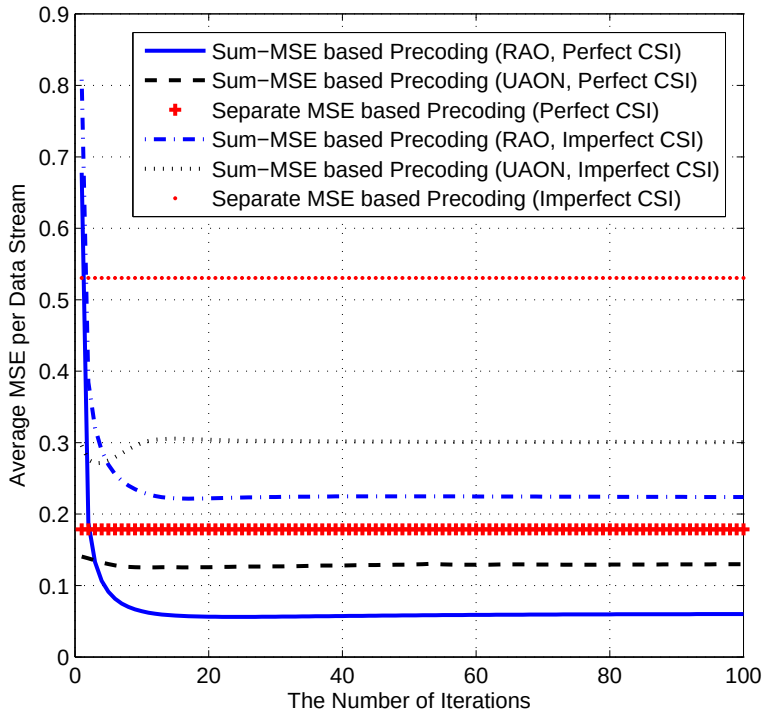

Figure 3.2: The average MSE per data stream learning curve over 100 runs ($K = 8$, $P_{\text{BS}} = 46$ dBm, $\bar{\sigma}_h^2 = 1$).

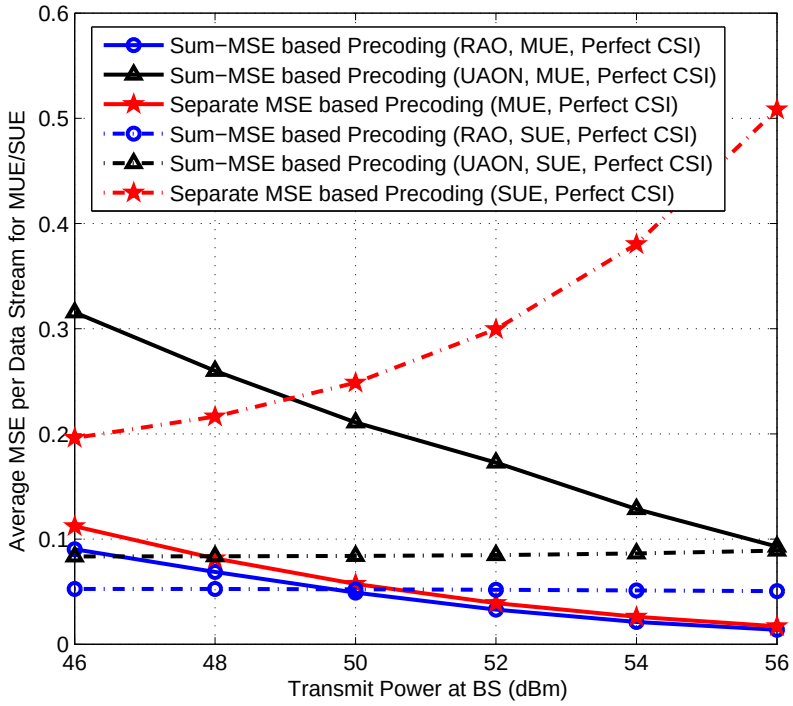

Figure 3.3: The average MSE per data stream for MUE/SUE versus transmit power at BS ($K = 8$, Perfect CSI).

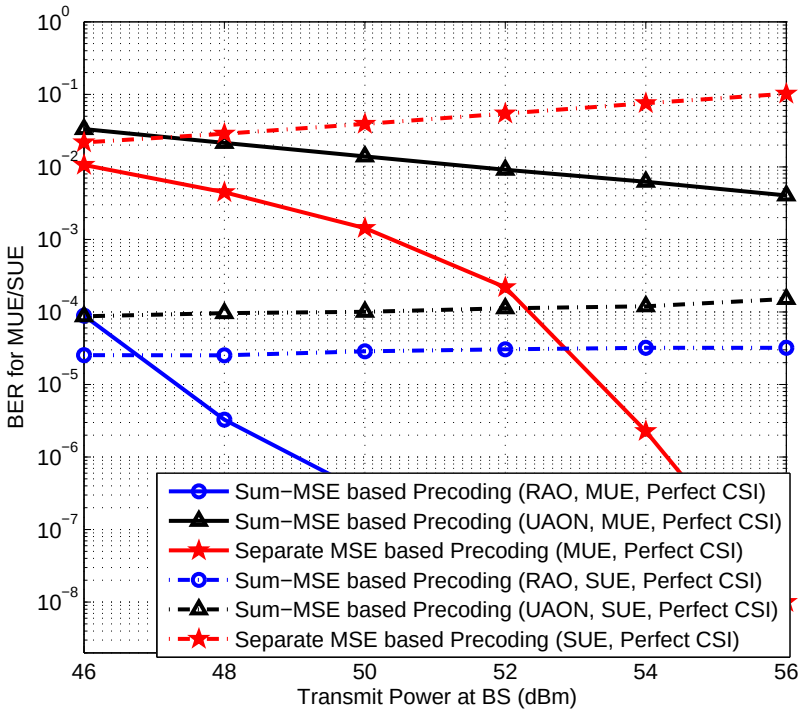

Figure 3.4: The BER per data stream for MUE/SUE versus transmit power at BS ($K = 8$, Perfect CSI).

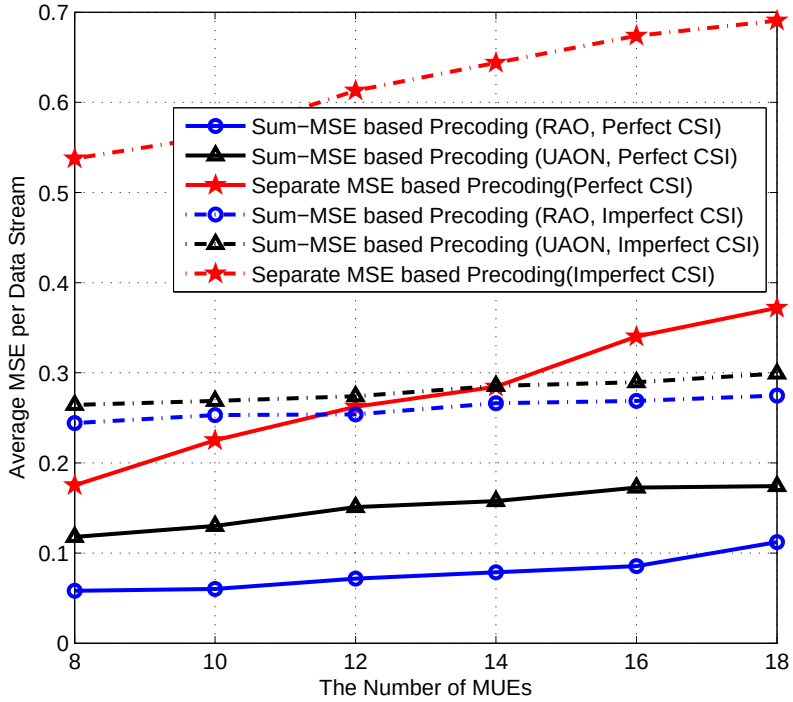

Figure 3.5: The average MSE per data stream versus the number of MUEs K ($P_{\mathrm{BS}} = 46$ dBm, $\bar{\sigma}_h^2 = 1$).

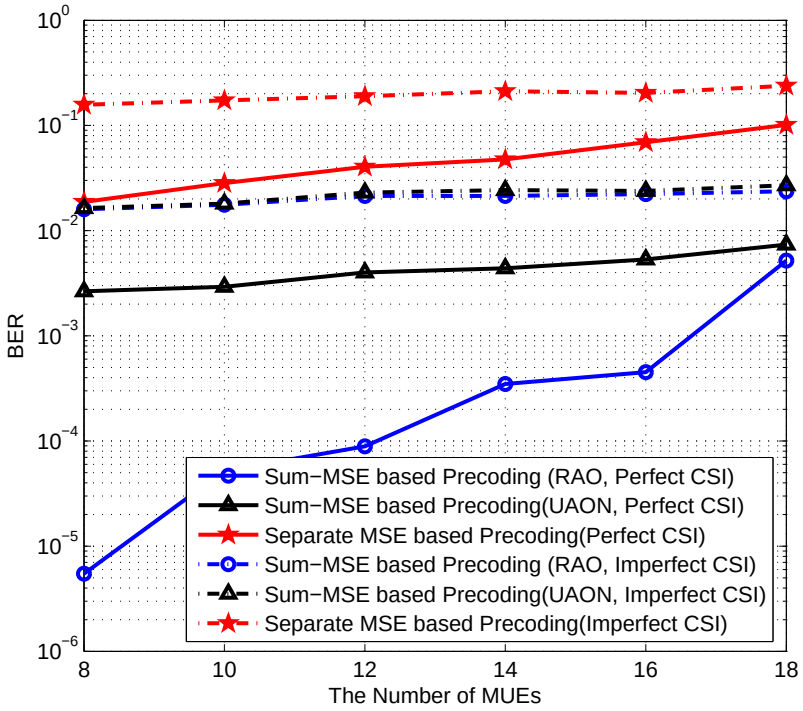

Figure 3.6: The BER per data stream for MUE/SUE versus the number of MUEs K ($P_{\mathrm{BS}} = 46$ dBm, $\bar{\sigma}_h^2 = 1$).

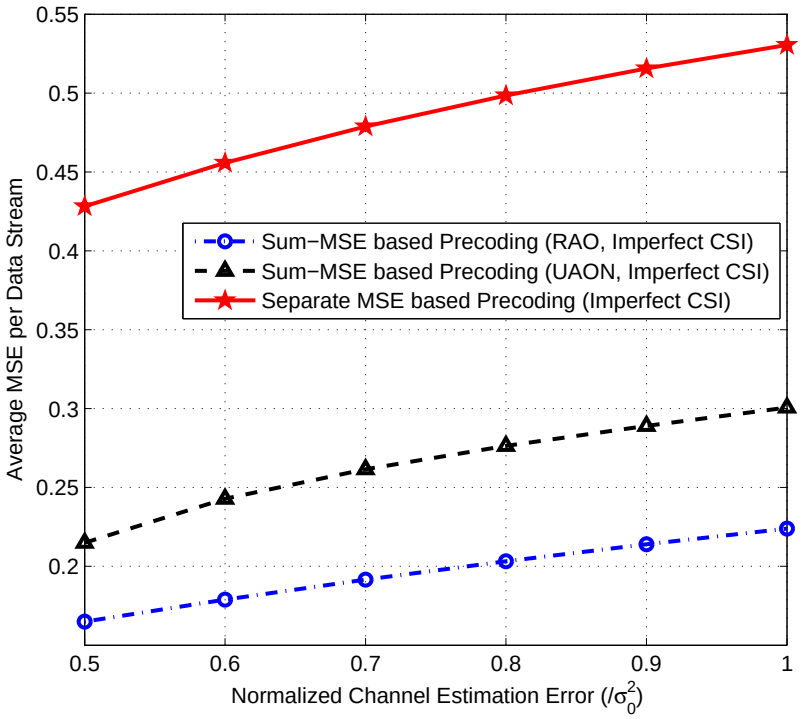

Figure 3.7: The average MSE per data stream versus normalized channel estimation error $\bar{\sigma}_h^2$ ($K = 8$, $P_{\mathrm{BS}} = 46$ dBm, Imperfect CSI).

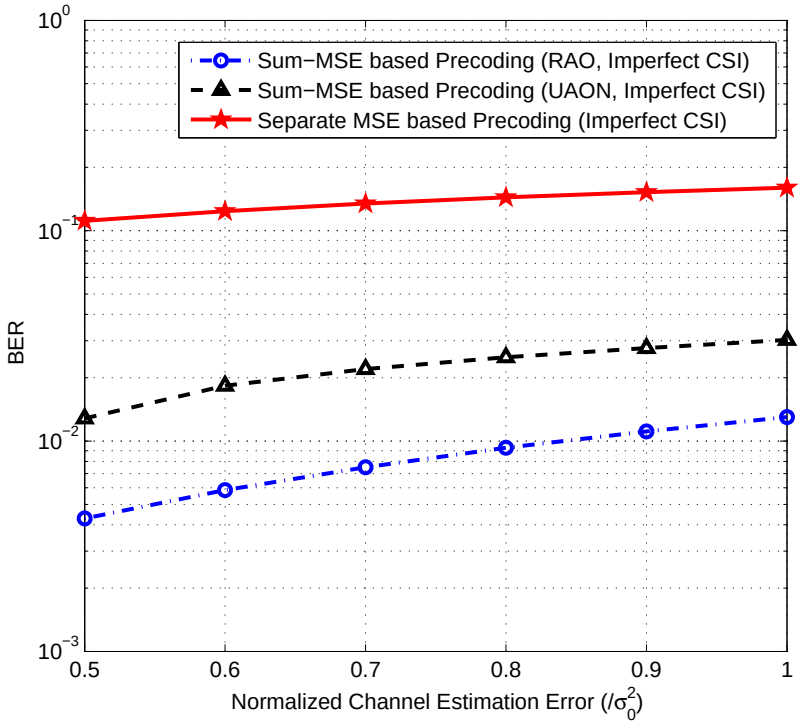

Figure 3.8: The BER per data stream for MUE/SUE versus normalized channel estimation error $\bar{\sigma}_h^2$ ($K = 8$, $P_{\text{BS}} = 46$ dBm, Imperfect CSI).

in Fig. 3.2 depict the average MSE per data stream obtained by the non-iterative algorithm based on separate MSE. From the curves in the figure, it is observed that between the sum-MSE based precoding schemes RAO offers a better performance with a lower average MSE than UAON, but its convergence rate is always slower than the simpler UAON. Also note that the MSE performance of the separate MSE based precoding scheme obtained from the non-iterative algorithm is inferior to that of RAO. Moreover, the performance curves in Fig. 3.2 reveal that when imperfect CSI is utilized, the MSE differences between the separate MSE based and Sum-MSE based precoding are more pronounced relative to those in the perfect CSI case, for all the three proposed schemes, and the convergence of the robust UAON and robust RAO appears to be slower than their perfect CSI counterpart.

Efforts were made to investigate how the average MSE is related to the transmission power. With a fixed $P_{\text{SC}} = 24$ (dBm), Fig. 3.3 shows that the average MSE for MUEs of iterative algorithms decreases gradually as the transmit power at the BS increases with perfect CSI, while the average MSE for SUEs increases slightly due to the increased inter-cell interferences. Furthermore, the sum-MSE based RAO offers the smallest MSE gap between MUE and SUE, indicating better user fairness, while the separate MSE based precoding has the largest one under the

low transmit power. As for the separate MSE based non-iterative precoding scheme, the average MSE for MUEs approaches to that of RAO as the transmit power at BS increases, while the average MSE curve for SUEs goes up gradually. Subsequently, Fig. 3.4 illustrates the corresponding BER performance of the proposed schemes, indicating the same relationships as those of the average MSE performance revealed in Fig. 3.3.

To further illustrate the factors that affect the MSE and BER performance, Fig. 3.5 provides the average MSE curves for the three different precoding schemes when the number of MUEs K increases from 8 to 18 under both perfect and imperfect CSI cases. It can be seen that the average MSE increases as the number of MUEs gets larger, indicating the higher interferences from other MUEs, and that the sum-MSE based RAO always outperforms both the UAON algorithm and the separate MSE based precoding on the MSE performance under different configurations. Also note that the MSE performance gaps between the separate MSE and sum-MSE based schemes become larger as the number of MUEs increases, i.e., the macro-BS has more antennas relative to the number of the MUEs. Furthermore, the BER performances of the proposed three schemes are given in Fig. 3.6, showing a consistent trend with those of Fig. 3.5. We remark that the BER reported here was averaged over all users in the MC and SCs.

In Fig. 3.7, the MSE performance of the three proposed schemes with imperfect CSI are depicted versus the normalized channel estimation error $\bar{\sigma}_h^2$, where $K = 8$ and the transmit power at BS was fixed to $P_{\text{BS}} = 46$ (dBm). From the figure, it is intuitively clear that the average MSE deteriorates as channel estimation error increases. Similarly, the obtained BER curves Fig. 3.8 are consistent to those in Fig. 3.7.

In summary, the sum-MSE based precoding scheme RAO proposed in Section 3.5 outperforms the separate MSE based precoding scheme described in Section 3.6 in terms of the average MSE per user. On the other hand, RAO requires the information of all channels in the HetNet and its superior performance is achieved at the cost of increased computational complexity relative to that of non-iterative separate MSE based precoding. Furthermore, when the macro-BS has a large number of antennas relative to the number of the MUEs, the performance gap between these two schemes shrinks. As a tradeoff algorithm, the sum-MSE based UAON is much simpler and faster than RAO, with a performance slightly better than the separate MSE based scheme in most configurations with reasonable number of BS antennas.

3.9 Conclusion

This chapter has developed three new MSE-based precoding schemes for MIMO downlinks in a HetNet architecture consisting of a macro tier overlaid with a second tier of SCs. The first two are both based on the same sum-MSE minimization problem focusing on the joint design of a set of BS and SC transmit precoding matrices or vectors by minimizing the total user MSE under individual transmit power constraints at each cell. On the other hand, we have also proposed a separate MSE minimization based two-level precoder by a non-iterative algorithm in which BD technique is employed as its first-level precoder and each cell designs its own second-level precoder separately without the need to exchange user data or channel state information over the backhaul. On the basis of the estimated imperfect CSI, corresponding robust precoding schemes have been proposed. Simulation results have shown that the sum-MSE based RAO algorithm always outperforms UAON and the separate MSE-based precoding on the MSE performance. When the number of antennas at the macro-BS is large enough relative to the number of MUEs, the average MSE of the low complexity separate MSE-based precoding can come close to those of RAO and UAON. Furthermore, the UAON algorithm has higher convergence rate and lower computation complexity compared to RAO, thus is a worthy trade-off between efficiency and performance.

3.10 Appendices

3.10.1 Proof of Equation (3.25)

To obtain the non-negative multipliers λ_0 and λ_s ($s \in \Omega$) in the above equations, we substitute (3.22) into (3.20) and write

$$\begin{aligned} L(\lambda) = & -\mathrm{tr}\left\{(\mathbf{\Phi}_{\mathrm{BS}} + \lambda_0 \mathbf{I}_{N_{\mathrm{BS}}})^{-1} \mathbf{G}_{\mathrm{B-M}} \mathbf{R}_{\mathrm{BS}}^H \mathbf{R}_{\mathrm{BS}} \mathbf{G}_{\mathrm{B-M}}^H\right\} - \lambda_0 + \kappa \\ & - \sum_{s=1}^{S} \left[\mathrm{tr}\left\{ \left(\mathbf{\Phi}_{\mathrm{SC}}^{(s)} + \lambda_t \mathbf{I}_{N_{\mathrm{SC}}} \right)^{-1} \mathbf{G}_{\mathrm{S-S}}^{(s,s)} \left(\mathbf{R}_{\mathrm{SC}}^{(s)} \right)^H \mathbf{R}_{\mathrm{SC}}^{(s)} \left(\mathbf{G}_{\mathrm{S-S}}^{(s,s)} \right)^H \right\} + \lambda_s \right] \end{aligned} \tag{3.49}$$

where $\kappa = \sigma_0^2 \mathrm{tr}\left\{\mathbf{R}_{\mathrm{BS}} \mathbf{R}_{\mathrm{BS}}^H\right\} + \sum_{s=1}^{S} \sigma_0^2 \mathrm{tr}\left\{ \mathbf{R}_{\mathrm{SC}}^{(s)} \left(\mathbf{R}_{\mathrm{SC}}^{(s)} \right)^H \right\} + K N_{\mathrm{S}} + \sum_{s=1}^{S} L_s N_{\mathrm{S}}$ is independent of λ. Then, we start from the expressions of $\mathbf{\Phi}_{\mathrm{BS}} = \mathbf{S}_{\mathrm{BS}}^H \mathbf{D}_{\mathrm{BS}} \mathbf{S}_{\mathrm{BS}}$ where $\mathbf{S}_{\mathrm{BS}}^H \mathbf{S}_{\mathrm{BS}} = \mathbf{I}_{N_{\mathrm{BS}}}$ and $\mathbf{D}_{\mathrm{BS}} = \mathrm{diag}\left\{ d_{\mathrm{BS}}^{(1)}, d_{\mathrm{BS}}^{(2)}, \ldots, d_{\mathrm{BS}}^{(N_{\mathrm{BS}})} \right\}$, and $\mathbf{\Phi}_{\mathrm{SC}}^{(s)} = \left(\mathbf{S}_{\mathrm{SC}}^{(s)} \right)^H \mathbf{D}_{\mathrm{SC}}^{(s)} \mathbf{S}_{\mathrm{SC}}^{(s)}$ ($s \in \Omega$) where $\left(\mathbf{S}_{\mathrm{SC}}^{(s)} \right)^H \mathbf{S}_{\mathrm{SC}}^{(s)} = \mathbf{I}_{N_{\mathrm{SC}}}$ and $\mathbf{D}_{\mathrm{SC}}^{(s)} = \mathrm{diag}\left\{ d_{\mathrm{SC}}^{(s,1)}, \ldots, d_{\mathrm{SC}}^{(s,N_{\mathrm{SC}})} \right\}$.

By substituting the above two expressions into (3.49), we obtain

$$L(\lambda) = -\mathrm{tr}\left\{(\mathbf{D}_{\mathrm{BS}} + \lambda_0 \mathbf{I}_{N_{\mathrm{BS}}})^{-1} \mathbf{S}_{\mathrm{BS}}^H \mathbf{G}_{\mathrm{B-M}} \mathbf{R}_{\mathrm{BS}}^H \mathbf{R}_{\mathrm{BS}} \mathbf{G}_{\mathrm{B-M}}^H \mathbf{S}_{\mathrm{BS}}\right\} - \lambda_0 + \kappa$$
$$- \sum_{s=1}^{S} \left[\mathrm{tr}\left\{ \left(\mathbf{D}_{\mathrm{SC}}^{(s)} + \lambda_t \mathbf{I}_{N_{\mathrm{SC}}}\right)^{-1} \left(\mathbf{S}_{\mathrm{SC}}^{(s)}\right)^H \mathbf{G}_{\mathrm{S-S}}^{(s,s)} \left(\mathbf{R}_{\mathrm{SC}}^{(s)}\right)^H \mathbf{R}_{\mathrm{SC}}^{(s)} \left(\mathbf{G}_{\mathrm{S-S}}^{(s,s)}\right)^H \mathbf{S}_{\mathrm{SC}}^{(s)} \right\} + \lambda_s \right]. \tag{3.50}$$

Defining $\mathbf{A}_{\mathrm{BS}} = \mathbf{S}_{\mathrm{BS}}^H \mathbf{G}_{\mathrm{B-M}} \mathbf{R}_{\mathrm{BS}}^H \mathbf{R}_{\mathrm{BS}} \mathbf{G}_{\mathrm{B-M}}^H \mathbf{S}_{\mathrm{BS}}$ with the (n,n)-th entry denoted as $a_{\mathrm{BS}}^{(n)}$, and $\mathbf{A}_{\mathrm{SC}}^{(s)} = \left(\mathbf{S}_{\mathrm{SC}}^{(s)}\right)^H \mathbf{G}_{\mathrm{S-S}}^{(s,s)} \left(\mathbf{R}_{\mathrm{SC}}^{(s)}\right)^H \mathbf{R}_{\mathrm{SC}}^{(s)} \left(\mathbf{G}_{\mathrm{S-S}}^{(s,s)}\right)^H \mathbf{S}_{\mathrm{SC}}^{(s)}$ with the (n,n)-th entry denoted as $a_{\mathrm{SC}}^{(s,n)}$, we have

$$L(\lambda) = -\sum_{n=1}^{N_{\mathrm{BS}}} \frac{a_{\mathrm{BS}}^{(n)}}{d_{\mathrm{BS}}^{(n)} + \lambda_0} - \lambda_0 - \sum_{s=1}^{S} \left(\sum_{n=1}^{N_{\mathrm{SC}}} \frac{a_{\mathrm{SC}}^{(s,n)}}{d_{\mathrm{SC}}^{(s,n)} + \lambda_s} + \lambda_s \right) + \kappa. \tag{3.51}$$

Using (3.51), computing the partial derivative of L w.r.t. λ_0 and λ_s ($s \in \Omega$) becomes straightforward, hence the proof of (3.25).

Chapter 4

Performance Analysis for Pilot-Reused HetNets with Large-Scale Antenna Arrays in Downlink Systems

In the previous chapters, massive multiple-input multiple-output (MIMO) combined with cooperative relaying and the mean square error precoding based heterogeneous networks (HetNets) have been investigated. As variable structure of antenna arrays requires less space now, large-scale antenna arrays set at both base stations and small cells become realizable. In this chapter, the performance analysis for large-scale antennas equipped HetNets with pilot reuse will be done in downlink systems.

4.1 Introduction

As a viable and cost-effective way to increase network capacity, heterogeneous networks (HetNets) that embed a large number of low-power nodes, called small cells (SCs), into an existing macro network has emerged with the aim to offload traffic from the macro cell (MC) to small cells [4, 32–34, 46–48] in hot spots or to solve coverage holes in MC. Conventionally deploying more macro BSs in already dense networks may be prohibitively expensive and result in severe inter-cell interference [34]. However, due to the large number of potentially interfering nodes in the network, mitigating both the inter-cell and intra-cell interference becomes a crucial issue facing HetNet. Interference control has been intensively studied and applied in HetNet [35–37], including the coordinated multi-point (CoMP) trans-

mission [35]. Although the CoMP transmission was shown to provide high spectral efficiency [38] with the backhaul among the coordinated tiers enabling both user data and channel state information (CSI) exchange, the high signaling overhead results in practical implementation limitations.

4.2 Related Work

Recently, multiple-input multiple-output (MIMO) transmission with large-scale antenna arrays at the base station (BS) has attracted substantial interest from both academia and industry. Using simple linear processing, such large-scale antenna arrays were proved to be able to substantially reduce the effects of the uncorrelated noise, small-scale fading and intracell interference [5, 6]. Then, the energy and spectral efficiency of very large multiuser MIMO uplink systems were investigated in [7], which showed that the power radiated by each terminal could be made inversely proportional to either the number of BS antennas or at least its square-root, considering both perfect and imperfect CSI. In [4], it was stated that the potential benefits have elevated large-scale MIMO to a central position as a promising technology for the next generation of wireless systems. In a HetNet setting, [49, 50] proposed to use large scale antenna arrays at the BS and limited antennas at the SCs due to their smaller form factor. As variable structures of antenna arrays, such as a cylindrical array, requires less space [51], large-scale antenna arrays set at SCs becomes realizable. Recently, NEC Corporation announced that it has developed a prototype of A4-sized massive-element Active Antenna System for 5G small cell base stations, and it proposed the use of a massive-element antenna in small cells for capacity enhancement [52]. Up to date, few papers in the literature have studied the effect of employing massive MIMO at SCs. In [53], *HetNet with large-scale antenna arrays* was investigated on downlink performance with interference coordination, and random matrix theory was used to simplify the analysis significantly. It was shown in [54] that using large-scale antenna arrays in SC reduces both intra-tier interference and the cross-tier interference from other nodes in the HetNet system, leading to higher spectral efficiency and better coverage, especially for hot zones.

4.3 Contributions

This chapter presents a comprehensive study of a two-tier network with large-scale antenna arrays set at both BS and SCs. In our preliminary literature [54], maximum-ratio transmission (MRT) precoding was employed based on the esti-

mated channels obtained from the orthogonal training scheme, and downlink capacity lower bounds for a user in the MC and for a user in an SC were derived in closed-form expressions. However, there are still many critical yet unsolved problems. This paper makes the following contributions to address the remaining issues.

1) It was stated in [18, 55] that pilot overhead is proportional to the number of user equipment (UE) for the conventional orthogonal training scheme, i.e., the system performance will degrade as the UE number grows due to heavy pilot overhead. In [5, 6], the pilot reuse (PR) technique is utilized among the macro cells to reduce the pilot overhead, while UEs within a cell use orthogonal pilots. In a two-tier HetNet with multiple small cells, massive antenna arrays and large number of UEs, we propose to apply pilot reuse among the SCs in this chapter, i.e., the same set of orthogonal pilots is reused among the small cells in one macro-cell. Thus the number of orthogonal pilots is smaller than the total UE number in the whole network.

2) We present for the first time the downlink capacity lower bounds of the large-scale HetNet system, where simple linear precoding such as MRT or zero-forcing transmission (ZFT) is employed at each node, followed by detailed asymptotic analysis.

3) The design of an efficient and practical user scheduler for the large-scale HetNet is an important and challenging problem, because the required CSI exchange becomes prohibitively complicated due to the large-scale antenna arrays and the large number of UEs in the MC and SCs. Based on the obtained capacity bounds and asymptotic analysis, a greedy scheduling algorithm (GSA) and an asymptotic scheduling algorithm (ASA) are proposed, respectively, where GSA requires only statistical CSI (SCSI) shared between BS and SCs, and ASA even removes the need for any CSI exchange among nodes.

The rest of the chapter is organized as follows. We briefly describe the system model for HetNet with large-scale antenna arrays in Section 4.4. In Section 4.5, lower bounds for the achievable rate are derived with both imperfect CSI based MRT and ZFT, followed by corresponding asymptotic analysis. Then, two user scheduling algorithms are developed in Section 4.6. Moreover, simulation results under different system configurations are given in Section 4.7 to demonstrate the effectiveness of both the derived rate expressions and the developed schemes. Finally, conclusions are drawn in Section 4.8.

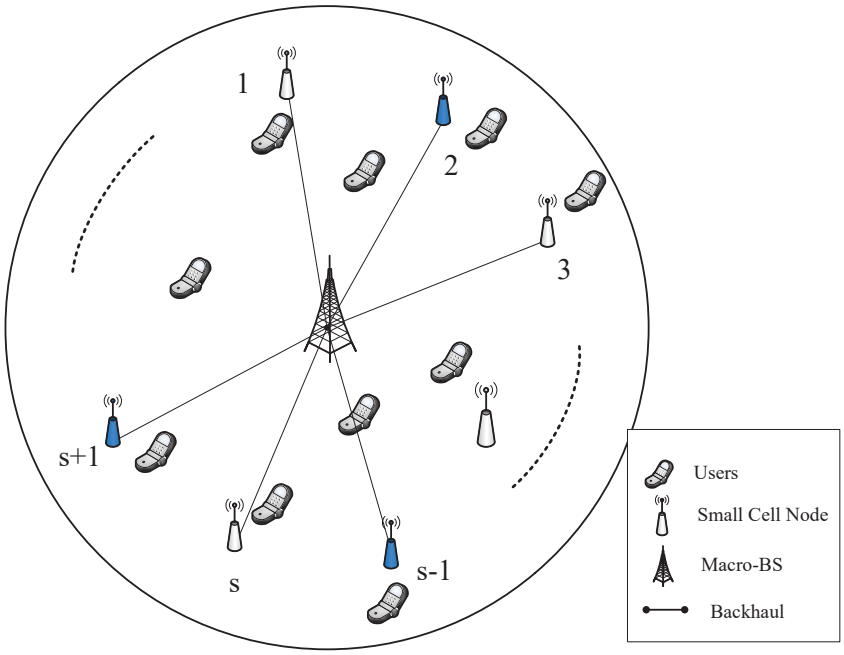

Figure 4.1: System model for HetNet with SCs deployment.

4.4 System Model

Fig. 4.1 shows the considered two-tier network architecture with one cell consisting of one macro BS, which is overlaid with a dense tier of S uniformly distributed SCs by sharing the same time-frequency resources. Assume that the BS and SCs are respectively equipped with large-scale arrays of N_{BS} and N_{SC} antennas, where $N_{\mathrm{BS}} > N_{\mathrm{SC}} \gg 1$, while each user has only one antenna due to the size or complexity constraint. Notably, uniform user distribution in the cell is focused here. Based on the biased user association [48], the users served by the macro BS are designated to a macro UE (MUE) set, and those served by each SC are designated to a small cell UE (SUE) set. Furthermore, suppose that the macro BS serves K MUEs simultaneously while each SC serves L SUEs with $K \leq N_{\mathrm{BS}}$ and $L \leq N_{\mathrm{SC}}$. Denote the MUE and SUE sets as U_{M} and $U_{\mathrm{S}}^{(m)}$ $(m \in \{1, \ldots, S\})$, respectively, then we have $K \leq |U_{\mathrm{M}}|$ and $L \leq \left|U_{\mathrm{S}}^{(m)}\right|$. The selected subsets of MUE and SUE after user scheduling are denoted by I and J_m $(m \in \{1, \ldots, S\})$, respectively.

For the channel matrices, they account for both small-scale fading and large-scale fading. Here, we assume that all the channels between the users and the nodes follow independent and identically distributed (i.i.d.) Rayleigh fading and time division duplex (TDD) is adopted with channel reciprocity satisfied. Denote the channel matrices from the BS and nth $(n \in \{1, \ldots, S\})$ SC to the K MUEs as $\mathbf{G}_{\mathrm{B-M}} = \left[\mathbf{g}_{\mathrm{B-M}}^{(1)}, \ldots, \mathbf{g}_{\mathrm{B-M}}^{(K)}\right] \in \mathbb{C}^{N_{\mathrm{BS}} \times K}$ and $\mathbf{G}_{\mathrm{S-M}}^{(n)} = \left[\mathbf{g}_{\mathrm{S-M}}^{(n,1)}, \ldots, \mathbf{g}_{\mathrm{S-M}}^{(n,K)}\right] \in \mathbb{C}^{N_{\mathrm{SC}} \times K}$, respec-

tively, and use $\mathbf{G}_{\mathrm{B-S}}^{(m)} = \left[\mathbf{g}_{\mathrm{B-S}}^{(m,1)}, \ldots, \mathbf{g}_{\mathrm{B-S}}^{(m,L)}\right] \in \mathbb{C}^{N_{\mathrm{BS}} \times L}$ and $\mathbf{G}_{\mathrm{S-S}}^{(n,m)} = \left[\mathbf{g}_{\mathrm{S-S}}^{(n,m,1)}, \ldots, \mathbf{g}_{\mathrm{S-S}}^{(n,m,L)}\right] \in \mathbb{C}^{N_{\mathrm{SC}} \times L}$ to represent the channel matrices from the BS and the nth SC to the L SUEs in the mth SC, respectively. We have $\mathbf{G}_{\mathrm{B-M}} = \mathbf{H}_{\mathrm{B-M}}\mathbf{D}_{\mathrm{B-M}}^{1/2}$, $\mathbf{G}_{\mathrm{B-S}}^{(m)} = \mathbf{H}_{\mathrm{B-S}}^{(m)}\left(\mathbf{D}_{\mathrm{B-S}}^{(m)}\right)^{1/2}$, $\mathbf{G}_{\mathrm{S-M}}^{(n)} = \mathbf{H}_{\mathrm{S-M}}^{(n)}\left(\mathbf{D}_{\mathrm{S-M}}^{(n)}\right)^{1/2}$ and $\mathbf{G}_{\mathrm{S-S}}^{(n,m)} = \mathbf{H}_{\mathrm{S-S}}^{(n,m)}\left(\mathbf{D}_{\mathrm{S-S}}^{(n,m)}\right)^{1/2}$ where $n,\ m \in \{1, \ldots, S\}$, the first items $\mathbf{H}_{\mathrm{B-M}} \in \mathbb{C}^{N_{\mathrm{BS}} \times K}$, $\mathbf{H}_{\mathrm{B-S}}^{(m)} \in \mathbb{C}^{N_{\mathrm{BS}} \times L}$, $\mathbf{H}_{\mathrm{S-M}}^{(n)} \in \mathbb{C}^{N_{\mathrm{SC}} \times K}$ and $\mathbf{H}_{\mathrm{S-S}}^{(n,m)} \in \mathbb{C}^{N_{\mathrm{SC}} \times L}$ include the i.i.d. $\mathcal{CN}(0,1)$ small-scale fading coefficients, and the second items are the large-scale fading diagonal matrices given by $\mathbf{D}_{\mathrm{B-M}} = \mathrm{diag}\left\{\beta_{\mathrm{B-M}}^{(1)}, \ldots, \beta_{\mathrm{B-M}}^{(K)}\right\}$, $\mathbf{D}_{\mathrm{B-S}}^{(m)} = \mathrm{diag}\left\{\beta_{\mathrm{B-S}}^{(m,1)}, \ldots, \beta_{\mathrm{B-S}}^{(m,L)}\right\}$, $\mathbf{D}_{\mathrm{S-M}}^{(n)} = \mathrm{diag}\left\{\beta_{\mathrm{S-M}}^{(n,1)}, \ldots, \beta_{\mathrm{S-M}}^{(n,K)}\right\}$, and $\mathbf{D}_{\mathrm{S-S}}^{(n,m)} = \mathrm{diag}\left\{\beta_{\mathrm{S-S}}^{(n,m,1)}, \ldots, \beta_{\mathrm{S-S}}^{(n,m,L)}\right\}$.

4.4.1 Channel Estimation with Pilot Reuse

Practically, the channel matrix from each node to its corresponding users, i.e., $\mathbf{G}_{\mathrm{B-M}}$ and $\mathbf{G}_{\mathrm{S-S}}^{(m,m)}$ ($m \in \{1, \ldots, S\}$), have to be estimated based on the uplink training. At the beginning of each coherence interval T, all users simultaneously transmit pilot sequences of length τ symbols. On account of the slight interferences between low-power SCs which are far away from each other, we present a pilot reuse pattern for small cells in a large-scale HetNet system.

First, we denote the reuse factor as γ, i.e., all SCs utilize γ sets of L pairwise orthogonal pilot sequences with a total of $SS = S/\gamma$ SCs sharing the same set. This requires $\tau \geq K + L \times \gamma$ to satisfy the orthogonality of the MC and SC pilot sets. Then, we group SCs into γ sets according to the maximum relative distance criterion and SCs in one set use the same pilot sequences. Since all low-power nodes are modeled as uniformly distributed in a circle with BS at the center as shown in Fig. 4.1, we can denote the nth ($n \in \{1, \cdots, \gamma\}$) SC set as $\mathcal{A}_n = \{n, n+\gamma, \cdots, n+(SS-1)\gamma\}$. Taking $S = 8$ and reuse factor $\gamma = 2$ for example, the SC sets are $\mathcal{A}_1 = \{1, 3, 5, 7\}$ and $\mathcal{A}_2 = \{2, 4, 6, 8\}$. The 4 SCs in each set share one pilot set which includes L pairwise orthogonal pilot sequences, and there are 2 pilot sets for the total of 8 SCs.

Then the training matrix received at the BS and the mth ($m \in \mathcal{A}_r$) SC can be written as

$$\begin{aligned} \mathbf{Y}_{\mathrm{BS}} &= \sqrt{\tau p_\tau}\mathbf{G}_{\mathrm{B-M}}\mathbf{\Phi}_{\mathrm{MUE}} + \sqrt{\tau p_\tau}\sum_{t=1}^{\gamma}\sum_{l \in \mathcal{A}_t}\mathbf{G}_{\mathrm{B-S}}^{(l)}\mathbf{\Phi}_{\mathrm{SUE}}^{(t)} + \mathbf{N}_{\mathrm{BS}} \\ \mathbf{Y}_{\mathrm{SC}}^{(m)} &= \sqrt{\tau p_\tau}\sum_{t=1}^{\gamma}\sum_{l \in \mathcal{A}_t}\mathbf{G}_{\mathrm{S-S}}^{(m,l)}\mathbf{\Phi}_{\mathrm{SUE}}^{(t)} + \sqrt{\tau p_\tau}\mathbf{G}_{\mathrm{S-M}}^{(m)}\mathbf{\Phi}_{\mathrm{MUE}} + \mathbf{N}_{\mathrm{SC}}^{(m)} \end{aligned} \tag{4.1}$$

respectively, where p_τ is the transmit power of each pilot symbol, $\mathbf{N}_{\mathrm{BS}}$ and $\mathbf{N}_{\mathrm{SC}}^{(m)}$ are the additive white Gaussian noise (AWGN) matrices with i.i.d. components following $\mathcal{CN}(0,\sigma_0^2)$, the training vectors transmitted by the ith ($i \in I$) MUE is denoted by the ith row of $\mathbf{\Phi}_{\mathrm{MUE}} \in \mathbb{C}^{K\times\tau}$, satisfying $\mathbf{\Phi}_{\mathrm{MUE}}\mathbf{\Phi}_{\mathrm{MUE}}^H = \mathbf{I}_K$, while the training vector transmitted by the jth ($j \in J_m$) SUE of one SC in the rth set $\mathcal{A}_r$ is represented by the jth row of $\mathbf{\Phi}_{\mathrm{SUE}}^{(r)} \in \mathbb{C}^{L\times\tau}$, satisfying $\mathbf{\Phi}_{\mathrm{SUE}}^{(r)}\left(\mathbf{\Phi}_{\mathrm{SUE}}^{(r)}\right)^H = \mathbf{I}_L$. Moreover, since the rows of pilot sequence matrices are pairwise orthogonal, we have $\mathbf{\Phi}_{\mathrm{MUE}}\left(\mathbf{\Phi}_{\mathrm{SUE}}^{(r)}\right)^H = \mathbf{0}_{K\times L}$ and $\mathbf{\Phi}_{\mathrm{SUE}}^{(r)}\left(\mathbf{\Phi}_{\mathrm{SUE}}^{(t)}\right)^H = \mathbf{0}_{L\times L}$ ($\forall r \neq t \in \{1,\ldots,\gamma\}$).

In order to estimate $\mathbf{G}_{\mathrm{B-M}}$ and $\mathbf{G}_{\mathrm{S-S}}^{(m,m)}$ ($m \in \mathcal{A}_r$), we employ the minimum mean-square-error (MMSE) estimation at each node [22]. The estimated channels are given by

$$\begin{aligned}
\hat{\mathbf{G}}_{\mathrm{B-M}} &= \frac{1}{\sqrt{\tau p_\tau}}\mathbf{Y}_{\mathrm{BS}}\mathbf{\Phi}_{\mathrm{MUE}}^H\tilde{\mathbf{D}}_{\mathrm{B-M}} = \mathbf{G}_{\mathrm{B-M}}\tilde{\mathbf{D}}_{\mathrm{B-M}} + \frac{1}{\sqrt{\tau p_\tau}}\tilde{\mathbf{N}}_{\mathrm{BS}}\tilde{\mathbf{D}}_{\mathrm{B-M}} \\
\hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)} &= \frac{1}{\sqrt{\tau p_\tau}}\mathbf{Y}_{\mathrm{SC}}^{(m)}\left(\mathbf{\Phi}_{\mathrm{SUE}}^{(r)}\right)^H\tilde{\mathbf{D}}_{\mathrm{S-S}}^{(m,m)} = \sum_{l\in\mathcal{A}_r}\mathbf{G}_{\mathrm{S-S}}^{(m,l)}\tilde{\mathbf{D}}_{\mathrm{S-S}}^{(m,m)} + \frac{1}{\sqrt{\tau p_\tau}}\tilde{\mathbf{N}}_{\mathrm{SC}}^{(m)}\tilde{\mathbf{D}}_{\mathrm{S-S}}^{(m,m)}
\end{aligned} \tag{4.2}$$

where

$$\begin{aligned}
\tilde{\mathbf{D}}_{\mathrm{B-M}} &\triangleq \left(\frac{\mathbf{D}_{\mathrm{B-M}}^{-1}\sigma_0^2}{\tau p_\tau} + \mathbf{I}_K\right)^{-1} \\
\tilde{\mathbf{D}}_{\mathrm{S-S}}^{(m,m)} &\triangleq \left[\left(\sum_{l\neq m, l\in\mathcal{A}_r}\mathbf{D}_{\mathrm{S-S}}^{(m,l)} + \frac{\sigma_0^2}{\tau p_\tau}\mathbf{I}_L\right)\left(\mathbf{D}_{\mathrm{S-S}}^{(m,m)}\right)^{-1} + \mathbf{I}_L\right]^{-1} \\
\tilde{\mathbf{N}}_{\mathrm{BS}} &\triangleq \mathbf{N}_{\mathrm{BS}}\mathbf{\Phi}_{\mathrm{MUE}}^H,\ \tilde{\mathbf{N}}_{\mathrm{SC}}^{(m)} \triangleq \mathbf{N}_{\mathrm{SC}}^{(m)}\left(\mathbf{\Phi}_{\mathrm{SUE}}^{(r)}\right)^H
\end{aligned} \tag{4.3}$$

with $m \in \mathcal{A}_r$. Due to the property of $\mathbf{\Phi}_{\mathrm{MUE}}$ and $\mathbf{\Phi}_{\mathrm{SUE}}^{(r)}$, $\tilde{\mathbf{N}}_{\mathrm{BS}}$ and $\tilde{\mathbf{N}}_{\mathrm{SC}}^{(m)}$ are also composed of i.i.d. $\mathcal{CN}(0,\sigma_0^2)$ elements. Then, we have

$$\begin{aligned}
\mathbf{G}_{\mathrm{B-M}} &= \hat{\mathbf{G}}_{\mathrm{B-M}} + \mathbf{\Xi}_{\mathrm{B-M}} \\
\mathbf{G}_{\mathrm{S-S}}^{(m,m)} &= \hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)} + \mathbf{\Xi}_{\mathrm{S-S}}^{(m,m)}
\end{aligned} \tag{4.4}$$

where $\mathbf{\Xi}_{\mathrm{B-M}}$ and $\mathbf{\Xi}_{\mathrm{S-S}}^{(m,m)}$ denote the estimation error matrices which are independent of $\hat{\mathbf{G}}_{\mathrm{B-M}}$ and $\hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)}$ from the property of MMSE channel estimation [22]. Hence, we have $\hat{\mathbf{G}}_{\mathrm{B-M}} \sim CN\left(0,\hat{\mathbf{D}}_{\mathrm{B-M}}\right)$ with $\hat{\mathbf{D}}_{\mathrm{B-M}} = \mathrm{diag}\left\{\hat{\beta}_{\mathrm{B-M}}^{(1)},\ldots,\hat{\beta}_{\mathrm{B-M}}^{(K)}\right\}$, $\hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)} \sim CN\left(0,\hat{\mathbf{D}}_{\mathrm{S-S}}^{(m,m)}\right)$ with $\hat{\mathbf{D}}_{\mathrm{S-S}}^{(m,m)} = \mathrm{diag}\left\{\hat{\beta}_{\mathrm{S-S}}^{(m,m,1)},\ldots,\hat{\beta}_{\mathrm{S-S}}^{(m,m,L)}\right\}$, $\mathbf{\Xi}_{\mathrm{B-M}} \sim CN\left(0,\mathbf{D}_{\mathrm{B-M}} - \hat{\mathbf{D}}_{\mathrm{B-M}}\right)$ with the ith column vector denoted by $\xi_{\mathrm{B-M}}^{(i)}$, and $\mathbf{\Xi}_{\mathrm{S-S}}^{(m,m)} \sim$

$CN\left(0, \mathbf{D}_{\mathrm{S-S}}^{(m,m)} - \hat{\mathbf{D}}_{\mathrm{S-S}}^{(m,m)}\right)$ with the jth column vector denoted by $\xi_{\mathrm{S-S}}^{(m,m,j)}$. Here, the estimated large-scale fading factors satisfy $\hat{\beta}_{\mathrm{B-M}}^{(i)} = \frac{\tau p_\tau \left(\beta_{\mathrm{B-M}}^{(i)}\right)^2}{\tau p_\tau \beta_{\mathrm{B-M}}^{(i)} + \sigma_0^2}$, $i \in I$ and $\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)} = \frac{\tau p_\tau \left(\beta_{\mathrm{S-S}}^{(m,m,j)}\right)^2}{\tau p_\tau \sum\limits_{l \in \mathcal{A}_r} \beta_{\mathrm{S-S}}^{(m,l,j)} + \sigma_0^2}$, where $j \in J_m$ and $m \in \mathcal{A}_r$.

4.4.2 Data Transmission

In the downlinks, the received signals at K MUEs and L SUEs in the mth small cell are

$$\mathbf{y}_{\mathrm{M}} = \mathbf{G}_{\mathrm{B-M}}^T \mathbf{W}_{\mathrm{BS}} \mathbf{x}_{\mathrm{BS}} + \sum_{n=1}^{S} \left(\mathbf{G}_{\mathrm{S-M}}^{(n)}\right)^T \mathbf{W}_{\mathrm{SC}}^{(n)} \mathbf{x}_{\mathrm{SC}}^{(n)} + \mathbf{n}_{\mathrm{M}} \tag{4.5}$$

$$\mathbf{y}_{\mathrm{S}}^{(m)} = \left(\mathbf{G}_{\mathrm{B-S}}^{(m)}\right)^T \mathbf{W}_{\mathrm{BS}} \mathbf{x}_{\mathrm{BS}} + \sum_{n=1}^{S} \left(\mathbf{G}_{\mathrm{S-S}}^{(n,m)}\right)^T \mathbf{W}_{\mathrm{SC}}^{(n)} \mathbf{x}_{\mathrm{SC}}^{(n)} + \mathbf{n}_{\mathrm{S}}^{(m)} \tag{4.6}$$

respectively, where $\mathbf{W}_{\mathrm{BS}}$ and $\mathbf{W}_{\mathrm{SC}}^{(n)}$ represent the linear precoding matrices at the BS and nth SC, respectively; $\mathbf{x}_{\mathrm{BS}} = \left[x_{\mathrm{BS}}^{(1)}, \ldots, x_{\mathrm{BS}}^{(K)}\right]^T$ and $\mathbf{x}_{\mathrm{SC}}^{(n)} = \left[x_{\mathrm{SC}}^{(n,1)}, \ldots, x_{\mathrm{SC}}^{(n,L)}\right]^T$ are the complex-valued data symbols from BS to its MUEs and from nth SC to its own SUEs, respectively, satisfying $\mathrm{E}\left[\mathbf{x}_{\mathrm{BS}} \mathbf{x}_{\mathrm{BS}}{}^H\right] = \mathbf{I}_{N_{\mathrm{BS}}}$ and $\mathrm{E}\left[\mathbf{x}_{\mathrm{SC}}^{(n)} \left(\mathbf{x}_{\mathrm{SC}}^{(n)}\right)^H\right] = \mathbf{I}_{N_{\mathrm{SC}}}$; and $\mathbf{n}_{\mathrm{M}} = \left[n_{\mathrm{M}}^{(1)}, \ldots, n_{\mathrm{M}}^{(K)}\right]^T$ and $\mathbf{n}_{\mathrm{S}}^{(m)} = \left[n_{\mathrm{S}}^{(m,1)}, \ldots, n_{\mathrm{S}}^{(m,L)}\right]^T$ involves the AWGN of variance σ_0^2.

4.4.3 MRT Precoding

Aiming to maximize the received signal-to-noise ratio, the MRT technique is utilized at both the BS and SCs to process the transmit signals towards the corresponding users. Given the estimated channel state information, the MRT precoding is expressed as [13]

$$\mathbf{W}_{\mathrm{BS}} = \alpha_{\mathrm{BS}} \hat{\mathbf{G}}_{\mathrm{B-M}}^*, \ \mathbf{W}_{\mathrm{SC}}^{(m)} = \alpha_{\mathrm{SC}}^{(m)} \left(\hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)}\right)^* \tag{4.7}$$

where α_{BS} and $\alpha_{\mathrm{SC}}^{(m)}$ are normalization constants, chosen to satisfy the transmit power constraints at the BS and SCs, respectively. On the basis of (4.7) and $\mathrm{Tr}\{\mathbf{AB}\} = \mathrm{Tr}\{\mathbf{BA}\}$, we have

$$\alpha_{\mathrm{BS}} = \sqrt{\frac{p_{\mathrm{BS}}}{N_{\mathrm{BS}} \Phi_{\mathrm{B-M}}}}, \ \alpha_{\mathrm{SC}}^{(m)} = \sqrt{\frac{p_{\mathrm{SC}}^{(m)}}{N_{\mathrm{SC}} \Phi_{\mathrm{S-S}}^{(m)}}} \tag{4.8}$$

where $\Phi_{\mathrm{B-M}} = \sum_{i=1}^{K} \hat{\beta}_{\mathrm{B-M}}^{(i)}$, and $\Phi_{\mathrm{S-S}}^{(m)} = \sum_{l=1}^{L} \hat{\beta}_{\mathrm{S-S}}^{(m,m,l)}$ with $m \in \{1, \ldots, S\}$.

4.4.4 ZFT Precoding

Likewise, when ZFT is employed based on imperfect CSI, in which the pseudo-inverse of the estimated channels in (4.4) are utilized for linear precoding, the precoder is given by [13]

$$\begin{aligned} \mathbf{W}_{\mathrm{BS}} &= \alpha_{\mathrm{BS}} \hat{\mathbf{G}}_{\mathrm{B-M}}^{*} \left(\hat{\mathbf{G}}_{\mathrm{B-M}}^{T} \hat{\mathbf{G}}_{\mathrm{B-M}}^{*} \right)^{-1} = \alpha_{\mathrm{BS}} \hat{\bar{\mathbf{G}}}_{\mathrm{B-M}}^{*} \\ \mathbf{W}_{\mathrm{SC}}^{(m)} &= \alpha_{\mathrm{SC}}^{(m)} \left(\hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)} \right)^{*} \left[\left(\hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)} \right)^{T} \left(\hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)} \right)^{*} \right]^{-1} = \alpha_{\mathrm{SC}}^{(m)} \left(\hat{\bar{\mathbf{G}}}_{\mathrm{S-S}}^{(m,m)} \right)^{*} \end{aligned} \tag{4.9}$$

where $\hat{\bar{\mathbf{G}}}_{\mathrm{B-M}} = \hat{\mathbf{G}}_{\mathrm{B-M}} \left(\hat{\mathbf{G}}_{\mathrm{B-M}}^{H} \hat{\mathbf{G}}_{\mathrm{B-M}} \right)^{-1}$, $\hat{\bar{\mathbf{G}}}_{\mathrm{S-S}}^{(m,m)} = \hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)} \left[\left(\hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)} \right)^{H} \hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)} \right]^{-1}$, α_{BS} and $\alpha_{\mathrm{SC}}^{(m)}$ are normalization constants. Similarly, based on (4.9) and $\mathrm{Tr}\{\mathbf{AB}\} = \mathrm{Tr}\{\mathbf{BA}\}$, we have

$$\alpha_{\mathrm{BS}} = \sqrt{\frac{(N_{\mathrm{BS}} - K - 1)\, p_{\mathrm{BS}}}{\Psi_{\mathrm{B-M}}}}, \quad \alpha_{\mathrm{SC}}^{(m)} = \sqrt{\frac{(N_{\mathrm{SC}} - L - 1)\, p_{\mathrm{SC}}^{(m)}}{\Psi_{\mathrm{S-S}}^{(m)}}} \tag{4.10}$$

where $\Psi_{\mathrm{B-M}} = \sum_{i=1}^{K} \frac{1}{\hat{\beta}_{\mathrm{B-M}}^{(i)}}$, and $\Psi_{\mathrm{S-S}}^{(m)} = \sum_{l=1}^{L} \frac{1}{\hat{\beta}_{\mathrm{S-S}}^{(m,m,l)}}$ with $m \in \{1, \ldots, S\}$. The detailed derivation is given in Appendix 4.9.1.

4.5 Achievable Rate Analysis

The exact rate analysis of the MUE and SUE in the pilot assisted massive MIMO heterogeneous network considered is highly complicated and intractable. In this section, we provide a closed-form capacity lower bound of each user for both MRT and ZFT precoding, respectively. The simple lower bounds can be applied to user scheduling and power allocation optimization as detailed in subsequent sections.

4.5.1 MRT Precoding

In practice, only imperfect CSI derived from transmitted pilots is available at each node for linear precoding. Utilizing MRT precoding in (4.5) and (4.6), the received signal of the ith ($i \in I$) MUE and jth ($j \in J_m$) SUE at the mth ($m \in$

$\{1,\ldots,S\}$) SC can be rewritten as

$$y_{\mathrm{M}}^{(i)} = \underbrace{\alpha_{\mathrm{BS}}\left(\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right)^{T}\left(\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right)^{*}x_{\mathrm{BS}}^{(i)}}_{\text{desired signal}} + \underbrace{\alpha_{\mathrm{BS}}\left(\xi_{\mathrm{B-M}}^{(i)}\right)^{T}\left(\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right)^{*}x_{\mathrm{BS}}^{(i)}}_{\text{estimation error induced interference}}$$
$$+ \underbrace{\sum_{k\neq i}^{K}\alpha_{\mathrm{BS}}\left(\mathbf{g}_{\mathrm{B-M}}^{(i)}\right)^{T}\left(\hat{\mathbf{g}}_{\mathrm{B-M}}^{(k)}\right)^{*}x_{\mathrm{BS}}^{(k)}}_{\text{intra-MC interference}} + \underbrace{\sum_{n=1}^{S}\sum_{l=1}^{L}\alpha_{\mathrm{SC}}^{(n)}\left(\mathbf{g}_{\mathrm{S-M}}^{(n,i)}\right)^{T}\left(\hat{\mathbf{g}}_{\mathrm{S-S}}^{(n,n,l)}\right)^{*}x_{\mathrm{SC}}^{(n,l)}}_{\text{cross-tier interference}} + \underbrace{n_{\mathrm{M}}^{(i)}}_{\text{noise at MUE}} \tag{4.11}$$

$$y_{\mathrm{S}}^{(m,j)} = \underbrace{\alpha_{\mathrm{SC}}^{(m)}\left(\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right)^{T}\left(\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right)^{*}x_{\mathrm{SC}}^{(m,j)}}_{\text{desired signal}} + \underbrace{\alpha_{\mathrm{SC}}^{(m)}\left(\xi_{\mathrm{S-S}}^{(m,m,j)}\right)^{T}\left(\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right)^{*}x_{\mathrm{SC}}^{(m,j)}}_{\text{estimation error induced interference}}$$
$$+ \underbrace{\sum_{i=1}^{K}\alpha_{\mathrm{BS}}\left(\mathbf{g}_{\mathrm{B-S}}^{(m,j)}\right)^{T}\left(\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right)^{*}x_{\mathrm{BS}}^{(i)}}_{\text{cross-tier interference}} + \underbrace{\sum_{l_1\neq j}^{L}\alpha_{\mathrm{SC}}^{(m)}\left(\mathbf{g}_{\mathrm{S-S}}^{(m,m,j)}\right)^{T}\left(\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,l_1)}\right)^{*}x_{\mathrm{SC}}^{(m,l_1)}}_{\text{intra-SC interference}}$$
$$+ \underbrace{\sum_{n\neq m}^{S}\sum_{l_2=1}^{L}\alpha_{\mathrm{SC}}^{(n)}\left(\mathbf{g}_{\mathrm{S-S}}^{(n,m,j)}\right)^{T}\left(\hat{\mathbf{g}}_{\mathrm{S-S}}^{(n,n,l_2)}\right)^{*}x_{\mathrm{SC}}^{(n,l_2)}}_{\text{inter-SC interference}} + \underbrace{n_{\mathrm{S}}^{(m,j)}}_{\text{noise at SUE}} . \tag{4.12}$$

Note that both the BS and small cell nodes treat the estimated channels as the true channels [7], and the first term is the desired signal. The remaining terms are considered as interferences and noise, including estimation error caused interference term. Accordingly, with imperfect CSI, the ergodic achievable rate of MUE i ($i \in I$) and SUE j ($j \in J_m$) in the mth ($m \in \{1,\ldots,S\}$) SC are given by

$$R_{\mathrm{M}}^{(i)} = \mathrm{E}\left[\log_2\left(1 + \frac{\alpha_{\mathrm{BS}}^2\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^4}{\mathrm{EEI}_i + \mathrm{IMI}_i + \mathrm{CTI}_i + \sigma_0^2}\right)\right] \tag{4.13}$$

$$R_{\mathrm{S}}^{(m,j)} = \mathrm{E}\left[\log_2\left(1 + \frac{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2\left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^4}{\mathrm{EEI}_{m,j} + \mathrm{CTI}_{m,j} + \mathrm{ISI}_{m,j} + \mathrm{SSI}_{m,j} + \sigma_0^2}\right)\right] \tag{4.14}$$

respectively, where EEI_i, IMI_i and CTI_i denote the estimation error induced interference, the intra-MC interference and the cross-tier interference for the ith MUE,

respectively, given by

$$\begin{aligned}\mathrm{EEI}_i =&\alpha_{\mathrm{BS}}^2\left|\left(\xi_{\mathrm{B-M}}^{(i)}\right)^H\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right|^2,\ \mathrm{IMI}_i=\sum_{k\neq i}^{K}\alpha_{\mathrm{BS}}^2\left|\left(\mathbf{g}_{\mathrm{B-M}}^{(i)}\right)^H\hat{\mathbf{g}}_{\mathrm{B-M}}^{(k)}\right|^2\\ \mathrm{CTI}_i=&\sum_{n=1}^{S}\sum_{l=1}^{L}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\left|\left(\mathbf{g}_{\mathrm{S-M}}^{(n,i)}\right)^H\hat{\mathbf{g}}_{\mathrm{S-S}}^{(n,n,l)}\right|^2,\end{aligned} \tag{4.15}$$

and $\mathrm{EEI}_{m,j}$, $\mathrm{CTI}_{m,j}$, $\mathrm{ISI}_{m,j}$ and $\mathrm{SSI}_{m,j}$ denote the estimation error induced interference, the cross-tier interference, the intra-SC interference and the inter-SC interference for the jth SUE in the mth SC, respectively, given by

$$\begin{aligned}\mathrm{EEI}_{m,j}=&\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2\left|\left(\xi_{\mathrm{S-S}}^{(m,m,j)}\right)^H\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right|^2,\ \mathrm{CTI}_{m,j}=\sum_{i=1}^{K}\alpha_{\mathrm{BS}}^2\left|\left(\mathbf{g}_{\mathrm{B-S}}^{(m,j)}\right)^H\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right|^2\\ \mathrm{ISI}_{m,j}=&\sum_{l_1\neq j}^{L}\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2\left|\left(\mathbf{g}_{\mathrm{S-S}}^{(m,m,j)}\right)^H\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,l_1)}\right|^2\\ \mathrm{SSI}_{m,j}=&\sum_{n\neq m}^{S}\sum_{l_2=1}^{L}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\left|\left(\mathbf{g}_{\mathrm{S-S}}^{(n,m,j)}\right)^H\hat{\mathbf{g}}_{\mathrm{S-S}}^{(n,n,l_2)}\right|^2.\end{aligned} \tag{4.16}$$

Notably, the inter-SC interference $\mathrm{SSI}_{m,j}$ includes the pilot contamination effect caused by pilot reuse. In the above achievable rate expressions, expectations over the estimated instantaneous CSI cannot be further derived into tractable forms. Therefore, we adopt a similar bounding technique of [7] to obtain closed form rate expressions, the result of which will provide insights on the impact of different system parameters and facilitate further optimizations.

By the convexity of $\log_2\left(1+\frac{1}{x}\right)$ and Jensen's inequality, from (4.13) and (4.14), a lower bound on the achievable rate is obtained as

$$R_{0,\mathrm{M,M}}^{(i)}=\log_2\left(1+\left(\mathrm{E}\left[\frac{\mathrm{EEI}_i+\mathrm{IMI}_i+\mathrm{CTI}_i+\sigma_0^2}{\alpha_{\mathrm{BS}}^2\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^4}\right]\right)^{-1}\right) \tag{4.17}$$

$$R_{0,\mathrm{S,M}}^{(m,j)}=\log_2\left(1+\left(\mathrm{E}\left[\frac{\mathrm{EEI}_{m,j}+\mathrm{CTI}_{m,j}+\mathrm{ISI}_{m,j}+\mathrm{SSI}_{m,j}+\sigma_0^2}{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2\left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^4}\right]\right)^{-1}\right). \tag{4.18}$$

Theorem 4.1: With imperfect CSI based MRT, $N_{\mathrm{BS}}\geq 2$ and $N_{\mathrm{SC}}\geq 2$, the downlink achievable rate of the ith ($i\in I$) MUE and jth ($j\in J_m$) SUE in the

mth ($m \in \{1, \ldots, S\}$) SC, for finite N_{BS} and N_{SC}, are lower bounded by

$$R_{0,\mathrm{M,M}}^{(i)} = \log_2 \left(1 + \frac{a_{\mathrm{MR}}^{(i)} p_{\mathrm{BS}}}{b_{\mathrm{MR}}^{(i)} p_{\mathrm{BS}} + \sum_{n=1}^{S} c_{\mathrm{MR}}^{(n,i)} p_{\mathrm{SC}}^{(n)} + \sigma_0^2} \right) \tag{4.19}$$

$$R_{0,\mathrm{S,M}}^{(m,j)} = \log_2 \left(1 + \frac{d_{\mathrm{MR}}^{(m,j)} p_{\mathrm{SC}}^{(m)}}{\sum_{n=1}^{S} e_{\mathrm{MR}}^{(n,m,j)} p_{\mathrm{SC}}^{(n)} + f_{\mathrm{MR}}^{(m,j)} p_{\mathrm{BS}} + \sigma_0^2} \right) \tag{4.20}$$

where

$$\begin{aligned}
a_{\mathrm{M}}^{(i)} &= \frac{(N_{\mathrm{BS}}-1)(N_{\mathrm{BS}}-2)\left(\hat{\beta}_{\mathrm{B-M}}^{(i)}\right)^2}{N_{\mathrm{BS}}\Phi_{\mathrm{B-M}}}, \; c_{\mathrm{M}}^{(n,i)} = \beta_{\mathrm{S-M}}^{(n,i)} \\
b_{\mathrm{M}}^{(i)} &= \beta_{\mathrm{B-M}}^{(i)} - \frac{2}{N_{\mathrm{BS}}}\hat{\beta}_{\mathrm{B-M}}^{(i)} - \frac{(N_{\mathrm{BS}}-4)\left(\hat{\beta}_{\mathrm{B-M}}^{(i)}\right)^2 + 2\beta_{\mathrm{B-M}}^{(i)}\hat{\beta}_{\mathrm{B-M}}^{(i)}}{N_{\mathrm{BS}}\Phi_{\mathrm{B-M}}} \\
d_{\mathrm{M}}^{(m,j)} &= \frac{(N_{\mathrm{SC}}-1)(N_{\mathrm{SC}}-2)\left(\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}\right)^2}{N_{\mathrm{SC}}\Phi_{\mathrm{S-S}}^{(m)}}, \; f_{\mathrm{M}}^{(m,j)} = \beta_{\mathrm{B-S}}^{(m,j)} \\
e_{\mathrm{M}}^{(n,m,j)} &= \begin{cases}
\beta_{\mathrm{S-S}}^{(n,m,j)}, \; n \neq m, \; n \notin \mathcal{A}_r \\
\frac{N_{\mathrm{SC}}\left(\beta_{\mathrm{S-S}}^{(n,m,j)}\right)^2\left(\hat{\beta}_{\mathrm{S-S}}^{(n,n,j)}\right)^2}{\left(\beta_{\mathrm{S-S}}^{(n,n,j)}\right)^2\Phi_{\mathrm{S-S}}^{(n)}} + \beta_{\mathrm{S-S}}^{(n,m,j)}, \; n \neq m, \; n \in \mathcal{A}_r \\
\beta_{\mathrm{S-S}}^{(m,m,j)} - \frac{2}{N_{\mathrm{SC}}}\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)} - \frac{(N_{\mathrm{SC}}-4)\left(\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}\right)^2 + 2\beta_{\mathrm{S-S}}^{(m,m,j)}\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}}{N_{\mathrm{SC}}\Phi_{\mathrm{S-S}}^{(m)}}, \; n = m.
\end{cases}
\end{aligned} \tag{4.21}$$

Proof: See Appendix 4.9.2.

Remark 4.1: Since the proof in Appendix 4.9.2 does not use any asymptotic assumptions on the antenna size, Theorem 1 is also valid for conventional scale MIMO systems. The capacity lower bounds for perfect CSI can be obtained by setting $\hat{\beta}_{\cdot}^{(\cdot)} = \beta_{\cdot}^{(\cdot)}$ in (4.19) and (4.20). Moreover, it can be observed from Appendix 4.9.2 that all the interferences contained in the received signals of the MUEs (i.e., EEI, IMI and CTI) can be significantly mitigated relative to the desired signals by increasing N_{BS}. Similarly, all the interference effect at the SUEs (i.e., EEI, CTI, ISI and SSI) is able to be reduced by increasing N_{SC}. These observations support the use of large scale antenna arrays at both BS and SC. In addition, the expression of $R_{0,\mathrm{M,M}}^{(i)}$ indicates that the MUE rate increases monotonically with N_{BS} but has no relationship with N_{SC}. Similarly, the SUE rate increases monotonically with N_{SC} and is independent of N_{BS}.

4.5.2 ZFT Precoding

For imperfect CSI based ZFT precoding, the received signal can be rewritten as

$$y_{\mathrm{M}}^{(i)} = \underbrace{\alpha_{\mathrm{BS}} x_{\mathrm{BS}}^{(i)}}_{\text{desired signal}} + \underbrace{\alpha_{\mathrm{BS}} \left(\xi_{\mathrm{B-M}}^{(i)}\right)^{T} \left(\hat{\tilde{\mathbf{g}}}_{\mathrm{B-M}}^{(i)}\right)^{*} x_{\mathrm{BS}}^{(i)}}_{\text{estimation error induced interference}} + \underbrace{\sum_{k \neq i}^{K} \alpha_{\mathrm{BS}} \left(\xi_{\mathrm{B-M}}^{(i)}\right)^{T} \left(\hat{\tilde{\mathbf{g}}}_{\mathrm{B-M}}^{(k)}\right)^{*} x_{\mathrm{BS}}^{(k)}}_{\text{intra-MC interference}}$$
$$+ \underbrace{\sum_{n=1}^{S} \sum_{l=1}^{L} \alpha_{\mathrm{SC}}^{(n)} \left(\mathbf{g}_{\mathrm{S-M}}^{(n,i)}\right)^{T} \left(\hat{\tilde{\mathbf{g}}}_{\mathrm{S-S}}^{(n,n,l)}\right)^{*} x_{\mathrm{SC}}^{(n,l)}}_{\text{cross-tier interference}} + \underbrace{n_{\mathrm{M}}^{(i)}}_{\text{noise at MUE}} \tag{4.22}$$

$$y_{\mathrm{S}}^{(m,j)} = \underbrace{\alpha_{\mathrm{SC}}^{(m)} x_{\mathrm{SC}}^{(m,j)}}_{\text{desired signal}} + \underbrace{\alpha_{\mathrm{SC}}^{(m)} \left(\xi_{\mathrm{S-S}}^{(m,m,j)}\right)^{T} \left(\hat{\tilde{\mathbf{g}}}_{\mathrm{S-S}}^{(m,m,j)}\right)^{*} x_{\mathrm{SC}}^{(m,j)}}_{\text{estimation error induced interference}}$$
$$+ \underbrace{\sum_{i=1}^{K} \alpha_{\mathrm{BS}} \left(\mathbf{g}_{\mathrm{B-S}}^{(m,j)}\right)^{T} \left(\hat{\tilde{\mathbf{g}}}_{\mathrm{B-M}}^{(i)}\right)^{*} x_{\mathrm{BS}}^{(i)}}_{\text{cross-tier interference}} + \underbrace{\sum_{l_1 \neq j}^{L} \alpha_{\mathrm{SC}}^{(m)} \left(\xi_{\mathrm{S-S}}^{(m,m,j)}\right)^{T} \left(\hat{\tilde{\mathbf{g}}}_{\mathrm{S-S}}^{(m,m,l_1)}\right)^{*} x_{\mathrm{SC}}^{(m,l_1)}}_{\text{intra-SC interference}}$$
$$+ \underbrace{\sum_{n \neq m}^{S} \sum_{l_2=1}^{L} \alpha_{\mathrm{SC}}^{(n)} \left(\mathbf{g}_{\mathrm{S-S}}^{(n,m,j)}\right)^{T} \left(\hat{\tilde{\mathbf{g}}}_{\mathrm{S-S}}^{(n,n,l_2)}\right)^{*} x_{\mathrm{SC}}^{(n,l_2)}}_{\text{inter-SC interference}} + \underbrace{n_{\mathrm{S}}^{(m,j)}}_{\text{noise at SUE}} \tag{4.23}$$

respectively, where $\left(\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right)^{T} \left(\hat{\tilde{\mathbf{g}}}_{\mathrm{B-M}}^{(i)}\right)^{*} = 1$, $\left(\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right)^{T} \left(\hat{\tilde{\mathbf{g}}}_{\mathrm{S-S}}^{(m,m,j)}\right)^{*} = 1$, and the intra-MC interference IMI_i and intra-SC interference $\mathrm{ISI}_{m,j}$ are reduced because ZFT precoding is able to null multi-user interference signals, i.e., $\left(\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right)^{T} \left(\hat{\tilde{\mathbf{g}}}_{\mathrm{B-M}}^{(k)}\right)^{*} = 0$ and $\left(\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right)^{T} \left(\hat{\tilde{\mathbf{g}}}_{\mathrm{S-S}}^{(m,m,l)}\right)^{*} = 0$ for $\forall k \neq i$ and $\forall l \neq j$. Similarly, by the convexity of $\log_2\left(1 + \frac{1}{x}\right)$ and Jensen's inequality, capacity lower bounds of MUE i ($i \in I$) and SUE j ($j \in J_m$) in the mth ($m \in \{1, \ldots, S\}$) SC in (4.17) and (4.18) become

$$R_{0,\mathrm{M,Z}}^{(i)} = \log_2 \left(1 + \left(\mathrm{E}\left[\frac{\mathrm{EEI}_i + \mathrm{IMI}_i + \mathrm{CTI}_i + \sigma_0^2}{\alpha_{\mathrm{BS}}^2}\right]\right)^{-1}\right) \tag{4.24}$$

$$R_{0,\mathrm{S,Z}}^{(m,j)} = \log_2 \left(1 + \left(\mathrm{E}\left[\frac{\mathrm{EEI}_{m,j} + \mathrm{CTI}_{m,j} + \mathrm{ISI}_{m,j} + \mathrm{SSI}_{m,j} + \sigma_0^2}{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2}\right]\right)^{-1}\right). \tag{4.25}$$

where

$$\begin{aligned}\mathrm{EEI}_i =&\alpha_{\mathrm{BS}}^2\left|\left(\xi_{\mathrm{B-M}}^{(i)}\right)^H\hat{\tilde{\mathbf{g}}}_{\mathrm{B-M}}^{(i)}\right|^2,\ \mathrm{IMI}_i=\sum_{k\neq i}^{K}\alpha_{\mathrm{BS}}^2\left|\left(\xi_{\mathrm{B-M}}^{(i)}\right)^H\hat{\tilde{\mathbf{g}}}_{\mathrm{B-M}}^{(k)}\right|^2\\ \mathrm{CTI}_i=&\sum_{n=1}^{S}\sum_{l=1}^{L}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\left|\left(\mathbf{g}_{\mathrm{S-M}}^{(n,i)}\right)^H\hat{\tilde{\mathbf{g}}}_{\mathrm{S-S}}^{(n,n,l)}\right|^2\end{aligned} \tag{4.26}$$

$$\begin{aligned}&\mathrm{EEI}_{m,j}=\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2\left|\left(\xi_{\mathrm{S-S}}^{(m,m,j)}\right)^H\hat{\tilde{\mathbf{g}}}_{\mathrm{S-S}}^{(m,m,j)}\right|^2,\ \mathrm{CTI}_{m,j}=\sum_{i=1}^{K}\alpha_{\mathrm{BS}}^2\left|\left(\mathbf{g}_{\mathrm{B-S}}^{(m,j)}\right)^H\hat{\tilde{\mathbf{g}}}_{\mathrm{B-M}}^{(i)}\right|^2\\ &\mathrm{ISI}_{m,j}=\sum_{l_1\neq j}^{L}\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2\left|\left(\xi_{\mathrm{S-S}}^{(m,m,j)}\right)^H\hat{\tilde{\mathbf{g}}}_{\mathrm{S-S}}^{(m,m,l_1)}\right|^2,\ \mathrm{SSI}_{m,j}=\sum_{n\neq m}^{S}\sum_{l_2=1}^{L}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\left|\left(\mathbf{g}_{\mathrm{S-S}}^{(n,m,j)}\right)^H\hat{\tilde{\mathbf{g}}}_{\mathrm{S-S}}^{(n,n,l_2)}\right|^2.\end{aligned} \tag{4.27}$$

Theorem 4.2: With imperfect CSI based ZFT, $N_{\mathrm{BS}}\geq 2$ and $N_{\mathrm{SC}}\geq 2$, the downlink achievable rate of the ith ($i\in I$) MUE and jth ($j\in J_m$) SUE in the mth ($m\in\{1,\ldots,S\}$) SC, for finite N_{BS} and N_{SC}, are lower bounded by

$$R_{0,\mathrm{M,Z}}^{(i)}=\log_2\left(1+\frac{a_{\mathrm{ZF}}^{(i)}p_{\mathrm{BS}}}{b_{\mathrm{ZF}}^{(i)}p_{\mathrm{BS}}+\sum_{n=1}^{S}c_{\mathrm{ZF}}^{(n,i)}p_{\mathrm{SC}}^{(n)}+\sigma_0^2}\right) \tag{4.28}$$

$$R_{0,\mathrm{S,Z}}^{(m,j)}=\log_2\left(1+\frac{d_{\mathrm{ZF}}^{(m,j)}p_{\mathrm{SC}}^{(m)}}{\sum_{n=1}^{S}e_{\mathrm{ZF}}^{(n,m,j)}p_{\mathrm{SC}}^{(n)}+f_{\mathrm{ZF}}^{(m,j)}p_{\mathrm{BS}}+\sigma_0^2}\right) \tag{4.29}$$

where

$$\begin{aligned}&a_{\mathrm{Z}}^{(i)}=\frac{N_{\mathrm{BS}}-K-1}{\Psi_{\mathrm{B-M}}},\ b_{\mathrm{Z}}^{(i)}=\xi_{\mathrm{B-M}}^{(i)},\ c_{\mathrm{Z}}^{(n,i)}=\beta_{\mathrm{S-M}}^{(n,i)},\ d_{\mathrm{Z}}^{(m,j)}=\frac{N_{\mathrm{SC}}-L-1}{\Psi_{\mathrm{S-S}}^{(m)}}\\ &e_{\mathrm{Z}}^{(n,m,j)}=\begin{cases}\beta_{\mathrm{S-S}}^{(n,m,j)},\ n\neq m,\ n\notin\mathcal{A}_r\\ \frac{\beta_{\mathrm{S-S}}^{(n,m,j)}\left(\Psi_{\mathrm{S-S}}^{(n)}-\frac{1}{\hat{\beta}_{\mathrm{S-S}}^{(n,n,j)}}\right)}{\Psi_{\mathrm{S-S}}^{(n)}}+\frac{(N_{\mathrm{SC}}-L-1)\left(\beta_{\mathrm{S-S}}^{(n,m,j)}\right)^2}{\Psi_{\mathrm{S-S}}^{(n)}\left(\beta_{\mathrm{S-S}}^{(n,n,j)}\right)^2},\ n\neq m,\ n\in\mathcal{A}_r\\ \xi_{\mathrm{S-S}}^{(m,m,j)},\ n=m\end{cases}\\ &f_{\mathrm{Z}}^{(m,j)}=\beta_{\mathrm{B-S}}^{(m,j)}.\end{aligned} \tag{4.30}$$

Proof: See Appendix 4.9.3.

Remark 4.2: Similar to MRT, conclusions in Remark 4.1 are also valid for ZFT based Theorem 4.2. Moreover, the capacity lower bounds in (4.28) and (4.29) indicate that $R_{0,\mathrm{M,ZF}}^{(i)}$ decreases monotonically as the estimation error $\xi_{\mathrm{B-M}}^{(i)}$ increases

for fixed $\Psi_{\mathrm{B-M}}$. Similar conclusions are drawn for the capacity lower bounds of SUEs.

Remark 4.3: From (4.28), we can conclude that the expression of $R_{0,\mathrm{M,Z}}^{(i)}$ for the ith MUE involves only N_{BS}, but no N_{SC}, which indicates that the capacity lower bound of MUE depends on N_{BS} but has no relationship with N_{SC}. Furthermore, the signal-to-interference-plus-noise ratio (SINR) involved in $R_{0,\mathrm{M,Z}}^{(i)}$ is approximately a linearly increasing function of N_{BS}, when $N_{\mathrm{BS}} \gg K$. Similarly, from (4.29), it can be concluded that the capacity lower bound of SUE increases monotonically with N_{SC} but is independent of N_{BS}. However, due to the pilot contamination effect caused by pilot reuse, the increase of the SINR in $R_{0,\mathrm{S,Z}}^{(m,j)}$ is not linear even when $N_{\mathrm{SC}} \gg L$.

4.5.3 Asymptotic Analysis with Massive Arrays

Having obtained the closed-form expressions for the achievable rate in (4.19) and (4.20), this subsection provides the asymptotic analysis under two different cases when the number of antennas approaches infinity. Suppose that all SCs have the same transmit power, i.e., $p_{SC}^{(1)} = \cdots = p_{SC}^{(S)} = p_{\mathrm{SC}}$, and $N_{\mathrm{BS}} = \lambda N_{\mathrm{SC}}$ with $\lambda \geq 10$.

Proposition 4.1: In case I where p_τ is fixed, $p_{\mathrm{SC}}^{(s)} = p_{\mathrm{SC}} = \frac{E_{\mathrm{SC}}}{N_{\mathrm{SC}}^{\chi_1}}$ $(s = 1, \cdots, S)$, $p_{\mathrm{BS}} = \frac{E_{\mathrm{BS}}}{N_{\mathrm{BS}}^{\eta_1}}$, and E_{SC} and E_{BS} are fixed, to achieve non-vanishing user rate as $N_{\mathrm{SC}} \to \infty$ with $N_{\mathrm{BS}} = \lambda N_{\mathrm{SC}}$, the SC and BS transmit power scaling factors χ_1 and η_1 must satisfy $0 \leq \chi_1 \leq 1$ and $0 \leq \eta_1 \leq 1$. When $\chi_1 = \eta_1 = 1$, the asymptotic achievable rate expressions of the ith $(i \in I)$ MUE and jth $(j \in J_m)$ SUE in the mth $(m \in \{1, \ldots, S\})$ SC for imperfect CSI based MRT and ZFT are

$$
\begin{aligned}
R_{0,\mathrm{M,M}}^{(i)} &\xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1 + \frac{E_{\mathrm{BS}}\left(\hat{\beta}_{\mathrm{B-M}}^{(i)}\right)^2}{\Phi_{\mathrm{B-M}}\sigma_0^2}\right) \\
R_{0,\mathrm{S,M}}^{(m,j)} &\xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1 + \frac{E_{\mathrm{SC}}\left(\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}\right)^2 / \Phi_{\mathrm{S-S}}^{(m)}}{\sigma_0^2 + \sum\limits_{\substack{n\neq m \\ n\in\mathcal{A}_r}} \frac{E_{\mathrm{SC}}\left(\hat{\beta}_{\mathrm{S-S}}^{(n,n,j)}\beta_{\mathrm{S-S}}^{(n,m,j)}\right)^2}{\Phi_{\mathrm{S-S}}^{(n)}\left(\beta_{\mathrm{S-S}}^{(n,n,j)}\right)^2}}\right)
\end{aligned}
\tag{4.31}
$$

$$R_{0,\mathrm{M,Z}}^{(i)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{E_{\mathrm{BS}}}{\Psi_{\mathrm{B-M}}\sigma_0^2}\right), R_{0,\mathrm{S,Z}}^{(m,j)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{E_{\mathrm{SC}}/\Psi_{\mathrm{S-S}}^{(m)}}{\sigma_0^2+\sum\limits_{\substack{n\neq m\\ n\in\mathcal{A}_r}}\frac{E_{\mathrm{SC}}\left(\beta_{\mathrm{S-S}}^{(n,m,j)}\right)^2}{\Psi_{\mathrm{S-S}}^{(n)}\left(\beta_{\mathrm{S-S}}^{(n,n,j)}\right)^2}}\right) \tag{4.32}$$

respectively, which show that the transmit powers at both BS and SCs can be scaled down by up to $\frac{1}{N_{\mathrm{SC}}}$ to maintain a given rate in case I. When $0 \leq \chi_1 < 1$ and $0 \leq \eta_1 < 1$, the asymptotic achievable rate of each user approaches to infinity as $N_{\mathrm{SC}} \to \infty$.

Remark 4.4: Obviously, when the pilot reuse factor $\gamma = S$, i.e., no pilot reuse, we have $R_{0,\mathrm{S,M}}^{(m,j)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{E_{\mathrm{SC}}\left(\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}\right)^2}{\Phi_{\mathrm{S-S}}^{(m)}\sigma_0^2}\right)$ and $R_{0,\mathrm{S,Z}}^{(m,j)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{E_{\mathrm{SC}}}{\Psi_{\mathrm{S-S}}^{(m)}\sigma_0^2}\right)$. For practical system configurations, we suppose that $\chi_1 \geq \eta_1$ in case I to guarantee $p_{\mathrm{SC}} < p_{\mathrm{BS}}$. When $0 < \chi_1 < 1$ and $0 < \eta_1 < 1$, the asymptotic achievable rate of both MUEs and SUEs approaches to infinity as $N_{\mathrm{SC}} \to \infty$ for both MRT and ZFT. When $\chi_1 = 1$ and $\eta_1 = 0$, we have

$$\begin{aligned}
&R_{0,\mathrm{M,M}}^{(i)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \infty,\ R_{0,\mathrm{M,Z}}^{(i)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \infty \\
&R_{0,\mathrm{S,M}}^{(m,j)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{E_{\mathrm{SC}}\left(\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}\right)^2/\Phi_{\mathrm{S-S}}^{(m)}}{\beta_{\mathrm{B-S}}^{(m,j)}E_{\mathrm{BS}}+\sum\limits_{\substack{n\neq m\\ n\in\mathcal{A}_r}}\frac{E_{\mathrm{SC}}\left(\hat{\beta}_{\mathrm{S-S}}^{(n,n,j)}\beta_{\mathrm{S-S}}^{(n,m,j)}\right)^2}{\left(\Phi_{\mathrm{S-S}}^{(n)}\left(\beta_{\mathrm{S-S}}^{(n,n,j)}\right)^2\right)}+\sigma_0^2}\right) \\
&R_{0,\mathrm{S,Z}}^{(m,j)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{E_{\mathrm{SC}}/\Psi_{\mathrm{S-S}}^{(m)}}{\beta_{\mathrm{B-S}}^{(m,j)}E_{\mathrm{BS}}+\sum\limits_{\substack{n\neq m\\ n\in\mathcal{A}_r}}\frac{E_{\mathrm{SC}}\left(\beta_{\mathrm{S-S}}^{(n,m,j)}\right)^2}{\left(\Psi_{\mathrm{S-S}}^{(n)}\left(\beta_{\mathrm{S-S}}^{(n,n,j)}\right)^2\right)}+\sigma_0^2}\right)
\end{aligned} \tag{4.33}$$

respectively, indicating that the cross-tier interferences at SUEs can not be eliminated when p_{SC} is scaled down proportionally to $\frac{1}{N_{\mathrm{SC}}}$ with fixed p_{BS} in case I.

Proposition 4.2: In case II where $p_\tau = \frac{E_\tau}{N_{\mathrm{SC}}^\theta}$, $p_{\mathrm{SC}}^{(s)} = p_{\mathrm{SC}} = \frac{E_{\mathrm{SC}}}{N_{\mathrm{SC}}^{\chi_2}}$ $(s = 1, \cdots, S)$, $p_{\mathrm{BS}} = \frac{E_{\mathrm{BS}}}{N_{\mathrm{BS}}^{\eta_2}}$, and E_τ, E_{BS} and E_{SC} are fixed, to achieve non-vanishing user rate as $N_{\mathrm{SC}} \to \infty$ with $N_{\mathrm{BS}} = \lambda N_{\mathrm{SC}}$ and the pilot reuse factor $\gamma = S$, the pilot, SC and BS transmit power scaling factors θ, χ_2 and η_2 must satisfy $0 < \theta \leq 1$, $0 \leq \chi_2 \leq 1-\theta$ and $0 \leq \eta_2 \leq 1-\theta$. When $0 < \theta < 1$ and $\chi_2 = \eta_2 = 1-\theta$, the asymptotic achievable rate expressions of the ith $(i \in I)$ MUE and jth $(j \in J_m)$ SUE in the

mth ($m \in \{1, \ldots, S\}$) SC for imperfect CSI based MRT and ZFT are

$$\begin{aligned} R_{0,\mathrm{M,M}}^{(i)} &\xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{\lambda^{\theta}\tau E_{\tau}E_{\mathrm{BS}}\left(\beta_{\mathrm{B-M}}^{(i)}\right)^4}{\sum\limits_{k=1}^{K}\left(\beta_{\mathrm{B-M}}^{(k)}\right)^2\sigma_0^4}\right) \\ R_{0,\mathrm{S,M}}^{(m,j)} &\xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{\tau E_{\tau}E_{\mathrm{SC}}\left(\beta_{\mathrm{S-S}}^{(m,m,j)}\right)^4}{\sum\limits_{l=1}^{L}\left(\beta_{\mathrm{S-S}}^{(m,m,l)}\right)^2\sigma_0^4}\right) \end{aligned} \tag{4.34}$$

$$R_{0,\mathrm{M,Z}}^{(i)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{\lambda^{\theta}\tau E_{\tau}E_{\mathrm{BS}}}{\sum\limits_{k=1}^{K}\left(\beta_{\mathrm{B-M}}^{(k)}\right)^{-2}\sigma_0^4}\right), R_{0,\mathrm{S,Z}}^{(m,j)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{\tau E_{\tau}E_{\mathrm{SC}}}{\sum\limits_{l=1}^{L}\left(\beta_{\mathrm{S-S}}^{(m,m,l)}\right)^{-2}\sigma_0^4}\right) \tag{4.35}$$

respectively, from which we conclude that the transmit powers of BS and SCs can only be reduced by up to $\frac{1}{N_{\mathrm{SC}}^{1-\theta}}$ with the pilot transmit power set as $p_{\tau} = \frac{E_{\tau}}{N_{\mathrm{SC}}^{\theta}}$ and pilot reuse factor $\gamma = S$ (no pilot reuse) in case II. When $0 < \theta < 1$, $0 \leq \chi_2 < 1-\theta$ and $0 \leq \eta_2 < 1-\theta$, the asymptotic achievable rate of each user approaches to infinity as $N_{\mathrm{SC}} \to \infty$.

Remark 4.5: To guarantee MUE and SUE achievable rate, $\chi_2 = \eta_2 = 0$ should be satisfied in case II when $\theta = 1$, which means that the pilot power can be scaled down by up to $\frac{1}{N_{\mathrm{SC}}}$ with fixed transmit power at both BS and SC nodes. Then, the asymptotic achievable rate can be expressed as

$$\begin{aligned} R_{0,\mathrm{M,M}}^{(i)} &\xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{\lambda\tau E_{\tau}E_{\mathrm{BS}}\left(\beta_{\mathrm{B-M}}^{(i)}\right)^4}{\sum\limits_{k=1}^{K}\left(\beta_{\mathrm{B-M}}^{(k)}\right)^2\sigma_0^2\left(\mathrm{RD}_i+\sigma_0^2\right)}\right) \\ R_{0,\mathrm{S,M}}^{(m,j)} &\xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{\tau E_{\tau}E_{\mathrm{SC}}\left(\beta_{\mathrm{S-S}}^{(m,m,j)}\right)^4}{\sum\limits_{l=1}^{L}\left(\beta_{\mathrm{S-S}}^{(m,m,l)}\right)^2\sigma_0^2\left(\mathrm{RD}_{m,j}+\sigma_0^2\right)}\right) \end{aligned}$$

$$R_{0,\mathrm{M,Z}}^{(i)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{\lambda\tau E_{\tau}\frac{E_{\mathrm{BS}}}{\sum\limits_{k=1}^{K}\left(\beta_{\mathrm{B-M}}^{(k)}\right)^{-2}}}{\sigma_0^2\left(\mathrm{RD}_i+\sigma_0^2\right)}\right), R_{0,\mathrm{S,Z}}^{(m,j)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} \log_2\left(1+\frac{\tau E_{\tau}\frac{E_{\mathrm{SC}}}{\sum\limits_{l=1}^{L}\left(\beta_{\mathrm{S-S}}^{(m,m,l)}\right)^{-2}}}{\sigma_0^2\left(\mathrm{RD}_{m,j}+\sigma_0^2\right)}\right) \tag{4.36}$$

for MRT and ZFT, respectively, where the residual items are $\mathrm{RD}_i = E_{\mathrm{BS}}\beta_{\mathrm{B-M}}^{(i)} + E_{\mathrm{SC}}\sum_{n=1}^{S}\beta_{\mathrm{S-M}}^{(n,i)}$ and $\mathrm{RD}_{m,j} = E_{\mathrm{SC}}\sum_{n=1}^{M}\beta_{\mathrm{S-S}}^{(n,m,l)} + E_{\mathrm{BS}}\beta_{\mathrm{B-S}}^{(m,j)}$. It indicates that the channel estimation error induced interference, cross-tier and inter-SC interferences can not be eliminated when p_τ is scaled down proportionally to $\frac{1}{N_{\mathrm{SC}}}$ with fixed p_{BS} and p_{SC} in case II.

Proposition 4.3: In case II as stated in Proposition 2, to achieve non-vanishing user rate as $N_{\mathrm{SC}} \to \infty$ with $N_{\mathrm{BS}} = \lambda N_{\mathrm{SC}}$ and the pilot reuse power $\gamma < S$, i.e., considering the pilot reuse introduced contamination, the pilot power scaling factor must satisfy $\theta = 0$. If $\theta > 0$, the MUE rate $R_{0,\mathrm{M,M}}^{(i)}$ and $R_{0,\mathrm{M,Z}}^{(i)}$ still follow (4.34) and (4.35), while the asymptotic achievable rate expressions of the jth ($j \in J_m$) SUE in the mth ($m \in \{1,\ldots,S\}$) SC for imperfect CSI based MRT and ZFT are

$$R_{0,\mathrm{S,M}}^{(m,j)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} 0, \quad R_{0,\mathrm{S,Z}}^{(m,j)} \xrightarrow[N_{\mathrm{SC}}\to\infty]{a.s.} 0. \tag{4.37}$$

4.6 User Scheduling Algorithms

To maximize the sum rate of the scheduled MUEs and SUEs, exhaustive search in the whole MUE and SUE sets is one possible method to obtain optimal results. However, it is not practical since it has rather low searching speed with high complexity. In this section, $\gamma = S$, i.e., no pilot reuse, is assumed[1]. As a traditional suboptimal method, a greedy scheduling algorithm is proposed according to the derived capacity lower bounds in Section 4.5. Then, in comparison to the greedy scheduling algorithm, we propose a much simpler scheduling algorithm based on the obtained asymptotic results to maximize each cell's sum rate supposing that $N_{\mathrm{SC}} \to \infty$. It is called asymptotic scheduling algorithm which significantly reduces the computation complexity.

4.6.1 Greedy Scheduling Algorithm (GSA)

Aiming to obtain an optimal user scheduling algorithm, we formulate an optimization problem considering the maximization of total achievable rate for all the scheduled MUEs and SUEs, subject to the constraints on the scale of each UE set,

[1]For other pilot reuse factors, the corresponding user scheduling algorithm can be a similar way.

i.e.,

$$\max_{I\subseteq U_{\mathrm{M}},J_m\subseteq U_{\mathrm{S}}^{(m)}} R_{\mathrm{SUM}}\left(I,J_1,\ldots,J_S\right)\triangleq\frac{T-\tau}{T}\left(\sum_{i\in I}R_{0,\mathrm{M}}^{(i)}+\sum_{m=1}^{S}\sum_{j\in J_m}R_{0,\mathrm{S}}^{(m,j)}\right) \quad (4.38a)$$

$$\text{s.t.}\quad |I|=K,\ \ |J_m|=L,\ m=1,\ldots,S \quad (4.38b)$$

which can surely be solved by inefficient exhaustive search. To reduce the computation complexity, a greedy scheduling algorithm is proposed, as summarized in Algorithm 4.1. In one iteration, each cell schedules one additional user, the user that maximizes the total sum rate is added in one cell, given the scheduled users in all other cells. This process repeats until the number of scheduled users in each cell reaches the target value.

As shown above, the proposed GSA is a suboptimal solution for the maximization of the sum rate, but it still results in high computation complexity, since each user's achievable rate is determined by the SCSI of the downlink channels from the BS and all SC nodes to the users, i.e., the BS and each SC node should share SCSI with one another even for each cell's own user scheduling.

Algorithm 4.1: Greedy scheduling algorithm

Initialization: $N=\left\lfloor\frac{K}{L}\right\rfloor$, $I=\emptyset$, $J_m=\emptyset$, $\tilde{U}_{\mathrm{M}}=U_{\mathrm{M}}$ and $\tilde{U}_{\mathrm{S}}^{(m)}=U_{\mathrm{S}}^{(m)}$ for $m=1,\ldots,S$.
Repeat:
For1 $k=1$ to N
$i^*=\arg\max_{i\in\tilde{U}_{\mathrm{M}}} R_{\mathrm{SUM}}\left(I\cup\{i\},J_1,\ldots,J_S\right)$
$I=I\cup\{i^*\}$, $\tilde{U}_{\mathrm{M}}=\tilde{U}_{\mathrm{M}}\backslash\{i^*\}$
Endfor1
For2 $m=1$ to S
$j_m^*=\arg\max_{j_m\in\tilde{U}_{\mathrm{S}}^{(m)}} R_{\mathrm{SUM}}\left(I,J_1\ldots,J_m\cup\{j_m\}\ldots,J_S\right)$
$J_m=J_m\cup\{j_m^*\}$, $\tilde{U}_{\mathrm{S}}^{(m)}=\tilde{U}_{\mathrm{S}}^{(m)}\backslash\{j_m^*\}$
Endfor2
Until: If $|J_m|=L$ with $m=1,\ldots,S$, change $N=K$ mod (L), go through For1 Loop and then stop;
Output: Output I and J_m $(m=1,\ldots,S)$ as the solutions.

4.6.2 Asymptotic Scheduling Algorithm (ASA)

In order to further reduce the computation complexity of the traditional GSA, a new algorithm named ASA is presented for MRT and ZFT, respectively, based on

the asymptotic results as given in Subsection 4.5.3. From Propositions 1 and 2, it can be concluded that there is no inter-cell interference when $N_{SC} \to \infty$, i.e., each cell is able to do user scheduling according to its own statistical CSI and there is no information exchange requirement among the BS and SC nodes any more.

Asymptotic Scheduling Algorithm for MRT (ASA-M):

Since the Propositions 4.1 and 4.2 provide the asymptotic results for MRT, we define the achievable rate of the ith ($i \in U_{\mathrm{BS}}$) MUE and jth ($j \in U_{\mathrm{SC}}^{(m)}$) SUE in the mth ($m = 1, \ldots, S$) SC as $R_{0,\mathrm{M,M-AS}}^{(i)}$ and $R_{0,\mathrm{S,M-AS}}^{(m,j)}$ according to (4.31) or (4.34) in Subsection 4.5.3. Since maximizing the sum rate is equivalent to maximizing each cell's rate, optimization problems for the MC and each SC are proposed on the constraints of the sizes for each selected user subset, respectively, i.e.,

$$\max_{I \subseteq U_{\mathrm{M}}} \; R_{\mathrm{BS,M}}(I) \triangleq \frac{T-\tau}{T} \sum_{i \in I} R_{0,\mathrm{M,M-AS}}^{(i)} \tag{4.39a}$$

$$\text{s.t.} \quad |I| = K \tag{4.39b}$$

$$\max_{J_m \subseteq U_{\mathrm{S}}^{(m)}} \; R_{\mathrm{SC,M}}^{(m)}(J_m) \triangleq \frac{T-\tau}{T} \sum_{j \in J_s} R_{0,\mathrm{S,M-AS}}^{(m,j)} \tag{4.40a}$$

$$\text{s.t.} \quad |J_m| = L \tag{4.40b}$$

which can be solved separately by exhaustive search. Similarly, as a suboptimal solution, a simplified greedy search method is proposed, in which the user scheduling at each cell is operated separately, i.e., it can be completed by its own node without sharing any CSI with other cells. The reader is referred to Algorithm 4.2 for a step-by-step summary of the proposed method.

Asymptotic Scheduling Algorithm for ZFT (ASA-Z):

Likewise, since the Propositions 4.1 and 4.2 also provide the asymptotic results for ZFT, we can define the achievable rate of the ith ($i \in U_{\mathrm{BS}}$) MUE and jth ($j \in U_{\mathrm{SC}}^{(m)}$) SUE in the mth ($m = 1, \ldots, S$) SC as $R_{0,\mathrm{M,Z-AS}}^{(i)}$ and $R_{0,\mathrm{S,Z-AS}}^{(m,j)}$ according to (4.32) or (4.35) in Subsection 4.5.3. However, notably, $R_{0,\mathrm{M,Z-AS}}^{(1)} = R_{0,\mathrm{M,Z-AS}}^{(2)} = \cdots = R_{0,\mathrm{M,Z-AS}}^{(K)}$ for fixed $\sum_{k=1}^{K} \left(\beta_{\mathrm{B-M}}^{(k)}\right)^{-\eta}$, and $R_{0,\mathrm{S,Z-AS}}^{(m,1)} = R_{0,\mathrm{S,Z-AS}}^{(m,2)} = \cdots = R_{0,\mathrm{S,Z-AS}}^{(m,L)}$

Algorithm 4.2: Asymptotic scheduling algorithm for MRT

Initialization: $I = \emptyset$, $J_m = \emptyset$, $\tilde{U}_{\mathrm{M}} = U_{\mathrm{M}}$ and $\tilde{U}_{\mathrm{S}}^{(m)} = U_{\mathrm{S}}^{(m)}$ for $m = 1, \ldots, S$.
Repeat1:
$i^* = \arg\max_{i \in \tilde{U}_{\mathrm{M}}} R_{\mathrm{BS,M}}(I \cup \{i\})$
$I = I \cup \{i^*\}$, $\tilde{U}_{\mathrm{M}} = \tilde{U}_{\mathrm{M}} \backslash \{i^*\}$
Until1: Stop if $|I| = K$.
For $m = 1$ to S
Repeat2:
$j_m^* = \arg\max_{j_m \in \tilde{U}_{\mathrm{S}}^{(m)}} R_{\mathrm{SC,M}}^{(m)}(J_m \cup \{j_m\})$
$J_m = J_m \cup \{j_m^*\}$, $\tilde{U}_{\mathrm{S}}^{(m)} = \tilde{U}_{\mathrm{S}}^{(m)} \backslash \{j_m^*\}$
Until2: Stop if $|J_m| = L$.
Endfor
Output: Output I and J_m $(m = 1, \ldots, S)$ as the solutions.

for fixed $\sum_{l=1}^{L} \left(\beta_{\mathrm{S-S}}^{(m,m,l)}\right)^{-\eta}$ with $m = 1, \cdots, S$[2], which means that maximizing the sum rate is equivalent to

$$\min_{I \subseteq U_{\mathrm{M}}} \sum_{k \in I} \left(\beta_{\mathrm{B-M}}^{(k)}\right)^{-\eta} \tag{4.41a}$$

$$\text{s.t.} \quad |I| = K \tag{4.41b}$$

$$\min_{J_m \subseteq U_{\mathrm{S}}^{(m)}} \sum_{l \in J_m} \left(\beta_{\mathrm{S-S}}^{(m,m,l)}\right)^{-\eta} \tag{4.42a}$$

$$\text{s.t.} \quad |J_m| = L. \tag{4.42b}$$

Different from ASA-M, the optimal ASA-Z can be realized by the max-beta based scheduling algorithm, the process of which is similar to the Algorithm 2. To solve this optimization problems of (4.41) and (4.42), we just need to find K MUEs with the first K-maximum $\beta_{\mathrm{B-M}}^{(k)}$ and L SUEs with the first L-maximum $\beta_{\mathrm{S-S}}^{(m,m,l)}$, respectively. Hence, we replace the item $R_{\mathrm{BS,M}}(I \cup \{i\})$ and $R_{\mathrm{SC,M}}^{(m)}(J_m \cup \{j_m\})$ in the Algorithm 2 by $\beta_{\mathrm{B-M}}^{(i)}$ and $\beta_{\mathrm{S-S}}^{(m,m,j_m)}$, respectively, to realize the optimal ASA-Z, which is much simpler than ASA-M as neither achievable rate calculation nor SCSI exchange is needed at each node.

[2]Here, $\eta = 1$ for case I in (4.32) and $\eta = 2$ for case II in (4.35).

4.7 Numerical Results

In this section, we present numerical results to validate the derived achievable rate expressions and examine the performance of HetNet with large-scale antenna arrays in downlink channels. A dense tier of S SCs that are uniformly distributed in a circle as shown in Fig. 4.1. Suppose that the users in the cell are uniformly distributed with the total number 200 and 500 for $S = 8$ and $S = 20$, respectively, and that the MUEs and SUEs are identified by the bias user association. The number of scheduled MUEs and SUEs in each SC are $K = 20$ and $L = 4$, respectively, with training length $\tau = K + L \times \gamma$ and bias factors $\kappa_{\text{BS}} = 1$ and $\kappa_{\text{SC}} = 1.2$ [48] for BS and SC nodes, respectively. Also we assume that equal transmit power at the each SC node satisfies $p_{\text{SC}}^{(1)}$ (dBm)$= \cdots = p_{\text{SC}}^{(S)}$ (dBm)$= p_{\text{BS}}$ (dBm)-22 (dB) as p_{BS} changes, and that $N_{\text{BS}}/N_{\text{SC}} = \lambda = 10$ is fixed as N_{BS} changes. See Table 4.1 for simulation parameters and assumption details.

Table 4.1: Simulation Parameters

Parameters	Setting
Bandwidth	20 MHz
Macro Cell radius	1000 m
Distance between BS and SCs	800 m
The number of SCs	$S = 8$ or 20
Pilot Reuse factor	$\gamma = 1 \sim 8$ or $1 \sim 20$
Coherent interval	$T = 200$
Transmit power threshold (BS/SC)	$p_{\text{BS}} - p_{\text{SC}} = 22$ dB, $p_{\text{SC}}^{(1)} = \cdots = p_{\text{SC}}^{(S)} = p_{\text{SC}}$
Noise power density	-174 dBm/Hz
Number of MUEs/SUEs	$K = 20$, $L = 4$
Pathloss model (BS)	$\theta_{\text{BS}}(d) = 128.1 + 37.6\log_{10}(d)$, d (km) [45]
Pathloss model (SC)	$\theta_{\text{SC}}(d) = 140.7 + 36.7\log_{10}(d)$, d (km) [45]

4.7.1 Comparison between One-Tier and Two-Tier Network Topologies

First, the effectiveness of the considered large-scale two-tier model is demonstrated by comparing it with the one-tier network topology. Here, suppose that simple random scheduling algorithm (RSA) is utilized for user scheduling, $p_\tau = 0$ dBm, $S = 8$ and $\gamma = 8$. To be fair, in the one-tier network the number of antennas is assumed to satisfy $N = N_{\text{BS}} + S \times N_{\text{SC}}$ and the number of UEs is set to be $K + S \times L$ including both MUEs and SUEs in the two-tier network. Fig. 4.2 shows that the two-tier network outperforms one-tier network on the spectral efficiency

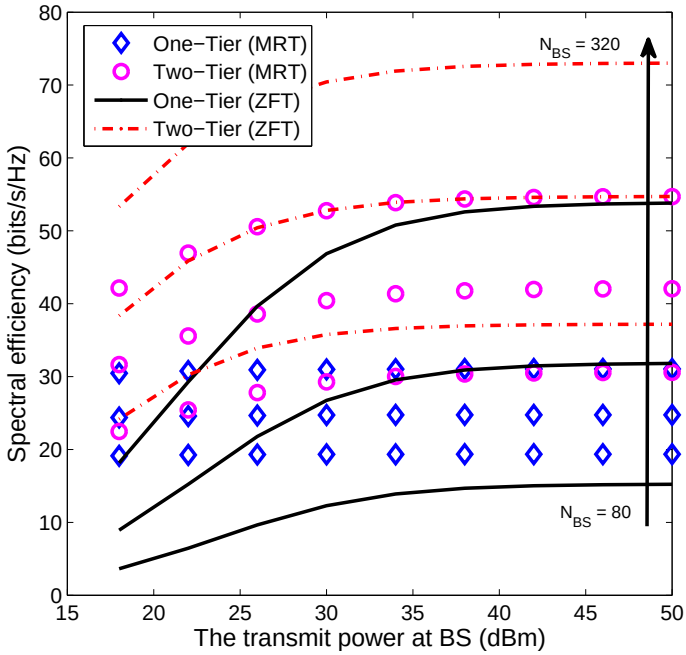

Figure 4.2: Spectral efficiency versus p_{BS} for one-tier and two-tier network topologies (RSA, $p_\tau = 0$ dBm, $S = 8$, $\gamma = 8$).

with both MRT and ZFT precoding schemes in the downlink channels. As an indication of coverage, Fig. 4.3 plots the cell boundary small cell user rate versus p_{BS} with fixed $N_{\text{BS}} = 80$, 160 and 320, where the cell boundary users are composed of all the SUEs identified by the biased user association [48]. For MRT the two-tier network has much higher cell boundary user rate, indicating better coverage than one-tier. However, for ZFT the two-tier network underperforms the one-tier for cell boundary SUEs, since neither the inter-SC interference nor the cross-tier interference can be cancelled in the non-cooperative two-tier HetNet systems. While in the one-tier network, the inter-user interferences could be totally eliminated by the ZFT precoding without inter-SC and cross-tier interferences. On the other hand, the ZFT precoding of the one-tier network is much more complex than that of the two-tier network due to the larger number of UEs ($K + S \times L$) served simultaneously by the BS.

4.7.2 Validation of Lower Capacity Bounds and Pilot Reuse Pattern

In this subsection, the effectiveness of the derived capacity lower bounds with imperfect CSI based MRT in (4.19) and (4.20) and ZFT in (4.28) and (4.29) is evaluated by comparing them with the Monte-Carlo simulation results. Then, the

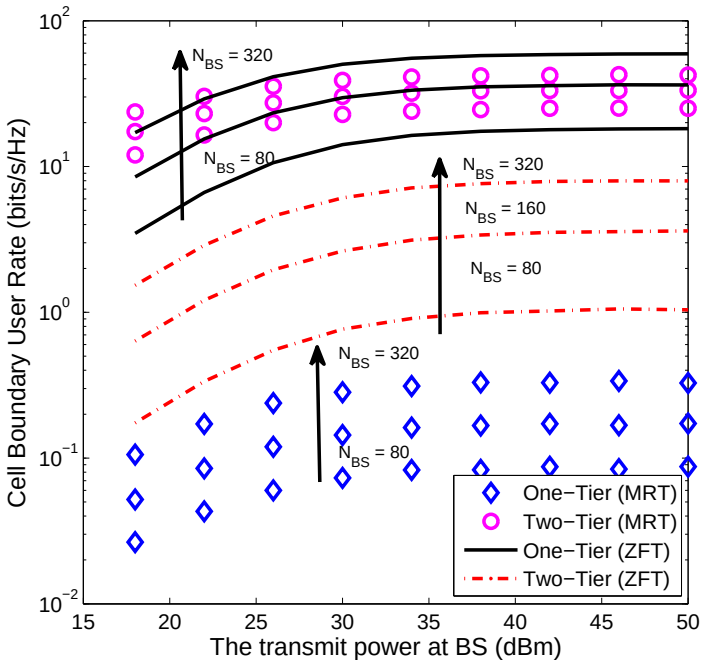

Figure 4.3: Cell boundary SC user rate versus p_{BS} for one-tier and two-tier network topologies (RSA, $p_\tau = 0$ dBm, $S = 8$, $\gamma = 8$).

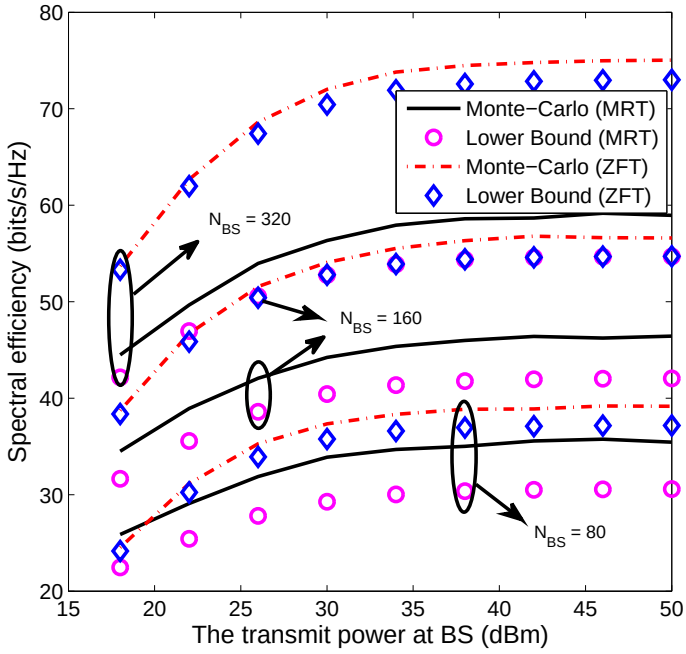

Figure 4.4: Spectral efficiency versus p_{BS} for Monte-Carlo results and lower bounds (RSA, $p_\tau = 0$ dBm, $S = 8$, $\gamma = 8$).

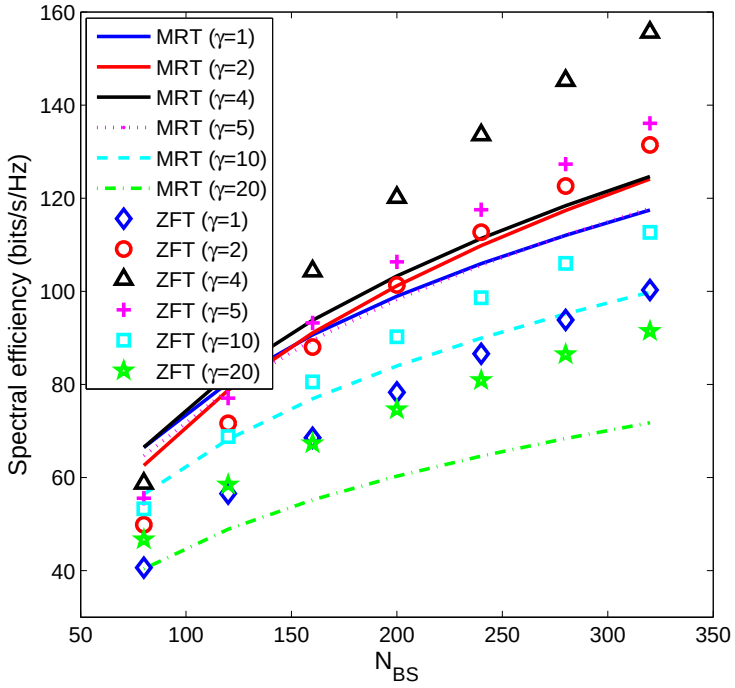

Figure 4.5: Spectral efficiency versus N_{BS} for different PR factors (RSA, $p_{\mathrm{BS}} = 46$ dBm, $p_\tau = 0$ dBm, $S = 20$).

performance of the presented simple PR pattern is evaluated. Besides, the asymptotic analyses with massive arrays for the two cases in Propositions 4.1 and 4.2 are examined in this subsection. Here, random scheduling algorithm (RSA) is utilized.

Firstly, $p_\tau = 0$ dBm, $S = 8$ and $\gamma = 8$ are set. In Fig. 4.4, the spectral efficiency versus p_{BS} curves with fixed $N_{\mathrm{BS}} = 80$, 160 and 320 for capacity lower bounds are compared with those obtained from (4.13) and (4.14) by Monte-Carlo simulation. "Lower Bound (MRT)" indicates the capacity lower bound obtained by (4.19) and (4.20), and "Lower Bound (ZFT)" is calculated from (4.28) and (4.29). It can be observed from Fig. 4.4 that the relative performance gap between "Lower Bound (ZFT)" and "Monte-Carlo (ZFT)" is quite smaller than that between "Lower Bound (MRT)" and "Monte-Carlo (MRT)", especially at lower transmit power with larger number of transmit antennas. Moreover, the spectral efficiency of ZFT increases much faster than that of MRT as p_{BS} increases, due to the fact that the effect of interference is much larger than that of the noise for higher SNR while ZFT is able to null multi-user interference signals [23]. In this way, the derived capacity lower bounds by Jensen's inequality are proved to be accurate predictors of the system performance.

Secondly, setting $p_{\mathrm{BS}} = 46$ dBm, $p_\tau = 0$ dBm and $S = 20$, the spectral efficiency of the presented PR pattern with different pilot reuse factor γ versus the number

of transmit antennas at BS is illustrated in Fig. 4.5. When the spectral efficiency is calculated by (4.19) and (4.20) taking the influence of the pilot overhead τ into consideration, Fig. 4.5 indicates that PR factor $\gamma = 4$ yields the best performance for both MRT and ZFT, which achieves the optimal trade off between pilot overhead and pilot contamination from the reuse of the pilot sets. Notably, when the PR factor is reduced to $\gamma = 2$ and $\gamma = 1$, the spectral efficiency of ZFT is no longer larger than that of MRT as shown in Fig. 4.5, which indicates that pilot reuse introduced pilot contamination effect is more severe for ZFT than that for MRT. It can be explained by the fact that more pilot reuse increases the channel estimation error, and thus monotonically decreases the SINR of the SUEs in the ZFT case as stated by Remark 4.2.

Finally, the asymptotic analyses with massive arrays for the two cases in Propositions 4.1 and 4.2 are examined with the MUE and SUE large-scale fading factors fixed as $\beta_{\mathrm{B-M}} = 1$, $\beta_{\mathrm{B-S}} = 0.2$, $\beta_{\mathrm{S-S}}^{(m,m)} = 5$, $\beta_{\mathrm{S-S}}^{(m,n)} = 0.6$ $(m \neq n)$ and $\beta_{\mathrm{S-M}} = 0.6$ considering pathloss, noise variance $\sigma_0^2 = 1$ and the normalized pilot power $p_\tau = 0$ dB. Fig. 4.6 shows the required MC and SC transmit power p_{BS} and p_{SC} to achieve 1 bit/s/Hz per MUE and SUE, respectively, for MRT in case I. It is obvious from Fig. 4.6 that in case I where the pilot power p_τ is fixed, the required p_{BS} and p_{SC} are significantly reduced as N_{SC} increases, and that the required p_{BS} with $\gamma = 1$ is the lowest and the p_{SC} with $\gamma = 4$ is the lowest in comparison to other PR factors. Moreover, for $\gamma < 4$ in Fig. 4.6, it is evident that the lower the PR factor γ is, the higher p_{SC} is required to achieve 1 bit/s/Hz per user. The observation indicates that the pilot contamination effect existed at the SCs increases the required transmit power p_{SC} to achieve 1 bit/s/Hz per SUE. Regarding the imperfect CSI effect, Fig. 4.7 shows that less transmit power in both MC and SCs are required when p_τ is high. For case II with $E_\tau = 0$ dB and the pilot power scaling down by $p_\tau = \frac{E_\tau}{N_{\mathrm{SC}}^\theta}$, Fig. 4.8 shows that higher θ leads to more slowly reduced P_τ, because the imperfect CSI effect becomes more severe when the pilot power is reduced much faster with the increase of N_{SC}.

4.7.3 User Scheduling

In this subsection, the proposed user scheduling algorithms are examined with respect to the spectral efficiency of the HetNet downlink systems. Here, we choose $p_\tau = 5$ dBm, $S = 8$ and $\gamma = 8$ for comparison fairness.

First, simulations are performed for MRT on the proposed two user scheduling algorithms in regard to the spectral efficiency versus p_{BS}, which is given in Fig. 4.9 with fixed $N_{\mathrm{BS}} = 80$, 160 and 320. We obtain the results that GSA outperforms

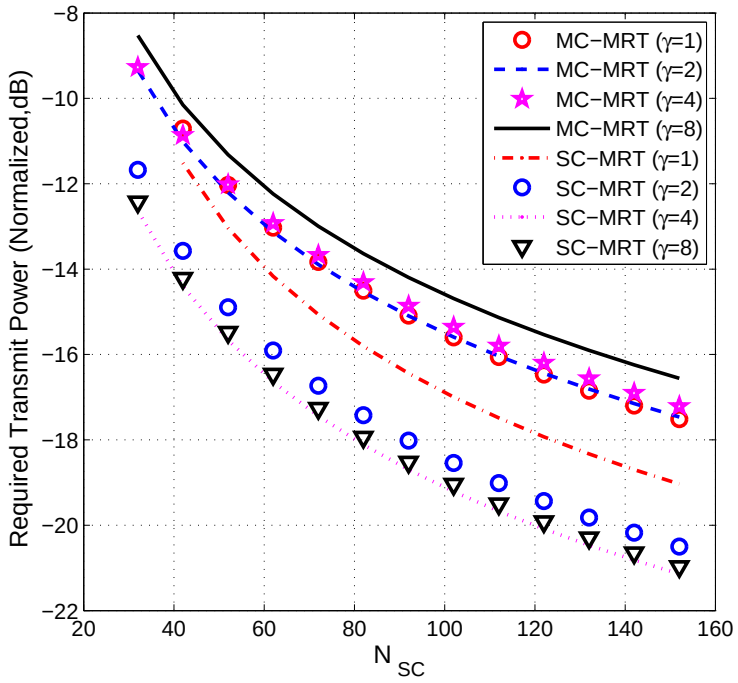

Figure 4.6: Transmit power required to achieve 1 bit/s/Hz per user for case I (normalized $p_\tau = 0$ dB, $S = 8$).

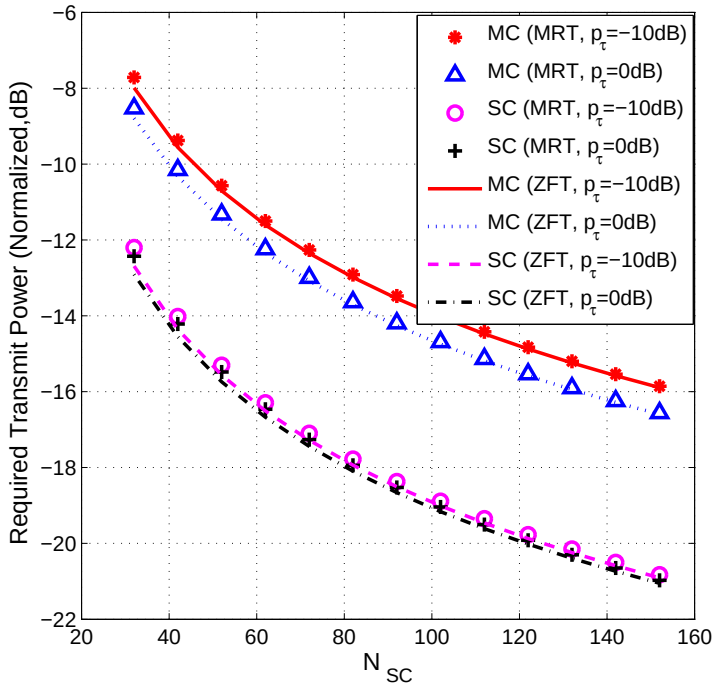

Figure 4.7: Transmit power required to achieve 1 bit/s/Hz per user for case I ($S = 8$, $\gamma = 8$).

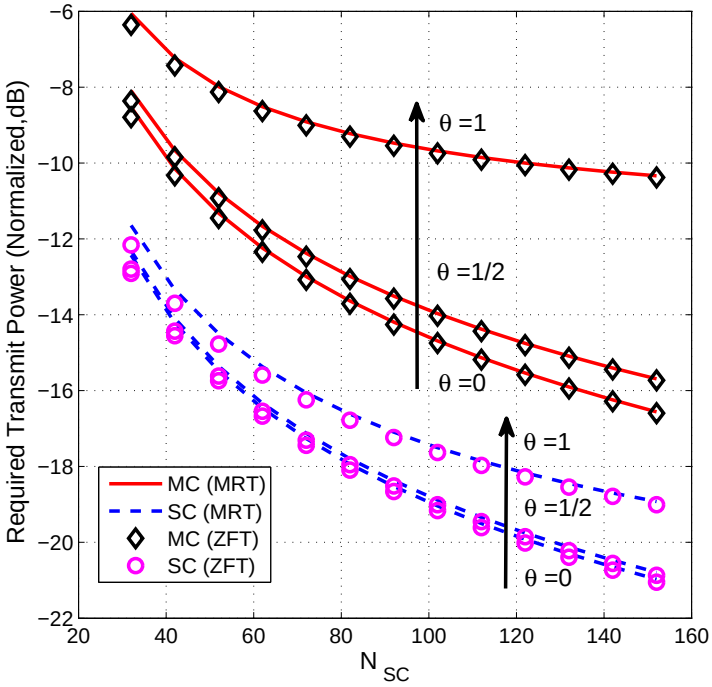

Figure 4.8: Transmit power required to achieve 1 bit/s/Hz per user for case II (normalized $E_\tau = 0$ dB, $S = 8$, $\gamma = 8$).

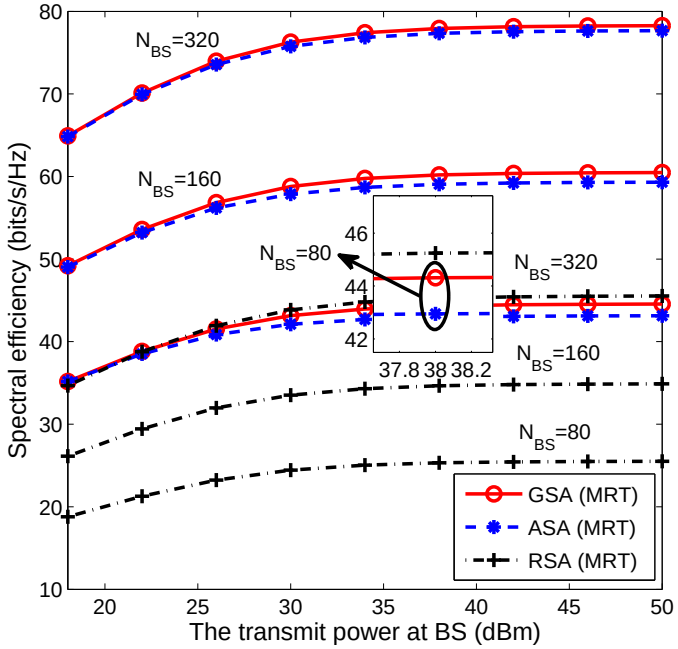

Figure 4.9: Spectral efficiency versus p_{BS} for different user scheduling algorithms (MRT, $p_\tau = 5$ dBm, $S = 8$, $\gamma = 8$).

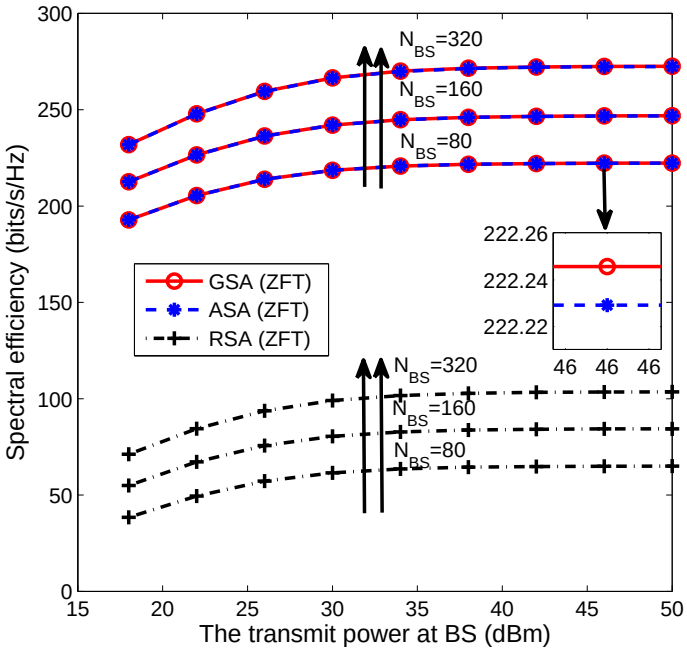

Figure 4.10: Spectral efficiency versus p_{BS} for different user scheduling algorithms (ZFT, $p_\tau = 5$ dBm, $S = 8$, $\gamma = 8$).

ASA and RSA even when N_{BS} is not sufficiently high, and that the curve of ASA approaches to the curve of GSA as N_{BS} increases. Furthermore, the performance gap between ASA and RSA for MRT is large. Fig. 4.10 shows the spectral efficiency versus p_{BS} for ZFT on two different user scheduling schemes. It can be observed that the two curves for GSA and ASA almost coincide with each other even when the number of antennas is not very large, which demonstrates the effectiveness of ASA. Hence, for user scheduling of both MRT and ZFT, ASA is a good choice that achieves better performance with lower complexity. Moreover, by comparing Fig. 4.4 with Fig. 4.9 and Fig. 4.10, it is shown that increasing p_τ from 0 dBm to 5 dBm introduces better performance for ZFT than MRT, indicating that different imperfect CSI effect on ZFT and MRT.

4.8 Conclusion

In this chapter, we have investigated the performance of a two-tier network with large-scale antenna arrays set at both BS and SCs. With MRT precoding employed at each node, we have derived capacity lower bounds with closed-form expressions for both imperfect CSI based MRT and ZFT cases, where a simple pilot reuse pattern is utilized for the channel estimation procedure to obtain the estimated

imperfect CSI, followed by asymptotic analyses. The benefits of employing large number of antennas at both BS and SCs have been demonstrated. Simulation results have shown that the derived closed-form expressions for the achievable rate are accurate predictors of the system performance for both MRT and ZFT, and that more pilot reuse is able to yield higher effective performance under different system configurations even though pilot contamination exists. As for user scheduling, two schemes have been proposed. The greedy scheduling algorithm designed based on the derived capacity lower bounds only requires the statistical CSI but not the instantaneous CSI exchange among BS and SCs. The asymptotic scheduling algorithm, based on the asymptotic analysis results, has even lower complexity by removing the need for any CSI exchange among nodes, and can still achieve near-optimal performance for both MRT and ZFT in the asymptotic regime of massive antenna arrays. Furthermore, it has been found that in the asymptotic regime and when ZFT is used, the capacity lower bound of a user is proportional to the large-scale fading factor of the channel from the user to its base station. Therefore, ASA-Z can further reduce complexity by circumventing the need to calculate the achievable rate.

4.9 Appendices

4.9.1 Proof of (4.10)

As to the detailed derivation of (4.10), we use the identity [30, 31]

$$\mathrm{E}\left[\mathbf{W}^{-1}\right] = \frac{\mathbf{\Sigma}^{-1}}{n-m-1} \tag{4.43}$$

where $\mathbf{W} \sim \mathcal{W}_m(\mathbf{\Sigma}, n)$ is an $m \times m$ central complex Wishart matrix with n $(n > m)$ degrees of freedom and the distribution of $\mathbf{W}^{-1}$ is called an inverted Wishart distribution, following $\mathcal{W}_m^{-1}(\mathbf{\Sigma}^{-1}, n)$. It can be easily concluded that $\left(\hat{\mathbf{G}}_{\mathrm{B-M}}^H \hat{\mathbf{G}}_{\mathrm{B-M}}\right)^{-1} \sim \mathcal{W}_K^{-1}(\hat{\mathbf{D}}_{\mathrm{B-M}}^{-1}, N_{\mathrm{BS}})$ and $\left[\left(\hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)}\right)^H \hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)}\right]^{-1} \sim \mathcal{W}_L^{-1}(\left(\hat{\mathbf{D}}_{\mathrm{S-S}}^{(m,m)}\right)^{-1}, N_{\mathrm{SC}})$ for $\forall m \in \{1, \cdots, S\}$, hence

$$\begin{aligned} \mathrm{E}\left[\left(\hat{\mathbf{G}}_{\mathrm{B-M}}^H \hat{\mathbf{G}}_{\mathrm{B-M}}\right)^{-1}\right] &= \frac{\hat{\mathbf{D}}_{\mathrm{B-M}}^{-1}}{N_{\mathrm{BS}}-K-1} \\ \mathrm{E}\left[\left[\left(\hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)}\right)^H \hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)}\right]^{-1}\right] &= \frac{\left(\hat{\mathbf{D}}_{\mathrm{S-S}}^{(m,m)}\right)^{-1}}{N_{\mathrm{SC}}-L-1} \end{aligned} \tag{4.44}$$

where $N_{\mathrm{BS}} > K$ and $N_{\mathrm{SC}} > L$. In this way, we have

$$\begin{aligned}
\alpha_{\mathrm{BS}} &= \sqrt{\frac{P_{\mathrm{BS}}}{\mathrm{E}\left[\mathrm{Tr}\left\{\left(\hat{\mathbf{G}}_{\mathrm{B-M}}^{H}\hat{\mathbf{G}}_{\mathrm{B-M}}\right)^{-1}\right\}\right]}} = \sqrt{\frac{\left(N_{\mathrm{BS}}-K-1\right)P_{\mathrm{BS}}}{\Psi_{\mathrm{B-M}}}} \\
\alpha_{\mathrm{SC}}^{(m)} &= \sqrt{\frac{P_{\mathrm{SC}}^{(m)}}{\mathrm{E}\left[\mathrm{Tr}\left\{\left[\left(\hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)}\right)^{H}\hat{\mathbf{G}}_{\mathrm{S-S}}^{(m,m)}\right]^{-1}\right\}\right]}} = \sqrt{\frac{\left(N_{\mathrm{SC}}-L-1\right)P_{\mathrm{SC}}^{(m)}}{\Psi_{\mathrm{S-S}}^{(m)}}}
\end{aligned} \tag{4.45}$$

where $\Psi_{\mathrm{B-M}}=\sum_{i=1}^{K}\frac{1}{\hat{\beta}_{\mathrm{B-M}}^{(i)}}$, and $\Psi_{\mathrm{S-S}}^{(m)}=\sum_{l=1}^{L}\frac{1}{\hat{\beta}_{\mathrm{S-S}}^{(m,m,l)}}$ with $m\in\{1,\ldots,S\}$.

4.9.2 Proof of Theorem 4.1

To derive the closed-form expression of the MUE achievable rate in (4.19) based on Jensen's inequality, we start from the expectation of the SINR's reciprocal, given by

$$\begin{aligned}
&\mathrm{E}\left[\frac{\mathrm{EEI}_i+\mathrm{IMI}_i+\mathrm{CTI}_i+\sigma_0^2}{\alpha_{\mathrm{BS}}^2\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^4}\right] \\
&= \mathrm{E}\left[\frac{\mathrm{EEI}_i}{\alpha_{\mathrm{BS}}^2\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^4}\right]+\mathrm{E}\left[\frac{\mathrm{IMI}_i}{\alpha_{\mathrm{BS}}^2\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^4}\right]+\mathrm{E}\left[\frac{1}{\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^4}\right]\frac{\mathrm{E}\left[\mathrm{CTI}_i\right]+\sigma_0^2}{\alpha_{\mathrm{BS}}^2}
\end{aligned} \tag{4.46}$$

where the item of $\frac{\mathrm{E}[\mathrm{CTI}_i]+\sigma_0^2}{\alpha_{\mathrm{BS}}^2}$ does not depend on $\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}$. Since large-scale antenna arrays are set at both the BS and SCs, some results from Gaussian distributed estimated channel in (4.4)[3] [28] can be utilized. Using [29, Lemma 2.9]

$$\begin{aligned}
\mathrm{E}\left[\mathrm{Tr}\left\{\mathbf{W}^{-1}\right\}\right] &= \frac{m}{n-m} \\
\mathrm{E}\left[Tr^2\left\{\mathbf{W}^{-1}\right\}\right] &= \frac{m}{n-m}\left(\frac{n}{(n-m)^2-1}+\frac{m-1}{n-m+1}\right)
\end{aligned} \tag{4.47}$$

[3]Due to the estimated channel model in (4.4), we have that $\hat{\mathbf{g}}_i$ and $\hat{\mathbf{g}}_j$ are mutually independent $N\times 1$ vectors with $\forall i\neq j$ whose elements are i.i.d. zero-mean Gaussian distributed with variances $\hat{\sigma}_i^2$ and $\hat{\sigma}_j^2$, respectively. Then, it can be concluded that $\mathrm{E}\left[\hat{\mathbf{g}}_i^H\hat{\mathbf{g}}_i\right]=N\hat{\sigma}_i^2$, $\mathrm{E}\left[\hat{\mathbf{g}}_j^H\hat{\mathbf{g}}_j\right]=N\hat{\sigma}_j^2$, and $\mathrm{E}\left[\hat{\mathbf{g}}_i^H\hat{\mathbf{g}}_j\right]=0$. Also, we can obtain that $\mathrm{E}\left[|\hat{\mathbf{g}}_i^H\hat{\mathbf{g}}_j|^2\right]=N\hat{\sigma}_i^2\hat{\sigma}_j^2$.

where $\mathbf{W} \sim \mathcal{W}_m(\mathbf{I}_n, n)$ is an $m \times m$ central complex Wishart matrix with n ($n > m+1$) degrees of freedom, we have

$$\mathrm{E}\left[\frac{1}{\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^2}\right] = \frac{1}{(N_{\mathrm{BS}}-1)\hat{\beta}_{\mathrm{B-M}}^{(i)}},\ \mathrm{E}\left[\frac{1}{\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^4}\right] = \frac{1}{(N_{\mathrm{BS}}-1)(N_{\mathrm{BS}}-2)\left(\hat{\beta}_{\mathrm{B-M}}^{(i)}\right)^2}. \tag{4.48}$$

Then, we define $\tilde{\xi}_{\mathrm{B-M}}^{(i)} \triangleq \frac{\left(\xi_{\mathrm{B-M}}^{(i)}\right)^H \hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}}{\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|} \sim \mathcal{CN}\left(0, \beta_{\mathrm{B-M}}^{(i)} - \hat{\beta}_{\mathrm{B-M}}^{(i)}\right)$ and $\hat{\tilde{g}}_{\mathrm{B-M}}^{(k)} \triangleq \frac{\left(\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right)^H \hat{\mathbf{g}}_{\mathrm{B-M}}^{(k)}}{\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|} \sim \mathcal{CN}\left(0, \hat{\beta}_{\mathrm{B-M}}^{(k)}\right)$. Conditioned on $\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}$, $\tilde{\xi}_{\mathrm{B-M}}^{(i)}$ and $\hat{\tilde{g}}_{\mathrm{B-M}}^{(k)}$ are Gaussian distributed and independent of $\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}$. Hence, we have

$$\mathrm{E}\left[\frac{\mathrm{EEI}_i}{\alpha_{\mathrm{BS}}^2\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^4}\right] = \mathrm{E}\left[\frac{1}{\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^2}\right]\mathrm{E}\left[\left|\tilde{\xi}_{\mathrm{B-M}}^{(i)}\right|^2\right] = \frac{\beta_{\mathrm{B-M}}^{(i)} - \hat{\beta}_{\mathrm{B-M}}^{(i)}}{(N_{\mathrm{BS}}-1)\hat{\beta}_{\mathrm{B-M}}^{(i)}} \tag{4.49}$$

and

$$\begin{aligned}
&\mathrm{E}\left[\frac{\mathrm{IMI}_i}{\alpha_{\mathrm{BS}}^2\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^4}\right] = \sum_{k\neq i}^{K}\mathrm{E}\left[\frac{\left|\left(\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}+\xi_{\mathrm{B-M}}^{(i)}\right)^H\hat{\mathbf{g}}_{\mathrm{B-M}}^{(k)}\right|^2}{\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^4}\right] \\
=&\mathrm{E}\left[\frac{1}{\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^2}\right]\sum_{k\neq i}^{K}\mathrm{E}\left[\left|\hat{\tilde{g}}_{\mathrm{B-M}}^{(k)}\right|^2\right] + \mathrm{E}\left[\frac{1}{\left\|\hat{\mathbf{g}}_{\mathrm{B-M}}^{(i)}\right\|^4}\right]\sum_{k\neq i}^{K}\mathrm{E}\left[\left|\left(\xi_{\mathrm{B-M}}^{(i)}\right)^T\left(\hat{\mathbf{g}}_{\mathrm{B-M}}^{(k)}\right)^*\right|^2\right] \\
=&\frac{\sum\limits_{k\neq i}^{K}\hat{\beta}_{\mathrm{B-M}}^{(k)}}{(N_{\mathrm{BS}}-1)\hat{\beta}_{\mathrm{B-M}}^{(i)}} + \frac{N_{\mathrm{BS}}\sum\limits_{k\neq i}^{K}\left(\beta_{\mathrm{B-M}}^{(i)}-\hat{\beta}_{\mathrm{B-M}}^{(i)}\right)\hat{\beta}_{\mathrm{B-M}}^{(k)}}{(N_{\mathrm{BS}}-1)(N_{\mathrm{BS}}-2)\left(\hat{\beta}_{\mathrm{B-M}}^{(i)}\right)^2}.
\end{aligned} \tag{4.50}$$

Also, Gaussian distribution leads to

$$\mathrm{E}\left[\mathrm{CTI}_i\right] = \sum_{n=1}^{S}\sum_{l=1}^{L}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\mathrm{E}\left[\left|\left(\mathbf{g}_{\mathrm{S-M}}^{(n,i)}\right)^T\left(\hat{\mathbf{g}}_{\mathrm{S-S}}^{(n,n,l)}\right)^*\right|^2\right] = \sum_{n=1}^{S}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2 N_{\mathrm{SC}}\Phi_{\mathrm{S-S}}^{(n)}\beta_{\mathrm{S-M}}^{(n,i)}. \tag{4.51}$$

By substituting (4.49), (4.50) and (4.51) into (4.46), (4.17) leads to

$$R_{0,\mathrm{M,M}}^{(i)} = \log_2\left(1 + \frac{\alpha_{\mathrm{BS}}^2 (N_{\mathrm{BS}}-1)(N_{\mathrm{BS}}-2)\left(\hat{\beta}_{\mathrm{B-M}}^{(i)}\right)^2}{\alpha_{\mathrm{BS}}^2 \hat{\chi}_{\mathrm{BS}}^{(i)} + \sum_{n=1}^{S}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2 N_{\mathrm{SC}} \Phi_{\mathrm{S-S}}^{(n)} \beta_{\mathrm{S-M}}^{(n,i)} + \sigma_0^2}\right) \tag{4.52}$$

where $\hat{\chi}_{\mathrm{BS}}^{(i)} = \Phi_{\mathrm{B-M}}\left(N_{\mathrm{BS}}\beta_{\mathrm{B-M}}^{(i)} - 2\hat{\beta}_{\mathrm{B-M}}^{(i)}\right) - (N_{\mathrm{BS}}-4)\left(\hat{\beta}_{\mathrm{B-M}}^{(i)}\right)^2 - 2\hat{\beta}_{\mathrm{B-M}}^{(i)}\beta_{\mathrm{B-M}}^{(i)}$.

For the derivation of the closed-form expression for the SUE achievable rate in (4.20), the expectation of the SINR's reciprocal can be written as

$$\begin{aligned}\mathrm{E}\left[\frac{\mathrm{EEI}_{m,j} + \mathrm{CTI}_{m,j} + \mathrm{ISI}_{m,j} + \mathrm{SSI}_{m,j} + \sigma_0^2}{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2 \left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^4}\right] &= \mathrm{E}\left[\frac{\mathrm{EEI}_{m,j}}{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2 \left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^4}\right] \\ &+ \mathrm{E}\left[\frac{\mathrm{ISI}_{m,j}}{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2 \left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^4}\right] + \mathrm{E}\left[\frac{1}{\left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^4}\right] \frac{\mathrm{E}\left[\mathrm{CTI}_{m,j}\right] + \mathrm{E}\left[\mathrm{SSI}_{m,j}\right] + \sigma_0^2}{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2}\end{aligned} \tag{4.53}$$

where $\frac{\mathrm{E}[\mathrm{CTI}_{m,j}]+\mathrm{E}[\mathrm{SSI}_{m,j}]+\sigma_0^2}{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2}$ is independent of $\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}$. According to the [29, Lemma 2.9] given in (4.47), we have

$$\begin{aligned}\mathrm{E}\left[\frac{1}{\left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^2}\right] &= \frac{1}{(N_{\mathrm{SC}}-1)\,\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}} \\ \mathrm{E}\left[\frac{1}{\left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^4}\right] &= \frac{1}{(N_{\mathrm{SC}}-1)\,(N_{\mathrm{SC}}-2)\left(\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}\right)^2}.\end{aligned} \tag{4.54}$$

Likewise, we define $\tilde{\xi}_{\mathrm{S-S}}^{(m,m,j)} \triangleq \frac{\left(\xi_{\mathrm{S-S}}^{(m,m,j)}\right)^H \hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}}{\left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|} \sim CN\left(0, \beta_{\mathrm{S-S}}^{(m,m,j)} - \hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}\right)$ and $\hat{\tilde{g}}_{\mathrm{S-S}}^{(m,m,l_1)} \triangleq \frac{\left(\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right)^H \hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,l_1)}}{\left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|} \sim CN\left(0, \hat{\beta}_{\mathrm{S-S}}^{(m,m,l_1)}\right)$, both of which do not depend on $\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}$ conditioned on it, then we obtain

$$\mathrm{E}\left[\frac{\mathrm{EEI}_{m,j}}{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2 \left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^4}\right] = \mathrm{E}\left[\frac{1}{\left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^2}\right] \mathrm{E}\left[\left|\tilde{\xi}_{\mathrm{S-S}}^{(m,m,j)}\right|^2\right] = \frac{\beta_{\mathrm{S-S}}^{(m,m,j)} - \hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}}{(N_{\mathrm{SC}}-1)\,\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}} \tag{4.55}$$

and

$$\mathrm{E}\left[\frac{\mathrm{ISI}_{m,j}}{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2\left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^4}\right]=\sum_{l_1\neq j}^{L}\mathrm{E}\left[\frac{\left|\left(\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}+\xi_{\mathrm{S-S}}^{(m,m,j)}\right)^T\left(\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,l_1)}\right)^*\right|^2}{\left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^4}\right]$$

$$=\mathrm{E}\left[\frac{1}{\left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^2}\right]\sum_{l_1\neq j}^{L}\mathrm{E}\left[\left|\hat{\tilde{g}}_{\mathrm{S-S}}^{(m,m,l_1)}\right|^2\right]+\mathrm{E}\left[\frac{1}{\left\|\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,j)}\right\|^4}\right]\sum_{l_1\neq j}^{L}\mathrm{E}\left[\left|\left(\xi_{\mathrm{S-S}}^{(m,m,j)}\right)^T\left(\hat{\mathbf{g}}_{\mathrm{S-S}}^{(m,m,l_1)}\right)^*\right|^2\right]$$

$$=\frac{\sum\limits_{l_1\neq j}^{L}\hat{\beta}_{\mathrm{S-S}}^{(m,m,l_1)}}{\left(N_{\mathrm{SC}}-1\right)\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}}+\frac{N_{\mathrm{SC}}\sum\limits_{l_1\neq j}^{L}\left(\beta_{\mathrm{S-S}}^{(m,m,j)}-\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}\right)\hat{\beta}_{\mathrm{S-S}}^{(m,m,l_1)}}{\left(N_{\mathrm{SC}}-1\right)\left(N_{\mathrm{SC}}-2\right)\left(\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}\right)^2} \tag{4.56}$$

Similarly, the law of large numbers [28] leads to

$$\mathrm{E}\left[\mathrm{CTI}_{m,j}\right]=\alpha_{\mathrm{BS}}^2N_{\mathrm{BS}}\Phi_{\mathrm{B-M}}\beta_{\mathrm{B-S}}^{(m,j)} \tag{4.57}$$

$$\mathrm{E}\left[\mathrm{SSI}_{m,j}\right]=\sum_{n\neq m}^{S}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2N_{\mathrm{SC}}\Phi_{\mathrm{S-S}}^{(n)}\beta_{\mathrm{S-S}}^{(n,m,j)}+\sum_{\substack{n\neq m\\ n\in\mathcal{A}_r}}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2N_{\mathrm{SC}}^2\left(\frac{\beta_{\mathrm{S-S}}^{(n,m,j)}\hat{\beta}_{\mathrm{S-S}}^{(n,n,j)}}{\beta_{\mathrm{S-S}}^{(n,n,j)}}\right)^2 \tag{4.58}$$

where $\hat{\beta}_{\mathrm{S-S}}^{(n,m,j)}\triangleq\frac{\tau p_\tau\left(\beta_{\mathrm{S-S}}^{(n,m,j)}\right)^2}{\tau p_\tau\sum\limits_{l\in\mathcal{A}_r}\beta_{\mathrm{S-S}}^{(n,l,j)}+\sigma_0^2}$. Then, substituting (4.55), (4.56), and (4.57) into (4.54), (4.18) leads to

$$R_{0,\mathrm{S,M}}^{(m,j)}=$$

$$\log_2\left(1+\frac{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2\left(N_{\mathrm{SC}}-1\right)\left(N_{\mathrm{SC}}-2\right)\left(\beta_{\mathrm{S-S}}^{(m,m,j)}\right)^2}{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2\hat{\chi}_{\mathrm{SC}}^{(m,j)}+\alpha_{\mathrm{BS}}^2\hat{\chi}_{\mathrm{BS}}^{(m,j)}+\sum\limits_{n\neq m}^{S}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\hat{\chi}_{\mathrm{SC},1}^{(n,m,j)}+\sum\limits_{\substack{n\neq m\\ n\in\mathcal{A}_r}}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\hat{\chi}_{\mathrm{SC},2}^{(n,m,j)}+\sigma_0^2}\right) \tag{4.59}$$

where $\hat{\chi}_{\mathrm{SC}}^{(m,j)}=\Phi_{\mathrm{S-S}}^{(m)}\left(N_{\mathrm{SC}}\beta_{\mathrm{S-S}}^{(m,m,j)}-2\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}\right)-\left(N_{\mathrm{SC}}-4\right)\left(\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}\right)^2-2\beta_{\mathrm{S-S}}^{(m,m,j)}\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}$, $\hat{\chi}_{\mathrm{BS}}^{(m,j)}=N_{\mathrm{BS}}\Phi_{\mathrm{B-M}}\beta_{\mathrm{B-S}}^{(m,j)}$, $\hat{\chi}_{\mathrm{SC},1}^{(n,m,j)}=N_{\mathrm{SC}}\Phi_{\mathrm{S-S}}^{(n)}\beta_{\mathrm{S-S}}^{(n,m,j)}$, and $\hat{\chi}_{\mathrm{SC},2}^{(n,m,j)}=N_{\mathrm{SC}}^2\beta_{\mathrm{S-S}}^{(n,m,j)}\hat{\beta}_{\mathrm{S-S}}^{(n,m,j)}$.

Finally, by substituting (4.8) into (4.52) and (4.59), (4.19) and (4.20) are obtained and thus Theorem 4.1 is demonstrated.

4.9.3 Proof of Theorem 4.2

Similar to the proof of Theorem 1 in Appendix 4.9.2, to derive the closed-form expression of the ZFT based MUE achievable rate in (4.28) based on Jensen's inequality, we start from the expectation of the SINR's reciprocal, given by

$$\mathrm{E}\left[\frac{\mathrm{EEI}_i+\mathrm{IMI}_i+\mathrm{CTI}_i+\sigma_0^2}{\alpha_{\mathrm{BS}}^2}\right]=\frac{\mathrm{E}\left[\mathrm{EEI}_i+\mathrm{IMI}_i\right]+\mathrm{E}\left[\mathrm{CTI}_i\right]+\sigma_0^2}{\alpha_{\mathrm{BS}}^2}. \tag{4.60}$$

Then, the property of the inverted Wishart Distribution in (4.44), the law of large numbers [28] and $\mathrm{Tr}\{\mathbf{AB}\}=\mathrm{Tr}\{\mathbf{BA}\}$ offer us

$$\begin{aligned}\mathrm{E}\left[\mathrm{EEI}_i+\mathrm{IMI}_i\right]&=\mathrm{E}\left[\alpha_{\mathrm{BS}}^2\left|\left(\xi_{\mathrm{B-M}}^{(i)}\right)^H\hat{\tilde{\mathbf{g}}}_{\mathrm{B-M}}^{(i)}\right|^2+\sum_{k\neq i}^{K}\alpha_{\mathrm{BS}}^2\left|\left(\xi_{\mathrm{B-M}}^{(i)}\right)^H\hat{\tilde{\mathbf{g}}}_{\mathrm{B-M}}^{(k)}\right|^2\right]\\&=\alpha_{\mathrm{BS}}^2\mathrm{E}\left[\left(\xi_{\mathrm{B-M}}^{(i)}\right)^H\hat{\tilde{\mathbf{G}}}_{\mathrm{B-M}}\hat{\tilde{\mathbf{G}}}_{\mathrm{B-M}}^H\xi_{\mathrm{B-M}}^{(i)}\right]\\&=\alpha_{\mathrm{BS}}^2\left(\beta_{\mathrm{B-M}}^{(i)}-\hat{\beta}_{\mathrm{B-M}}^{(i)}\right)\frac{\Psi_{\mathrm{B-M}}}{N_{\mathrm{BS}}-K-1}\end{aligned} \tag{4.61}$$

and

$$\begin{aligned}\mathrm{E}\left[\mathrm{CTI}_i\right]&=\mathrm{E}\left[\sum_{n=1}^{S}\sum_{l=1}^{L}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\left|\left(\mathbf{g}_{\mathrm{S-M}}^{(n,i)}\right)^H\left(\hat{\tilde{\mathbf{g}}}_{\mathrm{S-S}}^{(n,n,l)}\right)^*\right|^2\right]\\&=\sum_{n=1}^{S}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\mathrm{E}\left[\left(\mathbf{g}_{\mathrm{S-M}}^{(n,i)}\right)^H\hat{\tilde{\mathbf{G}}}_{\mathrm{S-S}}^{(n,n)}\left(\hat{\tilde{\mathbf{G}}}_{\mathrm{S-S}}^{(n,n)}\right)^H\mathbf{g}_{\mathrm{S-M}}^{(n,i)}\right]\\&=\sum_{n=1}^{S}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\beta_{\mathrm{S-S}}^{(n,i)}\mathrm{Tr}\left\{\mathrm{E}\left[\left[\left(\hat{\mathbf{G}}_{\mathrm{S-S}}^{(n,n)}\right)^H\hat{\mathbf{G}}_{\mathrm{S-S}}^{(n,n)}\right]^{-1}\right]\right\}\\&=\sum_{n=1}^{S}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\beta_{\mathrm{S-S}}^{(n,i)}\frac{\Psi_{\mathrm{S-S}}^{(n)}}{N_{\mathrm{SC}}-L-1}\end{aligned} \tag{4.62}$$

where $\Psi_{\mathrm{B-M}}=\sum_{i=1}^{K}\frac{1}{\hat{\beta}_{\mathrm{B-M}}^{(i)}}$, and $\Psi_{\mathrm{S-S}}^{(n)}=\sum_{l=1}^{L}\frac{1}{\hat{\beta}_{\mathrm{S-S}}^{(n,n,l)}}$ with $n\in\{1,\ldots,S\}$.

By substituting (4.61) and (4.62) into (4.60), then (4.17) leads to

$$R_{0,\mathrm{M,Z}}^{(i)}=\log_2\left(1+\frac{\alpha_{\mathrm{BS}}^2}{\alpha_{\mathrm{BS}}^2\frac{\left(\beta_{\mathrm{B-M}}^{(i)}-\hat{\beta}_{\mathrm{B-M}}^{(i)}\right)\Psi_{\mathrm{B-M}}}{N_{\mathrm{BS}}-K-1}+\sum_{n=1}^{S}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\beta_{\mathrm{S-S}}^{(n,i)}\frac{\Psi_{\mathrm{S-S}}^{(n)}}{N_{\mathrm{SC}}-L-1}+\sigma_0^2}\right). \tag{4.63}$$

For the derivation of the closed-form expression for the SUE achievable rate in

(4.29), the expectation of the SINR's reciprocal can be written as

$$\begin{aligned} &\mathrm{E}\left[\frac{\mathrm{EEI}_{m,j}+\mathrm{CTI}_{m,j}+\mathrm{ISI}_{m,j}+\mathrm{SSI}_{m,j}+\sigma_0^2}{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2}\right] \\ =\ &\frac{\mathrm{E}\left[\mathrm{EEI}_{m,j}+\mathrm{ISI}_{m,j}\right]+\mathrm{E}\left[\mathrm{CTI}_{m,j}\right]+\mathrm{E}\left[\mathrm{SSI}_{m,j}\right]+\sigma_0^2}{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2}. \end{aligned} \tag{4.64}$$

Similarly, according to the property of the inverted Wishart Distribution in (4.44), we have

$$\begin{aligned} \mathrm{E}\left[\mathrm{EEI}_{m,j}+\mathrm{ISI}_{m,j}\right] &= \mathrm{E}\left[\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2\left|\left(\xi_{\mathrm{S-S}}^{(m,m,j)}\right)^H\hat{\bar{\mathbf{g}}}_{\mathrm{S-S}}^{(m,m,j)}\right|^2\right. \\ \left.+\sum_{l_1\neq j}^{L}\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2\left|\left(\xi_{\mathrm{S-S}}^{(m,m,j)}\right)^H\hat{\bar{\mathbf{g}}}_{\mathrm{S-S}}^{(m,m,l_1)}\right|^2\right] &= \frac{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2\left(\beta_{\mathrm{S-S}}^{(m,m,j)}-\hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}\right)\Psi_{\mathrm{S-S}}^{(m)}}{N_{\mathrm{SC}}-L-1} \end{aligned} \tag{4.65}$$

$$\mathrm{E}\left[\mathrm{CTI}_{m,j}\right] = \mathrm{E}\left[\sum_{i=1}^{K}\alpha_{\mathrm{BS}}^2\left|\left(\mathbf{g}_{\mathrm{B-S}}^{(m,j)}\right)^H\hat{\bar{\mathbf{g}}}_{\mathrm{B-M}}^{(i)}\right|^2\right] = \frac{\alpha_{\mathrm{BS}}^2\beta_{\mathrm{B-S}}^{(m,j)}\Psi_{\mathrm{B-M}}}{N_{\mathrm{BS}}-K-1} \tag{4.66}$$

and

$$\begin{aligned} \mathrm{E}\left[\mathrm{SSI}_{m,j}\right] &= \sum_{n\neq m}^{S}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\mathrm{E}\left[\sum_{l_2=1}^{L}\left|\left(\mathbf{g}_{\mathrm{S-S}}^{(n,m,j)}\right)^H\hat{\bar{\mathbf{g}}}_{\mathrm{S-S}}^{(n,n,l_2)}\right|^2\right] = \frac{\sum_{n\neq m}^{S}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\beta_{\mathrm{S-S}}^{(n,m,j)}\Psi_{\mathrm{S-S}}^{(n)}}{N_{\mathrm{SC}}-L-1} \\ &+\sum_{\substack{n\neq m\\ n\in\mathcal{A}_r}}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\left[\left(\frac{\beta_{\mathrm{S-S}}^{(n,m,j)}}{\beta_{\mathrm{S-S}}^{(n,n,j)}}\right)^2-\frac{\beta_{\mathrm{S-S}}^{(n,m,j)}}{\left(N_{\mathrm{SC}}-L-1\right)\hat{\beta}_{\mathrm{S-S}}^{(n,n,j)}}\right]. \end{aligned} \tag{4.67}$$

Then, substituting (4.65), (4.66) and (4.67) into (4.64), (4.18) leads to

$$R_{0,\mathrm{S,Z}}^{(m,j)} = \log_2\left(1+\frac{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2}{\left(\alpha_{\mathrm{SC}}^{(m)}\right)^2\hat{\eta}_{\mathrm{SC}}^{(m,j)}+\alpha_{\mathrm{BS}}^2\hat{\eta}_{\mathrm{BS}}^{(m,j)}+\sum_{n\neq m}^{S}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\hat{\eta}_{\mathrm{SC},1}^{(n,m,j)}+\sum_{\substack{n\neq m\\ n\in\mathcal{A}_r}}\left(\alpha_{\mathrm{SC}}^{(n)}\right)^2\hat{\eta}_{\mathrm{SC},2}^{(n,m,j)}+\sigma_0^2}\right) \tag{4.68}$$

where $\hat{\eta}_{\mathrm{SC}}^{(m,j)} = \frac{\left(\beta_{\mathrm{S-S}}^{(n,m,j)} - \hat{\beta}_{\mathrm{S-S}}^{(m,m,j)}\right)\Psi_{\mathrm{S-S}}^{(m)}}{N_{\mathrm{SC}}-L-1}$, $\hat{\eta}_{\mathrm{BS}}^{(m,j)} = \frac{\beta_{\mathrm{B-S}}^{(m,j)}\Psi_{\mathrm{B-M}}}{N_{\mathrm{BS}}-K-1}$, $\hat{\eta}_{\mathrm{SC},1}^{(n,m,j)} = \frac{\beta_{\mathrm{S-S}}^{(n,m,j)}\Psi_{\mathrm{S-S}}^{(n)}}{N_{\mathrm{SC}}-L-1}$, and $\hat{\eta}_{\mathrm{SC},2}^{(n,m,j)} = \left[\left(\frac{\beta_{\mathrm{S-S}}^{(n,m,j)}}{\beta_{\mathrm{S-S}}^{(n,n,j)}}\right)^2 - \frac{\beta_{\mathrm{S-S}}^{(n,m,j)}}{(N_{\mathrm{SC}}-L-1)\hat{\beta}_{\mathrm{S-S}}^{(n,n,j)}}\right]$. Finally, by substituting (4.10) into (4.63) and (4.68), (4.28) and (4.29) are obtained and thus Theorem 4.2 is demonstrated.

Chapter 5

Conclusions and Further Research Issues

5.1 Conclusions

In this dissertation, we have investigated the multi-Pair two-way AF massive MIMO relaying, MSE precoding based HetNet and large-scale antenna arrays equipped HetNet. The conclusions are drawn as follows:

- **Power Allocation for Multi-Pair Massive MIMO Two-Way AF Relaying with Linear Processing**

  We have investigated the performance of a multi-pair two-way AF massive MIMO relaying system. With imperfect CSI based MRC/MRT and ZFR/ZFT beamforming employed at the relay side, respectively, we have derived closed-form lower bound for the achievable rate, based on which the asymptotic analysis has been provided for two different cases under massive arrays configurations. Then, according to the derived capacity lower bound, power allocation schemes have also been proposed under certain practical power constraints. Simulation results have shown that optimal power allocations schemes outperform equal power allocation in various scenarios. It has also been found that the asymptotically optimal power solutions for MRC/MRT and ZFR/ZFT achieve almost the same performance as OPA when the SNR is high and the number of antennas at the relay is large. Both AOPA and OPA outperform EPA on the spectral efficiency. Besides, the allocated power of each user in AOPA is inverse to the large-scale fading factor of the channel from the user to the relay, and proportional to the channel from the relay to its destination for MRC/MRT under the condition that the link end-to-end large-scale fading factors among

all pairs are equal, and the AOPA for ZFR/ZFT tends to be EPA when the number of antennas at the relay is large.

- **MSE-based Precoding for MIMO Downlinks in Heterogeneous Networks**

  We have developed three new MSE-based precoding schemes for MIMO downlinks in a HetNet architecture consisting of a macro tier overlaid with a second tier of SCs. Based on the same sum-MSE minimization problem, the first two precoding schemes have been presented focusing on the joint design of a set of BS and SC transmit precoding matrices or vectors by minimizing the total user MSE with individual transmit power at each cell constrained. Besides, a separate MSE minimization based two-level precoder has been proposed by a non-iterative algorithm, where BD technique is employed as its first-level precoder and each cell designs its own second-level precoder separately. There is no need to exchange user data or channel state information over the backhaul. Furthermore, robust precoding schemes have been proposed on the basis of the estimated imperfect CSI. Simulation results have shown that the sum-MSE based RAO algorithm offers better performance on the MSE at the cost of considerable complexity in comparison to UAON and the separate MSE-based precoding. When the number of antennas at the macro-BS is large enough relative to the number of MUEs, the average MSE of the low complexity separate MSE-based precoding comes close to those of RAO and UAON. Moreover, the UAON algorithm is a worthy trade-off between efficiency and performance, as it has higher convergence rate and lower computation complexity compared to RAO.

- **Performance Analysis for Pilot-Reused HetNets with Large-Scale Antenna Arrays in Downlink Systems**

  We have investigated the performance of a two-tier network with large-scale antenna arrays set at both BS and SCs. The main contributions can be summarized as follows: First, we have derived capacity lower bounds with closed-form expressions for both imperfect CSI based MRT and ZFT precoding employed at each node, where a pilot reuse pattern has been utilized for the channel estimation procedure to obtain the estimated CSI, followed by asymptotic analyses. Second, we have presented two different kinds of user scheduling algorithms, GSA and ASA, to further improve the system performance without requirement of instantaneous CSI exchange between BS and SCs. Simulation results

have shown that the derived closed-form expressions for the achievable rate are accurate predictors of the system performance for both MRT and ZFT precoding, and that more pilot reuse is able to yield higher effective performance under different system configurations even though pilot contamination exists. Furthermore, the effectiveness of the proposed user scheduling algorithms has been demonstrated in comparison to the corresponding random scheduling algorithm.

5.2 Further Research Issues

There are many open issues for further research related to the topics in the dissertation.

- It is well known that pilot overhead is proportional to the number of user equipment for the conventional orthogonal training scheme, i.e., the system performance will degrade as the UE number grows due to heavy pilot overhead. In a two-tier HetNet with multiple small cells, massive antenna arrays and large number of UEs, we propose to apply pilot reuse among the SCs in Chapter 4, i.e., the same set of orthogonal pilots is reused among the small cells in one macro-cell. Thus the number of orthogonal pilots is smaller than the total UE number in the whole network. Simulation results demonstrate the effectiveness of the proposed pilot reuse method, but the introduced pilot contamination effect becomes the bottleneck of the performance improvement. Hence, we could focus on the pilot decontamination scheme design in the future work.

- Most of the existing work on pilot contamination in massive MIMO systems are focused on the traditional homogeneous networks. In Chapter 4, we have considered a single-cell setting with a two-tier architecture and performed analysis with MRT precoding utilized at each node. However, it is only reasonable with orthogonality between the pilots in different cells guaranteed. In practical cellular networks, channel coherence times are not long enough due to mobility to allow for such long training sequences. While the multi-cell setting for HetNet with large-scale antenna arrays set at both BS and small cells has not been studied yet. Therefore, a multi-cell setting with hexagonal structure with both MRT and ZFT precoding could be employed at each node in the future work. And our goal is to propose pilot reuse strategies and complete performance analysis for pilot contamination in the multi-cell scenario. Moreover, we could propose user scheduling algorithms for multi-cell model and do the corresponding evaluation.

Publications

Published papers:

Y. Dai and X. Dong, "Power allocation for multi-pair massive MIMO two-way AF relaying with linear processing," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 5932–5946, Jun. 2016.

Y. Dai, X. Dong, and X. Dong, "Bi-directional cooperative relay strategies for transmitted reference pulse cluster UWB systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4512–4524, Oct. 2015.

Y. Dai and X. Dong, "Hybrid PPM-BPSK for transmitted reference pulse cluster systems in UWB and 60 GHz channels," *IEEE Wireless Commun. Lett.*, vol. 6, no. 3, pp. 657–660, Oct. 2014.

Y. Dai and X. Dong, "Downlink performance and user scheduling of HetNet with large-scale antenna arrays," In 2015 *IEEE GLOBECOM*, pp. 1–6, Dec. 2015.

Y. Dai, L. Pan, and X. Dong, "Physical-layer network coding aided bi-directional cooperative relays for transmitted reference pulse cluster UWB systems," In 2014 *IEEE ICC*, pp. 5825–5830, Jun. 2014.

L. Pan, **Y. Dai**, W. Xu, and X. Dong, "A novel block-shifted pilot design for multipair massive MIMO relaying," In 2016 *EUSIPCO*, Aug. 2016.

L. Liang, **Y. Dai**, W. Xu, and X. Dong, "How to approach zero-forcing under RF chain limitations in large mmWave multiuser systems?" In 2014 *IEEE/CIC ICCC*, Best Paper, pp. 518–522, Aug. 2014.

Submitted papers:

Y. Dai and X. Dong, "Performance of heterogeneous networks with large-scale antenna arrays in the downlink systems," **submitted to** *IEEE Trans. Commun.*, Jul. 2016.

Y. Dai, X. Dong, and W. S. Lu, "MSE-based precoding for MIMO downlinks in heterogeneous networks," **submitted to** *IEEE Trans. Veh. Technol.*, Aug. 2016.

L. Pan, **Y. Dai**, W. Xu, and X. Dong, "Multipair massive MIMO relaying with pilot-data transmission overlay," **submitted to** *IEEE Trans. Wireless Commun.*, Jun. 2016.

Y. Shao, **Y. Dai**, T. A. Gulliver, and X. Dong, "Precoding for MIMO full-duplex relay communication systems," **submitted to** *IEEE Trans. Wireless Commun.*, Jul. 2016.

Finished manuscripts:

Y. Dai, L. Pan, and X. Dong, "Physical-layer network coding aided two-way cooperative relays for transmitted reference pulse cluster UWB systems," Jun. 2014.

L. Pan, **Y. Dai**, X. Dong, and T. Lu, "Multipair two-way decode-and-forward relaying with physical layer network coding in massive MIMO systems," Dec. 2015.

Bibliography

[1] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

[2] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 379–389, Feb. 2007.

[3] T. J. Oechtering and H. Boche, "Bidirectional regenerative half-duplex relaying using relay selection," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1879–1888, 2008.

[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 J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

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

[6] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, Feb. 2013.

[7] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[8] H. Q. Ngo and E. G. Larsson, "Large-scale multipair two-way relay networks with distributed AF beamforming," *IEEE Commun. Lett.*, vol. 17, no. 12, pp. 2288–2291, Dec. 2013.

[9] H. A. Suraweera, H. Q. Ngo, T. Q. Duong, C. Yuen and E. G. Larsson, "Multi-pair amplify-and-forward relaying with large antenna arrays," in *Proc. IEEE Inter. Conf. Commun. (ICC'13)*, pp. 3228-3233, Jun. 2013.

[10] H. Cui, L. Song, and B. Jiao, "Multi-pair two-way amplify-and-forward relaying with very large number of relay antennas," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2636–2645, May 2014.

[11] M. Liu, J. Zhang, and P. Zhang, "Multipair two-way relay networks with very large antenna arrays," in *Proc. IEEE Veh. Tech. Conf. (VTC'14)*, pp. 1–5, Sept. 2014.

[12] S. Jin, X. Liang, X. Gao, K. Wong, and Q. Zhu, "Ergodic rate analysis for multipair massive MIMO two-way relay networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1480–1491, Mar. 2015.

[13] H. Q. Ngo, H. A. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, Sept. 2014.

[14] K. T. Phan, T. L. Ngoc, S. A. Vorobyov, and C. Telambura, "Power allocation in wireless multi-user relay networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2535–2545, May 2009.

[15] D. Nguyen and H. Nguyen, "Power allocation in wireless multiuser multi-relay networks with distributed beamforming," *IET Commun.*, vol. 5, no. 14, pp. 2040–2051, Mar. 2011.

[16] Q. Cao, Y. Jing, and H. V. Zhao, "Power allocation in multi-user wireless relay networks through bargaining," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2870–2882, Jun. 2013.

[17] Y. Dai and X. Dong, "Power allocation for multi-pair massive MIMO two-way AF relaying with linear processing," *IEEE Trans. Wireless Commun.*, vol. 99, no. 99, May 2016.

[18] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951-963, Apr. 2003.

[19] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 3640–3651, Aug. 2011.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[21] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming (web page and software)," *http://stanford.edu/boyd/cvx*, Feb. 2008.

[22] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Englewood Cliffs, NJ: Prentice Hall, 1993.

[23] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[24] Y. -C. Liang and R. Zhang, "Optimal analogue relaying with multi-antennas for physical layer network coding," in *Proc. IEEE Inter. Conf. Commun. (ICC'08)*, pp. 3893–3897, May 2008.

[25] T. Cover and J. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[26] P. C. Weeraddana, M. Codreanu, M. Latva-aho, and A. Ephremides, "Resource allocation for cross-layer utility maximization in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2790–2809, Jul. 2011.

[27] X. Wu, W. Xu, X. Dong, H. Zhang, and X. You, "Asymptotically optimal power allocation for massive MIMO wireless powered conmmunications," *IEEE Wireless Commun. Lett.*, Nov. 2015.

[28] H. Cramér, *Random Variables and Probability Distributions*, Cambridge, UK: Cambridge University Press, 1970.

[29] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Commun. Inf. Theory*, vol. 1, pp. 1–182, June 2004.

[30] N. H. Timm, *Applied Multivariate Analysis*, Springer, 2002.

[31] P. Graczyk, G. Letac, and H. Massam, "The complex Wishart distribution and the symmetric group," *Annals of Statistics*, vol. 31, pp. 287-309, 2003.

[32] 3GPP, TR 36.814, v9.0.0,"Evolved Universal Terrestrial Radio Access (E-UTRA), Further Advancements for E-UTRA Physical Layer Aspects," Mar. 2010.

[33] S. Brueck, "Heterogeneous networks in LTE-Advanced," in *Intern. Symp. Wireless Commun. Sys.* (*ISWCS'11*), pp. 171–175, Nov. 2011.

[34] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun. Mag.*, vol. 18, no. 3, pp. 10-21, June 2011.

[35] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, 2010.

[36] J. Zhu and H. Yang, "Interference control with beamforming coordination for two-tier femtocell networks and its performance analysis," in *Proc. IEEE Inter. Conf. Commun. (ICC'11)*, June 2011.

[37] Y. Dai, S. Jin, L. Pan, X. Gao, L. Jiang, and M. Lei, "Interference control based on beamforming coordination for heterogeneous network with RRH deployment," *IEEE Sys. J.*, vol. 99, no. 99, pp. 1–7, Apr. 2013.

[38] Z. Xu, C. Yang, G. Y. Li, Y. Liu, and S. Xu, "Energy-efficient CoMP precoding in heterogeneous networks," *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 1005-1017, Feb. 2014.

[39] H. Shen, B. Li, M. Tao, and X. Wang, "MSE-based transceiver designs for the MIMO interference channel," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3480–3489, Nov. 2010.

[40] R. Wang, M. Tao, and Y. Huang, "Linear precoding designs for amplify-and-forward multiuser two-way relay systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4457–4469, Dec. 2012.

[41] R. Wang and M. Tao, "Joint source and relay precoding designs for MIMO two-way relaying based on MSE criterion," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1352–1365, Mar. 2012.

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

[43] J. Hoydis, K. Hosseini, S. T. Brink, and M. Debbah, "Making smart use of excess antennas: massive MIMO, small cells and TDD," *Bell Labs Tech. J.*, vol. 18, no. 2, pp. 5-21, 2013.

[44] R. Zhang, "Cooperative multi-Cell block diagonalization with per-base-station power constraints," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1435–1445, 2010.

[45] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA), Further Advancements for E-UTRA Physical Layer Aspects," Mar. 2010 [Online]. Available: ftp.3gpp.org, TR 36.814 v9.0.0.

[46] 3GPP, R1-112433, NTT DOCOMO,"Scenarios and potential CSI feedback enhancements for DL MIMO in Rel-11," Aug. 2011.

[47] 3GPP, R1-112047, Huawei, HiSilicon,"Investigation on CSI feedback enhancements for closed-loop MIMO," Aug. 2011.

[48] Y. Lin, W. Bao, W. Yu, and B. Liang, "Optimizing user association and spectrum allocation in HetNets: a utility perspective," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1025–1039, June 2015.

[49] A. Adhikary, E. A. Safadi, and G. Caire, "Massive MIMO and inter-tier interference coordination," in *Proc. Inf. Theory and its Applications (ITA'14)*, Feb. 2014.

[50] K. Hosseini, J. Hoydis, S. Brink and M. Debbah, "Massive MIMO and small cells: how to densify heterogeneous networks," in *Proc. IEEE Int. Conf. Commun. (ICC'13)*, pp. 5442–5447, 2013.

[51] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large MIMIO channels," in *Proc. IEEE Veh. Tech. Conf. (VTC'11)*, pp. 1–5, Sept. 2011.

[52] "Massive-element antenna for small cell solutions in 5G," *White Paper* by NEC Coorperation, Feb. 2016.

[53] A. Adhikary, H. S. Dhillon, and G. Caire, "Massive-MIMO meets HetNet: interference coordination through spatial blanking," *arXiv preprint arXiv:1407.5716v1*, 2014.

[54] Y. Dai and X. Dong, "Downlink performance and user scheduling of HetNet with large-scale antenna arrays," in *Proc. IEEE Global Commun. Conf.* (*GLOBECOM'15*), pp. 1–6, Dec. 2015.

[55] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.