Partial Zero-Forcing Design and Achievable Rate Analysis for MIMO Relay Networks

by

Samira Rahimian

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Communications

Department of Electrical and Computer Engineering
University of Alberta

Abstract

The ever-increasing demand for wireless communication networks has led an evolution in this technology. The next generation which is responsible for significantly higher data rates, energy efficiency and coverage is the 5th Generation (5G). 5G has introduced many infrastructural changes, hardware innovations, and signal processing techniques to enable high quality-of-service, energy, and bandwidth efficiency. This thesis is intended to investigate the achievable rate of relay networks considering new techniques and concepts in 5G and propose novel signal processing methods to pave the path for more efficient and greener telecommunications. We have accomplished three projects on signal processing design and performance analysis of 5G relay networks.

For multi-way relay networks (MWRNs), where multiple users mutually exchange information, we propose a novel relay processing design called partial zero-forcing which combines zero-forcing processing at the relay and successive interference cancellation at the users in the decoding process. The invented method provides extra degrees-of-freedom in the relay signal processing compared to zero-forcing, thus can significantly improve the system rate performance. We also propose a new optimization method called modified gradient-ascent based on the gradient-ascent method to maximize the achievable rate using the extra degrees-of-freedom.

The second work is on massive multiple-input multiple-output (mMIMO) MWRNs. We consider the practical energy-saving and low-cost mMIMO systems with low-resolution analog-to-digital converters (ADCs). A novel mathe-

matical framework is proposed to derive a closed-form expression for the achievable rate of the network using random matrix theory including properties of Wishart and Haar matrices. This method can be applied and extended to other scenarios of mMIMO systems.

In the third work, we testify the novel application of simultaneous wireless information and power transfer (SWIPT) in a mMIMO relay network. Further, instead of the traditional linear antenna array, the case of planar antenna array at the relay is studied to exploit the elevation angle in the antennas beam pattern, a technique called 3D beamforming. We conduct performance analysis of the network, and then, the optimization of the 3D beam pattern for the maximum achievable rate. Significant rate improvement is achieved due to the extra degrees-of-freedom provided by SWIPT and 3D beamforming.

Preface

The results presented in Chapter 3 were published in IEEE Transactions on Communications in April 2018. The results of Chapter 4 were presented in the IEEE International Conference on Communications (ICC 2019) and also published in IEEE Transactions on Wireless Communications in June 2020. The results of Chapter 5 have been submitted to the Physical Communication Journal of Elsevier in July 2020.

Dedicated to
my beloved parents and lovely husband, Hadi,
who have been a constant source of support and encouragement during the challenges of this journey.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my co-supervisors, Profs. Yindi Jing and Masoud Ardakani. Their guidance and support were invaluable during my PhD study. This past four years have been a very formative and constructive part of my life, both professionally and personally. This to a large extent, is due to to them. They were always ready to answer my questions and give me a wider perspective of the subject. Also, their professionalism and enthusiasm influenced me with lifetime benefits.

I am also thankful to the members of my examining committee, Profs. Chintha Tellambura, Majid Khabbazian, and Jun Cai for their valuable comments and suggestions on improving this thesis.

Further, I wish to thank all my colleagues and friends at the UoA for sharing their time and expertise with me. In particular, Fudong Li for support on laboratory-related activities, and Sara Eghbali and Parvaneh Saffari for the inspirational coffee breaks. They have all made my time at the UoA an enjoyable experience.

I want to thank Alberta Innovates for providing me with the funding through my first three years. I am also thankful to National Sciences and Engineering Research Council of Canada and Telus Communications Inc for their support in the last year of this research. Further appreciation goes to FGSR, IEEE Communications Society, and GSA for their financial support in the form of travel grants that enabled me to attend an international conference.

Finally, I would like to express my gratitude to my parents for their constant love, support, and encouragement throughout my life that gives me the strength to step forward and reach my goals. Last but not least, my deepest gratitude goes to my husband, Hadi, for his unconditional love and support during the ups and downs of the PhD journey while experiencing the same odyssey himself. Thank you for always being there for me.

Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Definition
5G	The 5th Generation
3D	Three-Dimensional
3GPP	The 3rd Generation Partnership Project
ADC	Analog-to-Digital Converter
AF	Amplify-and-Forward
AQNM	Additive Quantization Noise Model
AWGN	Additive White Gaussian Noise
BC	Broadcast
BS	Base Station
CSI	Channel State Information
DAC	Digital-to-Analog Converter
dBi	Decibel Isotropic
DC	Direct Current
DF	Decode-and-Forward
EB	Exabyte
EH	Energy Harvesting
GHz	Gigahertz
ID	Information Decoding
i.i.d.	Independent and Identically Distributed
IoT	Internet-of-Things
LTE	Long Term Evolution
MF	Matched Filter
MIMO	Multiple-Input Multiple-Output
mMIMO	Massive Multiple-Input Multiple-Output
MMSE	Minimum Mean-Squared Error
MR	Maximum Ratio
MRC/MRT	Maximum Ratio Combining/Maximum Ratio Transmission
MSE	Mean-Squared Error
MWRN	Multi-Way Relay Network
OFDM	Orthogonal Frequency-Division Multiplexing

Abbreviation	Definition
pdf	probability density function
PS	Power Splitting
PZF	Partial Zero-Forcing
Q1/Q2/Q3/Q4	The 1st/2nd/3rd/4th Quarter of Year
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
RF	Radio Frequency
RZF	Regularized Zero-Forcing
SER	Symbol Error Rate
SINR	Signal-to-Interference-Plus-Noise Ratio
SLL	Side Lobe Level
SNR	Signal-to-Noise-Ratio
SVD	Singular-Value Decomposition
SWIPT	Simultaneous Wireless Information and Power Transfer
TS	Time Switching
ZF	Zero-Forcing

Chapter 1

Introduction

The dramatic growth of the number of subscriptions to the Internet and cellular systems, besides the rapid development of data-intensive wireless applications, such as online gaming, high quality voice over Internet Protocol, and video streaming, have led the wireless technology to undergo an evolution. According to recent Ericsson mobility report [1], in Q3 2019, mobile data traffic grew 68 percent year-on-year. In general, traffic growth is being driven by both the rising number of smartphone subscriptions and an increasing average data volume per subscription, fueled primarily by more viewing of video content. Figure 1.1 shows the total global monthly data and voice traffic from Q1 2014 to Q3 2019, along with the year-on-year percentage change for mobile data traffic. By the end of 2025, the total mobile data traffic is expected to grow by a factor of 6 of 2014, reaching 160 EB/month, with near 9 billion subscriptions. It is expected that nearly half of this traffic will be carried by the next generation of wireless technology, the 5th Generation (5G) networks.

The 5G system is envisioned to have significantly higher data rates, energy efficiency, reliability, coverage, and security in addition to less latency, error rate, and hardware cost than the current long term evolution (LTE) system. Researchers around the world have proposed different infrastructural changes, hardware, and signal processing enhancements to make 5G wireless technology a reality [2]. Many key technologies have been proposed to help achieve the 5G requirements, including densification and offloading, cooperative networking, millimeter wave, large scale antenna arrays or massive multiple-input multiple-output (mMIMO) systems, three-dimensional (3D) beamforming, and simultaneous wireless information and power transfer (SWIPT) [2]. This thesis is mainly focused on cooperative relay networks, or in short relay networks, mMIMO, 3D beamforming, and SWIPT. In the following section, we briefly introduce each of these technologies.

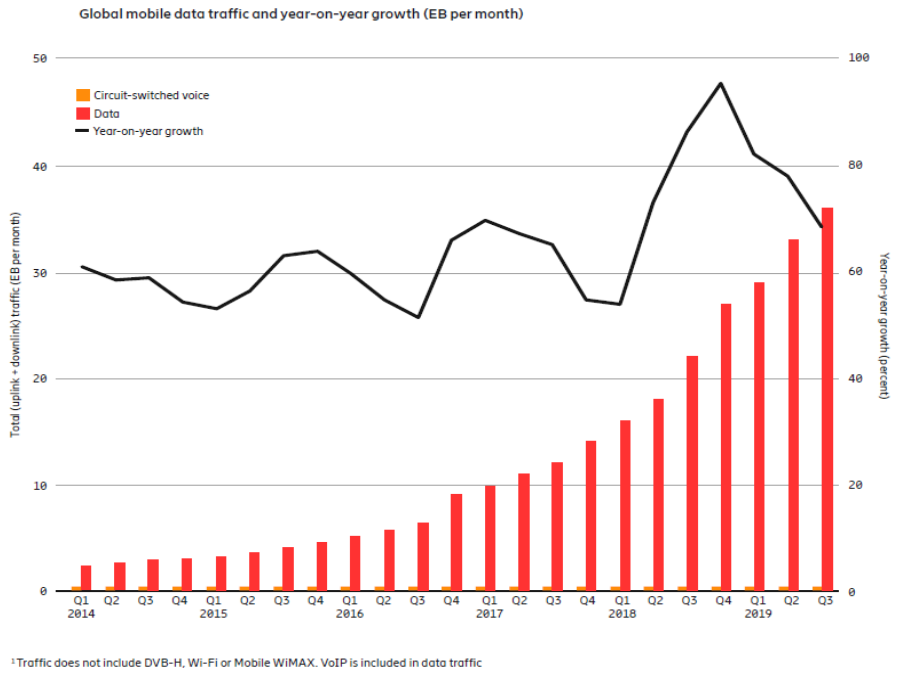


Figure 1.1: Total global monthly data and voice traffic from Q1 2014 to Q3 2019, and year-on-year percentage change for mobile data traffic [1].

1.1 Several 5G Key Technologies

In this section, we introduce the 5G technologies including relay networks, mMIMO, 3D beamforming, and SWIPT.

Relay Networks

When there is no reliable link between a number of source and destination users, one way to make the communications possible is employing cooperative intermediate nodes, called relays [3], that help with the communications. Therefore, relays extend the coverage of wireless networks. Usually, a relay station is employed to assist the communication between the users and the base station (BS) or just between the users. Compared to a BS, a relay station has a smaller size, less transmit power, and wireless backhaul connections. Hence, relays are less expensive and much easier to be employed in networks [3]. In particular, advanced cellular relays in 5G cellular networks enhance the topology of the cellular system by improving the robustness, decreasing power consumption, and providing ubiquitous coverage even for users in far and less populated areas or at the cell edge of a cellular network [4]. Additionally, the

multi-hop structure of relays can efficiently support the access of large numbers of 5G machine-to-machine terminals [5]. In the following, more details are presented on how relays can help in 5G systems.

- Cell-edge performance: In cellular systems, cell-edge users suffer from severe path-loss because of the long distance from the BS. Usually, to assure the quality-of-service (QoS) of those users, very high transmit power is used. In a heterogeneous cellular network, relay stations can be located in cells [5] which will extend the coverage and reduce the transmit power of both the users and BS [3].

- Mobile communications: Users in a mobile vehicle, e.g., a moving train, suffer from severe penetration-loss. Employing a relay station on the moving vehicle can mitigate the penetration-loss effect [5]. Moreover, the transmit power of users will get reduced as the relay is located closer than the BS.

- Device-to-device (D2D) communications: The D2D communication is an important part of heterogeneous cellular networks that makes the internet-of-things (IoT) possible [5]. In D2D systems, devices communicate with each other with no help from the BS. D2D networks reduce the traffic congestion on the core networks and improve the energy and spectral efficiency. When there is no reliable direct link between users, a relay station can be used [6]. Sometimes the nodes between source and destination users can play the relay role and cooperate in the communications [3].

Traditionally several single-antenna relays were designed to cooperate as a group to help the communication between a pair of source and destination users [7, 8]. However, with the advent of the concept of multiple-input multiple-output (MIMO) technology, they integrated into single central multi-antenna relays which reduce the large amount of handshaking overhead among single-antenna relays. Nowadays, central MIMO relay networks are being employed [9]. Depending on the number of users from which the relay can simultaneously receive signals, relay networks are divided into three categories of one-way, two-way, and multi-way. As its name suggests, in a one-way network the relay

only receives signal from one specific user. While in two-way and multi-way networks, the relay receives signals from two and more than two users, respectively. Two-way relay networks are more spectral efficient than one-way ones and multi-way relay networks (MWRNs) are the most efficient ones.

Massive MIMO

Systems with multiple antennas implemented at the transceivers are referred to as MIMO or multi-antenna systems. MIMO systems exploit the spatial diversity to provide high data rate and link reliability. In conventional MIMO systems, the number of antennas is usually moderate (e.g., up to 8 antennas for the LTE standard) [10, 11]. Recently, as one of the key technologies for 5G, large-scale MIMO or mMIMO systems where antenna arrays with hundreds of antenna elements are implemented at the transceivers have attracted a lot of attention, as they bring enormous data rate and energy efficiency improvements [12]. It is shown that mMIMO has the following advantages over conventional MIMO [2, 13].

- Reduced effect of small-scale fading: Because of the large spatial diversity provided by mMIMO, the randomness of communication channels caused by the small-scale fading effect can be reduced or averaged out by using simple processing techniques. In fact, all small-scale randomness abates as the number of channel observations grows. Thus, the effect of small-scale fading can be eliminated.

- Reduced interference: With a large number of antennas, the random communication channels between different users and the BSs become quasi-orthogonal to each other. Thus, the interference can be effectively reduced with simple signal processing techniques.

- Higher energy efficiency: With a large number of antennas, signal energy can be focused with extreme sharpness into small areas of desire and thus less waste of energy occurs during communications.

All the above mentioned features makes mMIMO a perfect candidate for 5G networks [2, 5].

One limiting factor of mMIMO systems is the dramatic increase in the hardware cost and energy consumption as the number of antennas increases

[14]. Recently, massive antenna arrays with low-resolution analog-to-digital converters (ADCs) have attracted lots of research interests [15] to help solve this problem.

ADCs create digital values to represent the original analog signal. Generally, in communication systems, each receive antenna is equipped with a radio frequency chain including a pair of ADCs for the quantization of the in-phase and quadrature components of the received signal [16]. While higher resolution ADCs have smaller quantization distortion and improve the communication performance, they also consume more energy and require more memory space. In fact, the power consumption increases exponentially with the quantization bit value [17]. Thus, mMIMO technique with low-resolution ADCs is an important candidate for 5G systems.

3D Beamforming

Conventional BSs usually feature linear horizontal arrays, which can only accommodate a limited number of antennas in tower structures due to form factors, and which only exploit the azimuth angle dimension [2]. Employing planar two-dimensional arrays and further exploiting the elevation angle, the so called 3D beamforming, help accommodate many more antennas with the same form factor [18]. As a side benefit, well-designed vertical beams increase the signal power and reduce interference to users in neighboring cells [2]. Thus, 3D beamforming is a promising technology in 5G networks. Initial practical results on the applications and improvements that 3D beamforming can bring are provided in [19, 20]. The channel model for 3D beamforming needs to incorporate elevation in addition to azimuth, i.e., should be a 3D channel model [18, 21, 22]. Only little data exists concerning power spectra and angle spreads on the elevation dimension. A 3D channel modeling study currently under way within 3GPP is expected to shed light on these various issues [23].

SWIPT

With the advent of IoT [24], an emerging solution for prolonging the lifetime of energy constrained relays is SWIPT which is a critical enabler of green communications in 5G networks. In a SWIPT-enabled system, the received signal is divided into two distinct parts: one part is used for energy harvesting (EH) and the other for information decoding (ID) [25]. An efficient technique for systems with co-located receivers for information and energy signals is power

splitting (PS) which splits the power of the received signal into two parts for EH and ID [26]. The other SWIPT reception technique is time switching (TS) where the duration of data reception is divided into two separate parts for EH and ID [27].

Other 5G Technologies

Other 5G technologies include millimeter wave communications, cognitive radio, and non-orthogonal multiple access which are briefly introduced in the following. Interested readers can find further 5G technologies in [2, 5].

Electromagnetic waves of frequencies below 3 GHz, called microwave, support desirable propagation characteristics. Almost all commercial wireless communications, such as radio broadcast, cell phone system, and satellite communications, work in this area. However, to meet the increasing traffic demands, electromagnetic waves with higher frequencies of 3-300 GHz, called millimeter wave, are now being exploited in the recent 5G systems [28]. The change to higher frequencies leads to higher attenuation and propagation loss which should be taken care of in the new systems designs.

Another way to efficiently use the available spectrum is cognitive radio. In cognitive radio [29], when licensed users do not use their bands, they can pass the spectrum to unlicensed users. Thus, a band is shared between users, rather than divided among them, making it easier for the operators to cope with temporary peaks in traffic.

Another candidate for the spectral efficiency improvement in 5G systems is non-orthogonal multiple access where multiple users are multiplexed in the power domain at the transmitter and at the receivers decoding is conducted based on successive interference cancellation [30]. Therefore, in NOMA-based networks, the achievable rate is affected by the power allocation at the transmitter and the decoding order at the receivers [31, 32].

1.2 Motivations and Contributions

Of the 5G requirements, the most important ones include the need for extremely higher data rates (1000x) and decreasing the energy and cost (10-100x) [2]. In this thesis, we study relay networks accompanied by the above introduced 5G key technologies to make the data rate, energy, and cost requirements of 5G reachable. This thesis follows two main goals; the first goal is to design

a novel relay beamforming scheme that brings higher spectral and energy efficiencies than the current beamforming schemes, and the second one is the performance analysis of relay networks considering the key 5G technologies. In this regard, three different problems are formulated and solved as the following.

1. **Beamforming Design for MIMO MWRNs**
   MIMO MWRNs allow a number of users to mutually exchange information providing high spectral efficiency; thus, they are promising candidates for 5G nested small cells [33], such as picocells (range under 100 meters) and femtocells (WiFi-like range) [34]. In the first contribution, we propose a novel beamforming design that improves the spectral and energy efficiency of MWRNs compared to existing beamforming designs. The new beamforming design is called partial zero-forcing (PZF). Zero-forcing (ZF) relay beamforming forces the interference from all interfering users to be zero [35]. In our design, the user-interference is carefully and jointly coordinated to be nullified in the broadcast (BC) phase by combining ZF beamforming with self-interference cancellation and successive interference cancellation at the users. This combination allows more degrees-of-freedom in the beamforming design compared to ZF. The sum-rate optimization problem is formulated under the PZF constraints in the BC phase. A numerical method which is a modification of the gradient-ascent optimization method is proposed to find the solution. Our numerous simulation results show that the proposed PZF beamforming leads to significantly higher sum-rates than ZF and linear beamforming schemes in [35, 36].

2. **Massive MIMO MWRNs with Mixed-Resolution ADCs**
   Massive MIMO MWRNs combine both mMIMO and MWRN technologies and are expected to inherit the advantages of both making them competitive candidates for the 5G wireless industry [36, 37], especially in macrocells. High power consumption and hardware cost have motivated the use of low-resolution ADCs for practical mMIMO systems. In this thesis, we consider mMIMO MWRNs with low-resolution ADCs. To the best of our knowledge, there had been no work on the performance analysis of mMIMO MWRNs with low-resolution ADC structure. In this research, we examine the general mixed-ADC receiver architec-

ture in mMIMO MWRNs with ZF beamforming for both reception and transmission under both perfect and imperfect channel state information (CSI). Using the results from random matrix theory, we derive a closed-form asymptotic approximation for the ergodic achievable rate of each pair of users. Our derived expression reveals the performance behavior of the system with respect to different system parameters, such as the number of antennas at the relay, the ADC resolution profile, the number of users in the system, the user and relay transmission power, and the CSI error.

3. **Performance Analysis and Optimization of 3D Massive MIMO Multi-Pair Relaying with SWIPT**
   Given the growing importance of 3D beamforming and SWIPT in stretching the durability of energy limited 5G IoT relay networks [24], we study a SWIPT-enabled mMIMO multi-user one-way relay network with planar two-dimensional arrays. There has been no existing work on the performance analysis and optimization of such networks. Under the PS protocol at the relay and considering maximum ratio combining/maximum ratio transmission (MRC/MRT) and matched filter (MF) beamforming, we derive closed-form expressions that serve as lower-bounds on the average signal-to-interference-plus-noise ratio (SINR). Based on that, asymptotic average achievable sum-rate expressions are obtained. Joint optimization problems over the relay PS ratio and antenna array tilt to maximize the average achievable sum-rate are formulated and solved. Monte-Carlo simulation results are presented to verify our theoretical analysis. Further, the simulations present the gains that the optimized setups bring.

1.3 Organization of Thesis

The remaining of this thesis is organized as follows. In Chapter 2, the necessary background on relaying schemes, beamforming designs, mMIMO, 3D channel model, and properties of special random matrices is given. Chapter 3 is focused on our first problem. This chapter presents the design steps for PZF beamforming and our two optimization methods based on gradient-ascent to find beamforming solutions. Further, the computational complexity of the PZF designs is compared with the conventional beamforming schemes. Sim-

ulation results are provided for the comparison of sum-rate performance of PZF scheme with the existing schemes considering different transmission scenarios. Chapter 4 is on our second problem. In this chapter, we introduce our novel method to find a closed-form approximate expression for the sum-rate of mMIMO MWRNs with low-resolution mixed-ADCs. Further, the effect of different structures of low-resolution mixed-ADCs and CSI error on the rate performance is presented through simulations. Chapter 5 is focused on the third problem, where performance analysis on 3D mMIMO multi-pair relaying with PS based SWIPT is provided. Further, an optimization is performed on the elevation angle and PS ratio to maximize the sum-rate. We also compare the sum-rate results for MRC/MRT and MF beamforming schemes through simulations. Finally, Chapter 6 concludes the thesis and proposes possible directions for future work.

1.4 Notations

In this thesis, bold upper case letters and bold lower case letters are used to denote matrices and vectors, respectively. For an arbitrary matrix $\mathbf{A}$, the transpose, conjugate, Hermitian, inverse, Moore-Penrose pseudoinverse and trace are denoted by $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$, $\mathbf{A}^+$ and $\mathrm{tr}\{\mathbf{A}\}$ respectively. Also, a_{ij} and $\mathbf{a}_i$ denote the (i,j)th entry and the ith column of $\mathbf{A}$. For vector $\mathbf{a}$, $\|\mathbf{a}\|$, and $|\mathbf{a}|$, denotes the 2-norm or Frobenius norm, and Euclidean norm, respectively. Also, diag$\{\mathbf{a}\}$ denotes a diagonal matrix with the elements of $\mathbf{a}$ as its diagonal entries. The $N \times N$ identity matrix reads as $\mathbf{I}_N$.

Also, $\mathrm{mod}_N(x)$ denotes x modulo N. When a is much less than b it is shown by $a \ll b$. Notations $\mathbb{E}\{\cdot\}$, and $\mathrm{Var}\{\cdot\}$ show the expectation and variance operators, respectively. Set of complex numbers is shown by $\mathbb{C}$. For a complex variable, $\Re\{.\}$ and $\Im\{.\}$ denote the real and imaginary parts, respectively. Gaussian and circularly symmetric complex Gaussian random variables with mean m and variance σ^2 are shown by $x \sim \mathcal{N}(m, \sigma^2)$, and $x \sim \mathcal{CN}(m, \sigma^2)$, respectively. The exponential and Gamma functions are denoted by $\exp(\cdot)$ and $\Gamma(\cdot)$, respectively. Logarithm function with base 2 is shown by $\log_2(\cdot)$. Finally, $\mathrm{erf}(\cdot)$ is the error function defined as $\mathrm{erf}(x) \triangleq \frac{1}{\sqrt{\pi}} \int_{-x}^{x} \exp(-t^2)dt$.

Chapter 2

Background

In this chapter, we review some of the existing works related to our contributions to provide the necessary background to follow this thesis. We first review important relaying schemes and introduce famous beamforming schemes. Then, the necessary details on quantization are provided, followed by an elaboration on the 3D channel model. At last, some important results on special random matrices are provided.

2.1 Relay Networks

This thesis is mainly focused on a central MIMO relay network where a single relay in the network is equipped with multiple antennas. Early research on the relay networks has been focused on single-user networks, e.g., [38–40]. However, nowadays, in order to respond to the high demands of simultaneous communications of multiple users and bandwidth shortage problem, multi-user communication is commonly employed [41]. In this thesis, a multi-user network is assumed where there are multiple users intended to have mutual communications.

Relays either operate in the half-duplex mode or the full-duplex mode, depending on whether they can receive and transmit simultaneously over the same frequency band. Traditionally, relays operate in half-duplex mode where the communication channels from the users to the relay are orthogonal to the reverse channels due to time division multiplexing. On the contrary, in the full-duplex mode, the relay and the users share a common time-frequency signal-space, so that the relay can transmit and receive simultaneously over the same frequency band. Full-duplex relaying brings higher spectral efficiency; however, it is hard to implement [42, 43]. In this thesis, the relays are assumed to operate in the half-duplex mode. A simple illustration of our considered relay network is depicted in Figure 2.1. In the following, we elaborate on other different

categories of central MIMO relay networks.

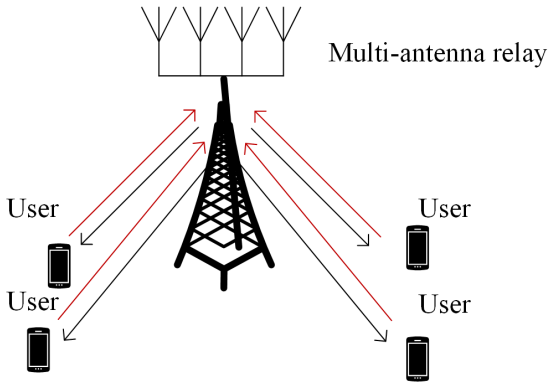


Figure 2.1: A MIMO multi-user relay network where users are having mutual communications in a half-duplex mode. The red and black arrows indicate the signal transmission directions from the users to the relay and the reverse, respectively.

2.1.1 One-Way and Two-Way Relay Networks

In conventional MIMO relay networks, messages are transmitted in one direction from the source users to the relay and then to the destination users, thus, they are called one-way relay networks. However, one-way relaying may not be efficient when two or more users communicate mutually. Borrowing the concept of network coding from [44], the next version of relay communications is two-way relaying, where the relay receives messages from the source and destination users simultaneously, and then broadcasts the summed messages back to them. Using its own message, each user is able to decode its intended message [45]. Since the mutual communication time is divided by two, two-way relaying is twice spectral efficient as one-way relaying.

In the following, assuming that there is no direct link between a pair of users, schematic examples of one-way and two-way relay networks are depicted. In Figure 2.3, a one-way relay communication is shown, where the users u_1 and u_2 are having a mutual communication by the help of the relay. The numbers assigned to the arrows show the order of events happening in the unit of time. In time slot 1, u_1 transmits its signal to the relay, then in time slot 2, the relay transmits u_1's signal to u_2. In time slots 3, and 4, u_2 transmits its

signal to the relay, and then, the relay transmits u_2's signal to u_1. Thus, for a mutual communication between a pair of users, it takes 4 time slots in a one-way relay network. Figure 2.3 shows a two-way relay network. In time slot 1, u_1 and u_2 simultaneously transmit their signals to the relay, and then in time slot 2, the relay transmits its received signal, which is the summation of the two users' signals back to the users. After that, each user is able to decode the other user's signal using the information of its own signal. So, in two-way relay communications, it only takes two time slots for a mutual communication between a pair of users, which is half of what it takes in one-way relay networks. Thus, two-way scheme is twice spectral and energy efficient and one-way relay networks are more suitable for unidirectional communications rather than mutual communications between the users. These setups can be extended to multi-pair networks with multiple pairs of source and destination users [46, 47].

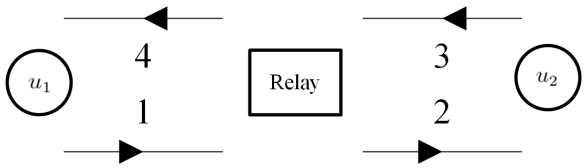


Figure 2.2: A one-way relay communication between u_1 and u_2 through the relay. The numbers assigned to the arrows, show the sequence of transmissions.

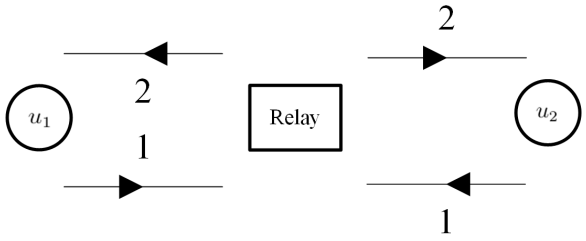


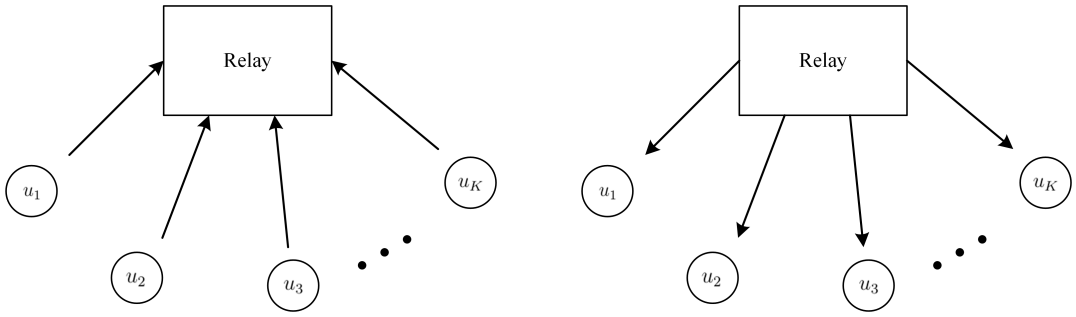
Figure 2.3: A two-way relay communication between u_1 and u_2 through the relay. The numbers assigned to the arrows, show the sequence of transmissions.

2.1.2 Multi-Way Relay Networks (MWRNs)

Multi-way relaying scheme is the recent version of relay networks and an extension of two-way relaying scheme [37]. By smartly leveraging user-interference, instead of completely avoiding it, MWRNs are able to significantly reduce

the number of time slots for full mutual communications among users which consequently improves the spectral and energy efficiencies [36, 37, 48–50]. In MWRNs, multiple interfering users exchange messages among themselves through a relay, such that each user multicasts its message to all the other users. Potential applications of multi-way communications include cellular and D2D communications, teleconferencing via satellites, wireless sensor networks, video conferences, multi-player online games, and generally wherever mutual communications among multiple users through a common relay happens [51, 52]. Early studies in MWRNs are mainly on networks with a single-antenna relay [37, 49, 51, 53, 54]. The performance of MWRNs can be further improved by employing multiple antennas at the relay [35, 55–60], which is the case considered in this thesis.

Consider a MWRN of K single-antenna users, $u_1, u_2, \cdots, u_K$, communicating with each other with the help of a half-duplex relay equipped with N antennas. Each user needs to communicate its information signal to all other $K-1$ users and receives their signals.



(a) The MAC phase. (b) A BC time slot.

Figure 2.4: Transceiver protocol of a MWRN.

Communications in multi-antenna MWRNs take two phases, one is the multiple access (MAC) phase and the other is the BC phase. For all users to send one symbol each to all other users, the MAC phase contains 1 time slot and the BC phase contains $K-1$ time slots, so in total K time slots are required. In the MAC phase, as shown in Figure 2.4a, all users transmit their information symbols simultaneously to the relay. Then, the BC phase starts where in each BC time slot, as shown in Figure 2.4b, the relay applies some linear signal processing to its received signal in the MAC phase. Then, the relay broadcasts the processed signal to all users. In each BC time slot, the

signal to be broadcast, changes in a way that after the BC phase every user has received the information symbol from all other users.

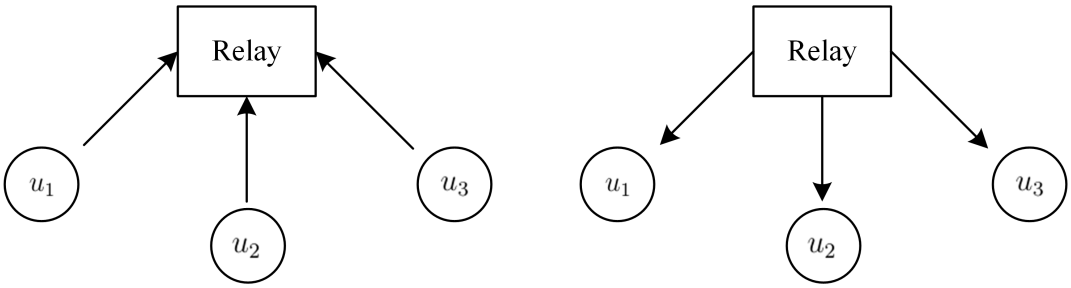
To better illustrate the protocol, a 3-user MWRN is shown in Figure 2.5. It takes the following steps for a full mutual communication.

1. In time slot 1 (MAC Phase), all users transmit their signals simultaneously to the relay.

2. In time slot 2 (first BC time slot), the relay transmits to all users, and u_1 decodes information of u_2, u_2 decodes information of u_3, and u_3 decodes information of u_1.

3. In time slot 3 (second BC time slot), the relay transmits to all users, and u_1 decodes information of u_3, u_2 decodes information of u_1, and u_3 decodes information of u_2.

After the BC phase, each user has decoded the information symbols from all other users.



(a) The MAC phase. (b) A BC time slot.

Figure 2.5: Transceiver protocol of a 3-user MWRN.

In MWRNs, each user's common rate is defined as the rate by which the user can reliably send information to all other users throughout the BC phase. In other words, each user's common rate is equal to the minimum transmission rate of the user through the BC time slots. Based on the users' common rates, the achievable sum-rate of a MWRN is defined as the summation of the users' common rates over the whole communication time slots.

In all MIMO relay networks, the transmission of signals from the source to destination users happens in three steps: 1) the source users transmit their

signals to the relay, 2) the relay processes the received signals, and 3) the relay transmits the processed signals to the destination users. The signal processing at the relay, or relaying scheme, is an important aspect in relay communication systems. Various relaying schemes and protocols have been studied [38, 61]. Two important ones are amplify-and-forward (AF) and decode-and-forward (DF).

- AF protocol: AF is a type of non-regenerative relaying [62], where the relay does not decode the received information from the source users and only performs some linear processing [63, 64] before forwarding it to the destination users.
- DF protocol: DF is a type of regenerative relaying where the relay first decodes the information of source users from the received signals, and then, it may apply some coding or linear processing on the decoded signals and retransmits the results to the destination users.

Compared to DF, AF is simpler, costs less to implement, and has less processing delay. Also, it avoids error propagation from the relay to the destination users. However, it causes noise and interference amplification, which causes performance degradation [3]. Moreover, in DF scheme, when the destination users decode the information, they only need to know the information of their own channel from the relay, thus, the resources used on channel estimation are reduced.

2.2 Beamforming Schemes

Beamforming is a signal processing technique that is used by the transmitters or receivers in order to form the beam directions considering certain goals on the signal to be transmitted or received. In a MIMO relay network, the relay serves as both the receiver (for the MAC phase) and transmitter (for the BC phase), thus its beamforming schemes include both receive and transmit beamforming, hence, called transceive beamforming.

The transmit and receive beamforming matrices are usually designed based on the CSI. In [35, 36], three well-know linear relay transceive beamforming designs named ZF, MRC/MRT sometimes referred to as MF, and minimum

mean-squared error (MMSE) are proposed. In the following, a brief introduction to the three beamforming designs is given.

- ZF: The principle idea of the ZF scheme is to fully nullify the multi-user interference. That is, in the t-th BC time slot the interference from all other users excepting u_i, i.e., the intended transmitter to u_k, is forced to zero at u_k. In ZF beamforming, first, the signals received at the relay antennas are combined by multiplying the received signal vector with the pseudoinverse of the users to relay channel matrix. Then, the combined signals are precoded by multiplying the combined signal vector with the pseudoinverse of the relay to users channel matrix. ZF beamforming is one of the most popular beamforming designs that brings high performance in MWRNs, especially for the high SNR range [65].

- MRC/MRT: This beamforming aims at maximizing the power of the desired signal. In this scheme, the signals received from all relay antennas are combined following the MRC rule which guarantees that the signals from the same source but received by different antennas can be combined coherently. This is done by multiplying the received signal vector with the Hermitian of the channel matrix of the MAC phase. Then, the combined signals are precoded following the MRT rule which is designed in a way that the intended signals received from different channels are added together coherently at the target destinations. This is done by MRT precoding, where the combined signal vector is multiplied with the conjugate of the MAC phase channel matrix before transmitted to the destinations.

  MF beamforming scheme has similar beamforming matrices to MRC/MRT, but with row-normalization. In the MF scheme, the same weight or power is allocated to all channels, while in the MRC/MRT scheme, strong channel gains are given higher weights or more transmit power.

- MMSE: The idea behind this beamforming technique is to minimize the mean-squared error (MSE) of the desired signals at the receiving users.

2.2.1 Beamforming for MWRNs

For a MWRN with K single-antenna users where the relay is equipped with N antennas, we denote the $K \times N$ receive beamforming matrix at the relay by

$\mathbf{G}_{\mathrm{RX}}$ and the $N \times K$ transmit beamforming matrix of the t-th BC time slot at the relay by $\mathbf{G}_{\mathrm{TX}}^{(t)}$, for $t \in \{1, 2, \cdots, K-1\}$. Further, we denote the $N \times N$ transceive beamforming matrix of the t-th BC time slot by $\mathbf{G}^{(t)}$. The matrix $\mathbf{G}^{(t)}$ has the following general structure [35],

$$\mathbf{G}^{(t)} = \mathbf{G}_{\mathrm{TX}}^{(t)} \mathbf{P}^t \mathbf{G}_{\mathrm{RX}}, \tag{2.1}$$

where $\mathbf{P}$ is obtained by circularly shifting the columns of identity matrix, $\mathbf{I}_K$, one position to the right. $\mathbf{P}^t$ is the permutation matrix that defines the relationship between an arbitrary receiving user, u_k, and the corresponding transmitting user, u_i, in the t-th BC time slot. In this thesis, the order of decoding is defined via the following relation among i, k, t:

$$i(k,t) \triangleq \mathrm{mod}_K(k+t-1)+1. \tag{2.2}$$

This is a function of the receiving user's index, k and the time slot, t. To help the presentation, this is simplified to $i(k)$ when there is no confusion.

We denote the channel matrix from the users to the relay by $\mathbf{H} \in \mathbb{C}^{N \times K}$. By assuming channel reciprocity, the channel matrix from the relay to the users equals $\mathbf{H}^T$. In the following, ZF, MMSE, and MRC/MRT beamforming designs for MWRNs are presented.

In ZF, $\mathbf{G}^{(t)}$ is designed such that the interference from all other users except $u_{i(k)}$, is forced to zero at u_k. $\mathbf{G}_{\mathrm{RX}}$ and $\mathbf{G}_{\mathrm{TX}}^{(t)}$ are defined as the following [66]:

$$\begin{aligned} \mathbf{G}_{\mathrm{RX}} &= (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H, \\ \mathbf{G}_{\mathrm{TX}}^{(t)} &= \frac{1}{p_{\mathrm{ZF}}^{(t)}} \mathbf{H}^* (\mathbf{H}^T \mathbf{H}^*)^{-1}, \end{aligned} \tag{2.3}$$

where $p_{\mathrm{ZF}}^{(t)}$ is used to fulfill the relay transmission power constraint.

MMSE beamforming minimizes the MSE of the signal. For MWRNs, the MMSE receive and transmit beamforming matrices [67] are:

$$\begin{aligned} \mathbf{G}_{\mathrm{RX}} &= \mathbf{R}_{\mathbf{x}} \mathbf{H}^H (\mathbf{H} \mathbf{R}_{\mathbf{x}} \mathbf{H}^H + \mathbf{I}_N)^{-1}, \\ \mathbf{G}_{\mathrm{TX}}^{(t)} &= \frac{1}{p_{\mathrm{MMSE}}^{(t)}} (\mathbf{H}^H \mathbf{H} + \frac{\mathbf{I}_N}{P_{\mathrm{R}}})^{-1} \mathbf{H}^H, \end{aligned} \tag{2.4}$$

where $\mathbf{R}_{\mathbf{x}} = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\}$ is the covariance matrix of the transmitted signal vector from the users to the relay, $\mathbf{x}$, P_{R} is the relay transmit power, and $p^{(t)}_{\mathrm{MMSE}}$ is used to fulfill the relay transmission power constraint. It is worth mentioning that regularized zero-forcing (RZF) beamforming [68] is a modification of MMSE. RZF replaces $\mathbf{I}_N$ in the MMSE receive beamforming formula ($\mathbf{G}_{\mathrm{RX}}$ in (2.4)) with $\alpha\mathbf{I}_N$ where α is the regularization coefficient.

MRC/MRT beamforming, is the optimal linear beamforming for maximizing the signal-to-noise-ratio (SNR) in the presence of additive noise. The MRC receive beamforming and MRT transmit beamforming matrices are:

$$\begin{aligned} \mathbf{G}_{\mathrm{RX}} &= \mathbf{H}^H, \\ \mathbf{G}^{(t)}_{\mathrm{TX}} &= \frac{1}{p^{(t)}_{\mathrm{MRC/MRT}}}\mathbf{H}^*, \end{aligned} \tag{2.5}$$

where $p^{(t)}_{\mathrm{MRC/MRT}}$ is used to fulfill the relay transmission power constraint.

In [67], the authors defined the following mixed design of MMSE and MRT as MF beamforming:

$$\begin{aligned} \mathbf{G}_{\mathrm{RX}} &= \mathbf{R}_{\mathbf{x}}\mathbf{H}^H(\mathbf{H}\mathbf{R}_{\mathbf{x}}\mathbf{H}^H + \mathbf{I}_N)^{-1}, \\ \mathbf{G}^{(t)}_{\mathrm{TX}} &= \frac{1}{p^{(t)}_{\mathrm{MF}}}\mathbf{H}^H, \end{aligned} \tag{2.6}$$

where $\frac{1}{p^{(t)}_{\mathrm{MF}}}$ is used to fulfill the relay power constraint. Notice that in the ZF, MRC/MRT, MF, and MMSE multi-way relaying schemes, other than the coordination of the decoding order which happens through the permutation matrix, beamforming in different BC time slots are designed separately instead of jointly. In addition to the above mentioned beamforming schemes, there are works on the joint design of the relay beamforming and user processing, e.g., [58–60].

2.3 ADCs and Quantization

The main function of an ADC is the quantization. Quantization is the process of converting an analog signal into a digital one. Consider an ADC with b bits of resolution, where b can take any positive integer value. The quantization of this ADC can be characterized by a set of $2^b + 1$ quantization

thresholds $\mathcal{T}_b = \{\tau_0, \tau_1, \cdots, \tau_{2^b}\}$, where $\tau_0 < \tau_1 < \cdots < \tau_{2^b}$, and a set of 2^b quantization labels $\mathcal{L}_b = \{l_0, l_1, \cdots, l_{2^b-1}\}$, where $l_i \in (\tau_i, \tau_{i+1}]$. The b-bit quantization operation is described by the function $\mathcal{Q}_b(\cdot) : \mathbb{R} \to \mathcal{L}_b$. Let s be a real analog signal to be quantized. Then, the quantization process $\hat{s} = \mathcal{Q}_b(s)$ maps s to $\hat{s}$ in a way that

$$\hat{s} = l_k, \quad \text{if } s \in (\tau_k, \tau_{k+1}].$$

In general, the optimal design of the sets of $\mathcal{T}_b$ and $\mathcal{L}_b$ that minimizes some error measure between the analog input signal and the quantizer output depends on the distribution of the input signal. In this thesis, we use Lloyd-Max [69, 70] algorithm to design the quantization thresholds and labels.

In Lloyd-Max algorithm, the optimization of the sets of labels and thresholds is performed with the goal to minimize the MSE between the analog input signal and the quantizer output, MSE is the mean of the square of the difference between s and $\hat{s}$. In this quantizer, the first and last thresholds are the minimum and maximum values of the range of the analog signal, respectively, while the rest of the thresholds are set to be the middle point of adjacent labels. Further, the labels are located on the centroid point between two adjacent thresholds. Denote the probability density function (pdf) of the quantizer input s by $f_S(s)$. The iterative Lloyd-Max quantization design algorithm has the following steps.

1. Assign an initial set of labels $\{l_0, l_1, \cdots, l_{2^b-1}\}$.
2. Calculate the thresholds using $t_q = \frac{1}{2}(l_{q-1} + l_q)$, for $q \in \{1, \cdots, 2^b - 1\}$.
3. Calculate the new labels using $l_q = \frac{\int_{t_q}^{t_{q+1}} s f_S(s) ds}{\int_{t_q}^{t_{q+1}} f_S(s) ds}$ for $q \in \{0, \cdots, 2^b - 1\}$.
4. Repeat items number 2 and 3, until no further reduction in MSE occurs.

2.3.1 Quantization Noise Model

When low-cost, low-resolution ADCs are used, the quantization error and its effect on the communication performance are non-negligible. In general, quantization of a signal using finite-resolution ADC causes a distortion that is correlated with the input to the quantizer. However, in most researches the quan-

tization error is approximated by an independent additive quantization noise model (AQNM) [71] for the sake of simplicity.

Another quantization noise model is based on Bussgang's decomposition theorem [72, 73]. When the input to a quantizer is Gaussian, the quantized signal can be decomposed into a term that is a linear function of the analog input and a distortion term that is uncorrelated to the analog input. Specifically, let s be a zero-mean Gaussian signal with variance v, $\hat{s}$ be the quantization output, and d be the quantization distortion, then, $\hat{s}$ can be written as

$$\hat{s} = \mathcal{Q}_b(s) = G_b s + d, \tag{2.7}$$

where G_b is the quantization coefficient and can be calculated as

$$G_b = \frac{1}{\sqrt{\pi v}} \sum_{i=0}^{2^b-1} l_i \left[\exp\left(-\frac{\tau_i^2}{v}\right) - \exp\left(-\frac{\tau_{i+1}^2}{v}\right) \right]. \tag{2.8}$$

To understand the effect of low-resolution ADCs on the communication over Gaussian channels, a general theoretical framework is presented by [74] which evaluates the mutual information under various distortion models. Several authors have considered mMIMO communication systems with low-resolution ADCs and investigated their performance. Among them, in [72], a tight approximation is obtained for the uplink achievable rate of mMIMO systems with MRC where the quantization error is modeled based on Bussgang's decomposition. A tight approximation for the uplink spectral efficiency is derived in [71] for the MRC receiver where all the ADCs have the same resolution, referred to as the uniform-ADC structure. A mixed-ADC receiver architecture is firstly proposed in [75] where the ADCs have multiple resolution levels. The spectral efficiency for the uplink of a two-level mixed-ADC mMIMO system is analyzed in [76]. For more detailed review of literature on this topic please refer to Section 4.1.

2.4 3D Channel Model

As mentioned in the previous chapter, planar antenna arrays lead to a 3D channel model which gives the opportunity to exploit the array elevation angle for a better performance. A simplified presentation of the 3D channel model,

or 3D antenna pattern, of a relay station is presented in Figure 2.6. The relay antenna array is located in the plane parallel to the ground and all antennas transmit precoded information signals with adjustable weights. The antenna array tilt can be tuned by adapting the weights. We assume that all antennas have a common tilt and approximate the antenna pattern by the 3D model of 3GPP in [77]. The observed gain from the relay antenna array at the kth user is expressed in dBi scale as the following

$$A_k^{\mathrm{dBi}}(\theta_{\mathrm{tilt}}) = -\Big(\min\Big[12\Big(\frac{\phi_k}{\phi_{3\mathrm{dB}}}\Big)^2, \mathrm{SLL}_{\mathrm{az}}\Big] + \min\Big[12\Big(\frac{\theta_k - \theta_{\mathrm{tilt}}}{\theta_{3\mathrm{dB}}}\Big)^2, \mathrm{SLL}_{\mathrm{el}}\Big]\Big), \tag{2.9}$$

where $0 < \theta_{\mathrm{tilt}} < \pi/2$ is the tilt angle between the horizontal line and the antenna array beam, θ_k is the angle between the horizon and the line connecting the user to the relay antenna array, ϕ_k is the angle between the X-axis and the line in the horizontal plane that connects the user to the projection point of the relay on the horizontal plane. Further, the side lobe levels (SLLs) of the antenna array pattern in the horizontal and vertical planes are set as $\mathrm{SLL}_{\mathrm{az}} = 25$ dB and $\mathrm{SLL}_{\mathrm{el}} = 20$ dB [77], respectively. Also, the half power beamwidth in the horizontal and vertical planes are $\phi_{3\mathrm{dB}} = 65°$ and $\theta_{3\mathrm{dB}} = 6°$ [77], respectively. Finally, it is assumed that the relay beam peak is fixed on $\phi = 0$ relative to the X-axis.

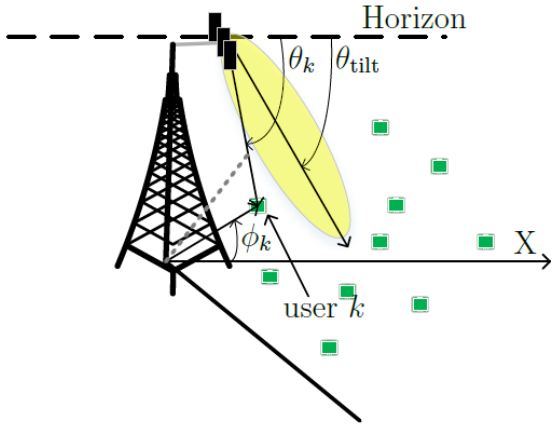


Figure 2.6: The green squares represent the users. The spherical angles of the kth user are illustrated. The coverage area in the horizontal plane spans an angular range of 120°.

2.5 Random Matrix Theory

Random matrix theory has been massively influenced by its applications in physics, statistics and engineering since its inception. Nowadays, random matrices find applications in various fields, such as stochastic differential equations, condensed matter physics, chaotic systems, numerical linear algebra, neural networks, multivariate statistics, information theory, and signal processing [78]. Specifically, with the emergence of mMIMO in wireless communications, the urge for analytical methods and results on the properties of random matrices with growing dimensions has made random matrix theory an important tool for the study of communications systems performance limits [79]. In most other methods than random matrix theory for the analysis of communications systems, the carrier frequency should be less than 6 GHz which is unlikely for millimeter wave communications. In this regard, recently, researchers use random matrix theory to analyze information theoretical problems associated with communications systems. For example, [80, 81] use it for the exploitation of antenna correlation diversity, and analyzing the effects of hardware impairments, respectively. Further, [82–84] use random matrix theory to design low complexity receivers or precoders. In this thesis, we use random matrix theory to conduct our performance analysis of mMIMO systems. Specifically, we employ results on Haar and Wishart random matrices which are explained in more details in the following.

2.5.1 Haar Matrix

In linear algebra, a complex square matrix $\mathbf{U}$ is unitary if $\mathbf{U}^{-1} = \mathbf{U}^H$, or

$$\mathbf{U}^H\mathbf{U} = \mathbf{U}\mathbf{U}^H = \mathbf{I}. \tag{2.10}$$

Unitary matrices are the complex analogy of real orthogonal matrices. For any unitary matrix $\mathbf{U}$, the following properties hold [85]:

- $\mathbf{U}^H$ is unitary.
- The rows and columns of $\mathbf{U}$ form an orthonormal basis with respect to the inner product determined by $\mathbf{U}$.

- Given two complex vectors $\mathbf{x}$ and $\mathbf{y}$, multiplication by $\mathbf{U}$ preserves their inner product; that is, $< \mathbf{Ux}, \mathbf{Uy} >=< \mathbf{x}, \mathbf{y} >$.
- $\mathbf{U}$ is a normal matrix, $\mathbf{U}^H\mathbf{U} = \mathbf{U}\mathbf{U}^H$, with eigenvalues lying on the unit circle.

A random matrix that is uniformly distributed on the manifold of unitary matrices is called Haar measure, also referred to as Haar matrix [78]. An important application of Haar matrices is their appearance in the singular-value decomposition (SVD) of Gaussian matrices. For instance, let $\mathbf{H} \in \mathbb{C}^{N\times K}$ be an $N \times K, N \geq K$ Gaussian matrix with independent and identically distributed (i.i.d.) entries of $\mathcal{CN}(0, 1)$. Then, consider the SVD

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \tag{2.11}$$

where $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{\Sigma}$ are $N \times N$, $N \times K$, and $K \times K$ matrices. $\mathbf{U}$ and $\mathbf{V}$ contain the left and right singular vectors of $\mathbf{H}$, respectively, and $\mathbf{\Sigma}$ contains the singular values of $\mathbf{H}$, $\{\sigma_1, \sigma_2, \cdots, \sigma_K\}$. According to Definition 2.5 in [78], $\mathbf{U}$, and $\mathbf{V}$ are Haar matrices. The following lemma is provided on the properties of Haar matrices.

Lemma 1. *If $1 \leq i, j, i', j' \leq N$, $i \neq i', j \neq j'$, and $\mathbf{U}$ is an $N \times N$ Haar matrix, then the following hold [86].*

$$1)\ \mathbb{E}(|u_{i,j}|^2) = \frac{1}{N}, \quad 2)\ \mathbb{E}(|u_{i,j}|^4) = \frac{2}{N(N+1)},$$
$$3)\ \mathbb{E}(|u_{i,j}|^2|u_{i',j}|^2) = \mathbb{E}(|u_{i,j}|^2|u_{i,j'}|^2) = \frac{1}{N(N+1)},$$
$$4)\ \mathbb{E}(|u_{i,j}|^2|u_{i',j'}|^2) = \frac{1}{N^2-1}, \quad 5)\ \mathbb{E}(u_{i,j}u_{i',j'}u^*_{i,j'}u^*_{i',j}) = -\frac{1}{N(N^2-1)}.$$

All other multiple moments up to the fourth order are zero.

2.5.2 Complex Wishart and Inverse Wishart Matrices

In the performance analysis of wireless communications systems, Wishart and inverse Wishart matrices also often appear. Assume that $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K$ are $N\times 1$, i.i.d. complex Gaussian vectors, i.e., $\mathbf{x}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$ for $i \in \{1, 2, \cdots, K\}$. Then, the $N \times N$ random matrix $\mathbf{W} = \sum_{k=1}^{K} \mathbf{x}_k\mathbf{x}_k^H$ is central complex Wishart

distributed with K degrees-of-freedom and parameter matrix $\mathbf{\Sigma}$; $\mathbf{W} \sim \mathcal{W}_N^C(K, \mathbf{\Sigma})$. The pdf of $\mathbf{W}$ and the characteristic function of its elements are known [87]. Further, if $K \geq N$, the $N \times N$ matrix $\mathbf{W}^{-1}$ is central complex inverse Wishart distributed with K degrees-of-freedom and parameter matrix $\mathbf{\Sigma}^{-1}$; $\mathbf{W}^{-1} \sim \mathcal{W}_N^{C^{-1}}(K, \mathbf{\Sigma}^{-1})$. In [78], the pdf of complex inverse Wishart distribution is calculated. As an example, considering the same $\mathbf{H}$ as in the example of above subsection, $\mathbf{H}^T\mathbf{H}^*$ and $\mathbf{H}^H\mathbf{H} \sim \mathcal{W}_K^C(N, \mathbf{I})$ are $K \times K$ central Wishart matrices of N degrees of freedom and their inverse matrices are central inverse Wishart matrices of N degrees of freedom. In the following lemma, we bring some important properties of central Wishart and inverse Wishart matrices that are used in this thesis.

Lemma 2. *If* $1 \leq i, j, l, k \leq N$, *and* $\mathbf{W} \sim \mathcal{W}_N^C(K, \mathbf{\Sigma})$, $K \geq N$ *then the following hold [88].*

1. $\mathbf{W}$ *is Hermitian.*
2. $\mathbb{E}\left[\mathbf{W}_{ij}\right] = \mathbf{\Sigma}_{ij}$
3. $\mathbb{E}\left[\left(\mathbf{W}^{-1}\right)_{ij}\right] = \frac{\left(\mathbf{\Sigma}^{-1}\right)_{ij}}{K-N}$
4. $\mathbb{E}\left[\left(\mathbf{W}^{-1}\right)_{ij}\left(\mathbf{W}^{-1}\right)_{lk}\right] = \frac{\left(\mathbf{\Sigma}^{-1}\right)_{ij}\left(\mathbf{\Sigma}^{-1}\right)_{lk} + \frac{\left(\mathbf{\Sigma}^{-1}\right)_{lj}\left(\mathbf{\Sigma}^{-1}\right)_{ik}}{K-N}}{(K-N)^2-1}$

Chapter 3

Partial Zero-Forcing for Multi-Way Relay Networks

In this chapter, we present our novel linear beamforming design, PZF, for MIMO MWRNs. Compared to ZF, PZF relaxes the constraints on the relay beamforming matrix such that only partial user-interference, instead of all, is canceled at the relay. The users eliminate the remaining interference through self-interference and successive interference cancellation. A sum-rate maximization problem is formulated and solved to exploit the extra degrees-of-freedom resulted from PZF. Simulation results show that the proposed PZF relay beamforming design achieves significantly higher network sum-rates than the existing linear beamforming designs.

3.1 Introduction and Literature Review

In this section, we review the most recent works on the performance analysis and design of different beamforming schemes for MWRNs. Also, we introduce our contributions in this chapter.

Several research works on the performance analysis of single-antenna, MIMO, and mMIMO MWRNs with different relay beamforming schemes have been carried out. Considering a MWRN with a single-antenna relay, [37] provides upper bounds on the common rate of symmetric Gaussian MWRNs and calculates the achievable symmetric rate for AF and DF relaying schemes. Further, other studies, e.g. [51] and [54], focus on improving the achievable data rates of MWRNs with single-antenna relays by suggesting new relaying approaches and scheduling the users' transmission order.

As mentioned in Chapter 2, employing multiple antennas at the relay improves the performance of MWRNs. In this regard, for three different relaying scenarios, called unicasting, multicasting, and hybrid uni/multicasting, linear

relay transceive beamforming designs based on ZF, MMSE, and MF are proposed in [35] and [36]. In another study, the situation when the CSI is not available at the relay is investigated [35]. For this case, the authors use space-time analog network coding transmission strategy for stationary channels, and repetition transmission strategy for non-stationary channels. In [52, 56], closed-form approximations for the spectral and engery efficiencies of MIMO and mMIMO MWRNs with ZF beamforming and perfect CSI are obtained. Further in [65, 89], for mMIMO MWRNs achievable rate analysis is performed under imperfect CSI for MR and ZF processing at the relay. It is shown in [52, 90], that if MRC/MRT or ZF is used in a mMIMO relay network, the transmit power of each user or relay (or both) can be made inversely proportional to the number of relay antennas while maintaining a given QoS. Thus, high power efficiency can be achieved with MRC/MRT or ZF at the relay.

Another relaying scenario, namely superimposed uni/multicasting, is reported in [58], which efficiently combines the MMSE beamforming at the relay with joint receive processing at the users. More specifically, by carefully designing the selection of uni/multicast signals at the relay and the interference cancellation order at the users, the proposed strategy improves the system sum-rate. The authors in [60], have designed joint relay beamforming and receiver processing matrices to maximize the minimum received SINR at the users. For the receiver processing, MRC, and ZF are considered. However, the proposed iterative algorithm can have high computational complexity, especially when there exists a large number of users and/or antennas.

In this work, similar to [35, 36], we consider a MIMO MWRN with beamforming design at the relay. Each user individually decodes only the information that is intended for that user. We consider an AF MWRN where a half-duplex relay equipped with N antennas helps K single-antenna users to receive information from each other. The goal of our work is to maximize the achievable sum-rate of the users. To this end, we introduce the novel idea of PZF. Unlike ZF relay beamforming, where in each relay BC time slot the interference from all interfering users is forced to be zero [35], our proposed PZF only forces partial interference (the interference from a carefully designed subset of the interfering users) to be zero. Thus, PZF allows more degrees-of-freedom in the relay beamforming design. With the help of self-interference cancellation

and successive interference cancellation at the users, the proposed PZF relay beamforming allows each user to obtain interference-free information from all other users.

The sum-rate maximization problem which is a constrained non-linear multi-dimensional optimization problem is formulated based on the PZF idea. We propose a numerical method, called modified gradient-ascent method, to jointly optimize the PZF relay beamforming matrices for all BC time slots. In addition, to reduce the computational complexity, another method to separately optimize the relay beamforming matrices of different BC time slots is proposed. Simulation results verify the significant sum-rate improvement that the proposed PZF design achieves compared to the existing ZF, MMSE, and MF beamforming designs in [35] and [36]. For example, for a homogeneous 3-user MWRN, where all channels from the users to the relay have the same large-scale fading, we report between 14% to 200% sum-rate improvements compared to ZF, MMSE, and MF schemes. In comparison to [60], we report slightly lower sum-rates, but it should be noted that our system models are different. Unlike [60], we do not allow joint information decoding at the users or joint relay beamforming and receiver processing. Although, these amendments can improve the performance, but this improvement comes at a high computational complexity cost. Also, its significantly higher processing requirement at the users and in the beamforming optimization stage can make it less attractive for most of the applications. The interesting observation in our work is that with a relatively simple PZF beamforming a significant sum-rate gain can be achieved.

3.2 System Model

The system model of MWRNs includes two parts: the network model and the transceiver protocol, which will be elaborated in the following two subsections.

3.2.1 Network Model

We consider a MWRN consisting of K single-antenna users, $u_1, u_2, \cdots, u_K$, and a multi-antenna relay equipped with N antennas. We assume that $N \geq K$ to make a fair comparison between the existing ZF, MMSE, and MF beamforming

designs which need to have enough degrees-of-freedom to cancel user interferences [35], [36]. The extension to the case of $N = K - 1$ will be considered in Section 3.6.

All users and the relay operate in the half-duplex mode. There are no direct links between the users and users communicate with each other with the help of the relay. Denote the channel vector between u_i and the relay by $\mathbf{h}_i = (h_{1,k}, h_{2,k}, ..., h_{N,k})^T$ for $i = 1, 2, ..., K$. Hence, the $N \times K$ channel matrix between the users and the relay is $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_K]$. Channels are assumed to follow independent frequency-flat Rayleigh fading, where $h_{n,k}$ follows $\mathcal{CN}(0, \beta_k)$. Which implies that the channels between the same user and different antennas at the relay have the same variance, while the channels between different users and the relay antennas may have different variances. Further, we assume channel reciprocity and that channels do not change in each communication block of K time slots.

3.2.2 Communication Protocol

As explained in Section 2.1.2, for all users to have a mutual communication with each other, K time slots are needed, including 1 MAC time slot and $K - 1$ BC time slots. As shown in Figure 2.4a, in the MAC phase, all users simultaneously transmit their information symbols to the relay. We denote the normalized vector of users' information symbols as $\mathbf{x} \in \mathbb{C}^{K\times 1}$. The kth element of $\mathbf{x}$ corresponds to the signal of u_k normalized to have unit power, $\mathbb{E}_{x_k \in \mathcal{S}_k}\{|x_k|^2\} = 1$ for $k \in \{1, 2, \cdots, K\}$, where $\mathcal{S}_k$ is the modulation set for u_k. The average transmit power of each user is set to p_u which implies that all users have the same transmit power. So, the received signal vector at the relay, $\mathbf{r}_{\mathrm{R}} \in \mathbb{C}^{N\times 1}$, is

$$\mathbf{r}_{\mathrm{R}} = \sqrt{p_u}\mathbf{H}\mathbf{x} + \mathbf{z}_{\mathrm{R}} = \sum_{i=1}^{K} \sqrt{p_u}\mathbf{h}_i x_i + \mathbf{z}_{\mathrm{R}}, \tag{3.1}$$

where $\mathbf{z}_{\mathrm{R}} \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{I}_N)$ is the vector of additive white Gaussian noise (AWGN) at the relay. As shown in Figure 2.4b, in the BC phase, the relay applies linear beamforming to the received signal vector $\mathbf{r}_{\mathrm{R}}$ and broadcasts the information back to the users. For the t-th BC time slot, $t = 1, \cdots, K-1$, $\mathbf{G}^{(t)}$ denotes the $N \times N$ relay beamforming matrix. Each user considers the received symbols

from the relay other than its intended one as interference. The symbol transmitted from the relay to each user changes in different BC time slots, so that after $K-1$ BC time slots each user receives the information from other ones. In this section, for the simplicity of presentation, unicasting transmission [35] is assumed, where in each BC time slot, the relay transmits different information symbols to different users. Each symbol is intended only for one receiving user in each BC time slot. The extension to hybrid uni/multicasting will be explained in Section 3.5.

Due to the channel reciprocity, the channel matrix from the relay to the users equals $\mathbf{H}^T$ [52]. By using (3.1), in the t-th BC time slot, the received signal vector of users, $\mathbf{r}_{\mathrm{u}}^{(t)}$, will be

$$\mathbf{r}_{\mathrm{u}}^{(t)} = \sqrt{p_u}\mathbf{H}^T\mathbf{G}^{(t)}\mathbf{H}\mathbf{x} + \mathbf{H}^T\mathbf{G}^{(t)}\mathbf{z}_{\mathrm{R}} + \mathbf{z}_{\mathrm{u}}^{(t)}, \tag{3.2}$$

where $\mathbf{z}_{\mathrm{u}}^{(t)} = [{z_1}^{(t)}, {z_2}^{(t)}, ..., {z_K}^{(t)}]^T$ is the noise vector at the users in the t-th BC time slot. The additive noise at the users are modeled as independent random variables with $\mathcal{CN}(0,1)$ entries.

The relay transmit power in each BC time slot can be written as

$$P_{\mathrm{R}} = \mathbb{E}\{\mathrm{tr}\{\mathbf{G}^{(t)}(\sqrt{p_u}\mathbf{H}\mathbf{x} + \mathbf{z}_{\mathrm{R}})\left[\mathbf{G}^{(t)}(\sqrt{p_u}\mathbf{H}\mathbf{x} + \mathbf{z}_{\mathrm{R}})\right]^H\}\}. \tag{3.3}$$

After straightforward calculations, it can be simplified as

$$P_{\mathrm{R}} = \mathrm{tr}\Big\{\mathbf{G}^{(t)}(\mathbb{E}\left[p_u\mathbf{H}\mathbf{H}^H\right] + \mathbf{I}_N)(\mathbf{G}^{(t)})^H\Big\}, \tag{3.4}$$

In the t-th BC time slot, u_k decodes $u_{i(k)}$'s information symbol, $x_{i(k)}$, according to (2.2). Thus, according to (3.2), in the t-th BC time slot, the received signal at u_k is

$$r_k^{(t)} = \sqrt{p_u}\mathbf{h}_k^T\mathbf{G}^{(t)}\mathbf{h}_{i(k)}x_{i(k)} + \sum_{j=1, j\neq i(k)}^{K}\sqrt{p_u}\mathbf{h}_k^T\mathbf{G}^{(t)}\mathbf{h}_j x_j + \mathbf{h}_k^T\mathbf{G}^{(t)}\mathbf{z}_{\mathrm{R}} + z_k^{(t)}. \tag{3.5}$$

Notice that in each BC time slot, the signal transmitted by the relay contains signals of all users sent in the MAC phase. In (3.5), the first term contains the intended signal from $u_{i(k)}$, the second term contains the interference from other

users than the intended user (including the receiver's own signal x_k), which are all forwarded to the user by the relay, the third term contains the propagated noise from the relay, and the last term is the noise at u_k. So, the SINR of the communication from $u_{i(k)}$ to u_k, denoted as $\gamma_{k,i(k)}$, can be written as

$$\gamma_{k,i(k)} = \frac{p_u|\mathbf{h}_k^T\mathbf{G}^{(t)}\mathbf{h}_{i(k)}|^2}{\sum_{j=1,j\neq i(k)}^{K} p_u|\mathbf{h}_k^T\mathbf{G}^{(t)}\mathbf{h}_j|^2+|\mathbf{h}_k^T\mathbf{G}^{(t)}|^2+1}. \tag{3.6}$$

However, if after each BC time slot, u_k performs interference cancellation by subtracting its self-interference and the interference of user symbols which have already been decoded in the previous BC time slots. The SINR after interference cancellation is

$$\gamma_{k,i(k)} = \frac{p_u|\mathbf{h}_k^T\mathbf{G}^{(t)}\mathbf{h}_{i(k)}|^2}{\sum_{j=1,j\neq i(k),j\neq k,j\notin\mathbb{L}_{k,t}}^{K} p_u|\mathbf{h}_k^T\mathbf{G}^{(t)}\mathbf{h}_j|^2+|\mathbf{h}_k^T\mathbf{G}^{(t)}|^2+1}, \tag{3.7}$$

where $\mathbb{L}_{k,t} = \{\mathrm{mod}_K(k+q-1)+1, q = 1, 2, ..., t-1\}$. $\mathbb{L}_{k,t}$ contains the indexes of the symbols already decoded by u_k from previous $t-1$ BC time slots which is determined by the order of detection defined in (2.2). Hence, the achievable rate from $u_{i(k)}$ to u_k, denoted as $R_{k,i(k)}$, is

$$R_{k,i(k)} = \log_2(1+\gamma_{k,i(k)}). \tag{3.8}$$

The common rate R_i that u_i can reliably send to all other users is:

$$R_i = \min_{k\neq i} R_{k,i}. \tag{3.9}$$

The average achievable sum-rate of the MWRN is thus [35]:

$$R_{\mathrm{sum}} = \frac{K-1}{K}\sum_{i=1}^{K} R_i. \tag{3.10}$$

The sum-rate of MWRNs is given by (3.6)-(3.10). It can be seen that the design of relay beamforming matrices $\mathbf{G}^{(t)}$, $t = 1, \cdots, K-1$ is crucial for the sum-rate performance. In Subsection 2.2.1,we discussed the details on existing relay beamforming schemes including ZF, MMSE, and MF proposed in [35]. In the following section, we elaborate on the design of PZF beamforming.

3.3 PZF Design

In this section, we design PZF relay beamforming based on the ZF relying scheme. First, we explain the main idea behind PZF, then, we formulate the sum-rate maximization problem for PZF. A numerical method is proposed to solve the optimization problem. Finally, simulation results on the performance of PZF and the comparison with existing beamforming designs are provided.

In ZF relay beamforming [35], the beamforming matrices of all $K-1$ BC time slots are designed such that the effects of transmitted signals of all users except for the desired one are forced to be zero in the received signal at each user. For instance, if u_k wants to receive $u_{i(k)}$'s message in the BC time slot, t, all interference signals from $u_j, j \neq i(k)$, (all terms in (3.5) with $x_j, j \neq i(k)$) are forced to be zero by $\mathbf{G}_{\text{ZF}}^{(t)}$, $t = 1, \cdots, K-1$. This puts heavy constraints on each $\mathbf{G}_{\text{ZF}}^{(t)}$, where $K(K-1)$ entries of $\mathbf{H}^T\mathbf{G}_{\text{ZF}}^{(t)}\mathbf{H}$ must be zero as shown in (2.3). However, such heavy constraints are not necessary to obtain interference-free signals at the users.

As each user has the knowledge of its own information symbol and the CSI, it can perform self-interference cancellation. Further, at the t-th BC time slot, each user has already decoded the symbols of $t-1$ users, through the previous $t-1$ relay broadcasts, thus, it can cancel the interference from these users' symbols without further help from the relay. So, the t-th BC time slot relay beamforming matrix only needs to cancel the interference from the remaining $K-t-1$ users. This constraint relaxation, which we refer to as PZF, allows more degrees-of-freedom in the design of beamforming matrices to improve the network sum-rate.

For better illustration of the PZF idea and to help later analysis, we define

$$\mathbf{A}^{(t)} = \mathbf{H}^T\mathbf{G}^{(t)}\mathbf{H}, \tag{3.11}$$

which as seen from the system equation in (3.2), is the equivalent channel matrix of the t-th BC time slot. As shown in (2.3), in the ZF design, $\mathbf{A}^{(t)}$ should be equal to the permutation matrix $\mathbf{P}^{(t)}$ where $K(K-1)$ entries are zero and K entries are 1. However, in our PZF design, only $(K-t-1)K$ entries of $\mathbf{A}^{(t)}$ need to be zero and other entries can take any complex value. This is because

of the fact that according to (3.2), $\mathbf{H}^T\mathbf{G}^{(t)}\mathbf{H}\mathbf{s} = \mathbf{A}^{(t)}\mathbf{s}$. Thus, each entry of $\mathbf{A}^{(t)}\mathbf{s}$ contains $t-1$ previously detected symbols, self-interference, and the new symbols that are to be detected in the future. In PZF beamforming, the idea is to eliminate the interference from symbols to be detected in future via the relay beamforming matrix design and eliminate the interference from previously detected symbols and self-interference via direct interference cancellation at the users. This means that in each row of $\mathbf{A}^{(t)}$, our design requires having $K-(t+1)$ zero entries at predetermined locations, while the rest $t+1$ entries can take any complex value. So, in total for all the K rows, $\mathbf{A}^{(t)}$ matrix should have $(K-t-1)K$ zero entries and other entries can take any complex number.

Assume the MWRN with $N = K = 4$ as an example. If ZF beamforming is used at the relay, $\mathbf{G}^{(1)}$, $\mathbf{G}^{(2)}$ and $\mathbf{G}^{(3)}$ are designed so that $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$, and $\mathbf{A}^{(3)}$ have the following forms

$$\mathbf{A}_{\mathrm{ZF}}^{(1)} = \frac{1}{p_{\mathrm{ZF}}^{(1)}}\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \mathbf{A}_{\mathrm{ZF}}^{(2)} = \frac{1}{p_{\mathrm{ZF}}^{(2)}}\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \mathbf{A}_{\mathrm{ZF}}^{(3)} = \frac{1}{p_{\mathrm{ZF}}^{(3)}}\begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{3.12}$$

$\mathbf{A}_{\mathrm{ZF}}^{(1)}$, $\mathbf{A}_{\mathrm{ZF}}^{(2)}$, and $\mathbf{A}_{\mathrm{ZF}}^{(3)}$ should have 12 zero-value entries, which means all the interference signals except the desired one are canceled through ZF relay beamforming. However, if PZF beamforming is used at the relay, $\mathbf{A}_{\mathrm{PZF}}^{(1)}$, $\mathbf{A}_{\mathrm{PZF}}^{(2)}$, and $\mathbf{A}_{\mathrm{PZF}}^{(3)}$ will have the following forms

$$\mathbf{A}_{\mathrm{PZF}}^{(1)} = \begin{pmatrix} * & * & 0 & 0 \\ 0 & * & * & 0 \\ 0 & 0 & * & * \\ * & 0 & 0 & * \end{pmatrix}, \mathbf{A}_{\mathrm{PZF}}^{(2)} = \begin{pmatrix} * & * & * & 0 \\ 0 & * & * & * \\ * & 0 & * & * \\ * & * & 0 & * \end{pmatrix}, \mathbf{A}_{\mathrm{PZF}}^{(3)} = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix}, \tag{3.13}$$

where "$*$" is used to show that the entry can have any complex value. This way the restrictions on $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$, and $\mathbf{A}^{(3)}$ are reduced, as only 8 and 4 entries in $\mathbf{A}_{\mathrm{PZF}}^{(1)}$, and $\mathbf{A}_{\mathrm{PZF}}^{(2)}$ should be zero, respectively, and all others can take any complex value. In the first and second BC time slots, the relay beamforming matrices only need to be designed to cancel part of the interference, and the rest can be canceled through self-interference and successive interference can-

cellation at the users. In the third BC time slot, the relay leaves the whole interference to be canceled by the users, as they have the knowledge of their own information symbols and also already decoded information symbols from the first and second BC time slot.

In the following, the PZF beamforming design is formulated and the relay beamforming matrix optimization problem is specified. First, we define the structure of $\mathbf{A}^{(t)}$ for PZF. The (i,j)-th element of $\mathbf{A}^{(t)}$ is denoted as $a_{ij}^{(t)}$. To clearly express the PZF constraints on $\mathbf{A}^{(t)}$, a set of 3-tuples of indexes are introduced as the following,

$$\mathbb{A} = \left\{ (i,j,t) \;\middle|\; \begin{array}{l} t = 1,2,\cdots,K-2; \\ i = 1,2,\cdots,K; \\ q = 1,2,\cdots,K-t-1 \\ j = \mathrm{mod}_K(i+q+t-1)+1, \end{array} \right\}, \tag{3.14}$$

which is a subset of the 3-tuples of the indexes (i,j,t) representing the receiving user, the transmitting/interfering user, and the BC time slot. A tuple is an element of $\mathbb{A}$, if in the t-th BC time slot, the interference of u_j to u_i needs to be canceled with the PZF design.

From (3.7)-(3.10), the sum-rate maximization problem can be stated mathematically, as

$$\max_{\mathbf{G}^{(1)},\cdots,\mathbf{G}^{(K-1)}} \sum_{i=1}^{K} \min_{k\neq i} \left\{ \log_2 \left(1 + \frac{p_u \left| \mathbf{h}_k^T \mathbf{G}^{(t)} \mathbf{h}_i \right|^2}{\left| \mathbf{h}_k^T \mathbf{G}^{(t)} \right|^2 + 1} \right) \right\} \tag{3.15}$$

$$\text{s.t.} \quad \mathrm{tr}\left\{ \mathbf{G}^{(t)} \left(p_u \mathbf{H}\mathbf{H}^H + \mathbf{I} \right) \left(\mathbf{G}^{(t)} \right)^H \right\} \leq P_{\mathrm{R}}, \tag{3.16}$$

$$\text{and} \quad [\mathbf{H}^T \mathbf{G}^{(t)} \mathbf{H}]_{(i,j)} = 0, \text{ for } (i,j,t) \in \mathbb{A}. \tag{3.17}$$

The non-linear constraint in (3.16) is due to the transmit power constraint at the relay and the linear constraints in (3.17) are forced by the PZF idea. This sum-rate maximization problem is a multi-dimensional non-linear optimization problem with linear and non-linear constraints. So, first we simplify the problem using the transformation in (3.11). While the optimization variables in (3.15) are beamforming matrices $\mathbf{G}^{(1)},\cdots,\mathbf{G}^{(K-1)}$, after applying the transformation in (3.11), the problem is converted to an optimization over

$\mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(K-1)}$, and $\mathbf{G}^{(t)}$ can be calculated from $\mathbf{A}^{(t)}$ using

$$\mathbf{G}^{(t)} = (\mathbf{H}^T)^+ \mathbf{A}^{(t)} \mathbf{H}^+. \tag{3.18}$$

This transformation makes the linear constraints in (3.17) simpler which in turn simplifies the optimization problem. Thus, the sum-rate maximization problem is transformed as

$$\max_{\mathbf{A}^{(1)},\cdots,\mathbf{A}^{(K-1)}} \sum_{i=1}^{N} \min_{k \neq i} \left\{ \log_2 \left(1 + \frac{p_u |\mathbf{h}_k^T \mathbf{G}^{(t)} \mathbf{h}_i|^2}{|\mathbf{h}_k^T \mathbf{G}^{(t)}|^2 + 1} \right) \right\} \tag{3.19}$$

$$\text{s.t.} \quad \text{tr}\left\{ \mathbf{G}^{(t)} \left(p_u \mathbf{H}\mathbf{H}^H + \mathbf{I} \right) (\mathbf{G}^{(t)})^H \right\} \leq P_{\text{R}}, \tag{3.20}$$

$$\text{and} \quad a_{ij}^{(t)} = 0, \text{for } (i,j,t) \in \mathbb{A}. \tag{3.21}$$

3.4 Optimization of PZF Matrix

In this section, first, a numerical method is provided to jointly optimize all $\mathbf{A}^{(t)}$ matrices, and then, a separate optimization method is provided.

Define

$$\mathbf{c}^{(t)} = [a_{11}^{(t)} \;\; a_{12}^{(t)} \;\; \cdots \;\; a_{ij}^{(t)} ((i,j,t) \notin \mathbb{A}) \;\; \cdots \;\; a_{KK}^{(t)}], \tag{3.22}$$

which includes all the nonzero entries in $\mathbf{A}^{(t)}$, and is U_t-dimensional where

$$U_t = (t+1)K, \text{for} \;\; t = 1, 2, \cdots, K-1. \tag{3.23}$$

For example, for $N = K = 4$, according to (3.13),

$$\mathbf{c}^{(1)} = [a_{11}^{(1)}, a_{12}^{(1)}, a_{22}^{(1)}, a_{23}^{(1)}, a_{33}^{(1)}, a_{34}^{(1)}, a_{41}^{(1)}, a_{44}^{(1)}], \tag{3.24}$$

$$\mathbf{c}^{(2)} = [a_{11}^{(2)}, a_{12}^{(2)}, a_{13}^{(2)}, a_{22}^{(2)}, a_{23}^{(2)}, a_{24}^{(2)}, a_{31}^{(2)}, a_{33}^{(2)}, a_{34}^{(2)}, a_{41}^{(2)}, a_{42}^{(2)}, a_{44}^{(2)}], \tag{3.25}$$

$$\mathbf{c}^{(3)} = [a_{11}^{(1)}, a_{12}^{(3)}, a_{13}^{(3)}, a_{14}^{(3)}, a_{21}^{(3)}, a_{22}^{(3)}, a_{23}^{(3)}, a_{24}^{(3)}, a_{31}^{(3)}, a_{32}^{(3)}, a_{33}^{(3)}, a_{34}^{(3)}, a_{41}^{(3)}, a_{42}^{(3)}, a_{43}^{(3)}, a_{44}^{(3)}]. \tag{3.26}$$

Further, define vector $\mathbf{c}$ formed by concatenating all the vectors $\mathbf{c}^{(t)}$, as

$$\mathbf{c} = [\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \cdots, \mathbf{c}^{(t)}, \cdots, \mathbf{c}^{(K-1)}]. \tag{3.27}$$

The vector contains $a_{ij}^{(t)}$s for $(i,j,t) \notin \mathbb{A}$ and is W-dimensional, where

$$W = (K+2)K(K-1)/2. \tag{3.28}$$

With these notations, the optimization problem in (3.19) to (3.21) can be transformed to an optimization problem over $\mathbf{c}$ and the constraints in (3.21) are consequently removed. Since the objective function in (3.19) is non-convex and the constraints in (3.20) are non-linear, the solution is generally hard to find. A common method to find sub-optimal solutions for such non-convex problems is to use the gradient-ascent method. However, due to the complicated non-linear constraints the conventional gradient-ascent method does not apply in this case. Actually, by moving toward the gradient direction even with a small step size, the new $\mathbf{c}$-vector may violate the power constraint. To avoid this, we propose a modification to the gradient-ascent method. Our modified gradient-ascent method updates the $\mathbf{c}$-vector towards the direction of the modified gradient, as shown in the following.

Denote the objective function in (3.19) as $f(\mathbf{c})$, and the power constraint in (3.20) as $\phi(\mathbf{c}^{(t)}) \leq P_{\mathrm{R}}$, where

$$\phi(\mathbf{c}^{(t)}) = \phi(\mathbf{A}^{(t)}) = \mathrm{tr}\left\{\mathbf{G}^{(t)}(p_u\mathbf{H}\mathbf{H}^H + \mathbf{I})(\mathbf{G}^{(t)})^H\right\}. \tag{3.29}$$

So, the optimization problem becomes

$$\max_{\mathbf{c}} f(\mathbf{c}) \tag{3.30}$$

$$\text{s.t. } \phi(\mathbf{c}^{(t)}) \leq P_{\mathrm{R}} \text{ for } t = 1, 2, \cdots, K-1. \tag{3.31}$$

Notice from the definitions in (3.22) and (3.27) that the m-th element in $\mathbf{c}$ is the l-th element of $\mathbf{c}^{(t)}$ with the relationship, $m = 2K + \cdots + tK + l$. Letting $\mathbf{e}_l$ be the l-th canonical basis vector, we define the following power normalization factors

$$\alpha_m^{\mathrm{Re}} = \frac{\phi(\mathbf{c}^{(t)} + \epsilon\mathbf{e}_l)}{P_{\mathrm{R}}} \text{ and } \alpha_m^{\mathrm{Im}} = \frac{\phi(\mathbf{c}^{(t)} + i\epsilon\mathbf{e}_l)}{P_{\mathrm{R}}}. \tag{3.32}$$

The modified partial derivative of f with respect to the m-th element of $\mathbf{c}$ is

given by

$$\begin{aligned} d(f, c_m) =& \lim_{\epsilon \to 0} \frac{f\left(\mathbf{c}^{(1)}, \cdots, \frac{\mathbf{c}^{(t)} + \epsilon \mathbf{e}_l}{\alpha_m^{\mathrm{Re}}}, \cdots, \mathbf{c}^{(K-1)}\right) - f(\mathbf{c})}{\epsilon} \\ &+ i \lim_{\epsilon \to 0} \frac{f\left(\mathbf{c}^{(1)}, \cdots, \frac{\mathbf{c}^{(t)} + i\epsilon \mathbf{e}_l}{\alpha_m^{\mathrm{Im}}}, \cdots, \mathbf{c}^{(K-1)}\right) - f(\mathbf{c})}{\epsilon}. \end{aligned} \tag{3.33}$$

Compared with the definition of normal partial derivative,

$$\begin{aligned} \frac{\partial f}{\partial c_m} =& \lim_{\epsilon \to 0} \frac{f\left(\mathbf{c}^{(1)}, \cdots, \mathbf{c}^{(t)} + \epsilon \mathbf{e}_l, \cdots, \mathbf{c}^{(K-1)}\right) - f(\mathbf{c})}{\epsilon} \\ &+ i \lim_{\epsilon \to 0} \frac{f\left(\mathbf{c}^{(1)}, \cdots, \mathbf{c}^{(t)} + i\epsilon \mathbf{e}_l, \cdots, \mathbf{c}^{(K-1)}\right) - f(\mathbf{c})}{\epsilon}, \end{aligned} \tag{3.34}$$

(3.33) takes the non-linear constraint $\phi(\mathbf{c}^{(t)}) \leq P_{\mathrm{R}}$ into account. In other words, to make sure that this constraint is not violated when $\mathbf{c}^{(t)}$ is modified to $\mathbf{c}^{(t)} + \epsilon \mathbf{e}_l$ or $\mathbf{c}^{(t)} + i\epsilon \mathbf{e}_l$, the vector is scaled by α_m^{Re} or α_m^{Im} whose definition guarantees the power constraint. The modified gradient of f is thus,

$$D(f, \mathbf{c}) = [d(f, c_1) \quad \cdots \quad d(f, c_m) \quad \cdots \quad d(f, c_W)]. \tag{3.35}$$

Algorithm 1 Joint optimization scheme.

1: Initialize $\alpha, tolerance, \mathbf{A}^{(t)}$s, $\mathbf{c}$ and calculate $D(f, \mathbf{c})$.
2: **while** $norm(D(f, \mathbf{c})) \geq tolerance$ **do**
3: Update $\mathbf{c}$: $\mathbf{c} = \mathbf{c} + \alpha D(f, \mathbf{c})$.
4: Construct $\mathbf{A}^{(t)}$s from $\mathbf{c}$.
5: Scale $\mathbf{A}^{(t)}$s based on the constraint and construct $\mathbf{c}$.
6: Calculate $D(f, \mathbf{c})$.
7: **end while**
8: Calculate $\mathbf{G}^{(1)}, \cdots, \mathbf{G}^{(K-1)}$ using (3.18).

In the proposed numerical method, the $\mathbf{c}$-vector is updated toward the modified gradient with a step size α. Also, a scaling step is done at every iteration to guarantee that each searched point satisfies the constraint. In fact, a new point is found by two moves. First, a move of $\mathbf{c}$ proportional to the modified gradient is made. Second, constructed from $\mathbf{c}$, $\mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(K-1)}$ are

scaled to make the power constraint satisfied. $\mathbf{c}$ is then moved to a new point accordingly. Once a solution for $\mathbf{c}$ is found, we reconstruct $\mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(K-1)}$, and then from (3.18) calculate $\mathbf{G}^{(1)}, \cdots, \mathbf{G}^{(K-1)}$. It should be noted that similar to the gradient-ascent method, the proposed modified gradient-ascent method does not guarantee the global optimum solution. However, we can use ZF relay beamforming matrices as the initial point to guarantee a solution better than ZF. The algorithm is described in Algorithm 1.

In the optimization method discussed above, the matrices $\mathbf{A}^{(1)}, \cdots, \mathbf{A}^{(K-1)}$ are jointly optimized and thus the algorithm can be computationally expensive for large MWRNs. In the following, we propose a separate optimization method where the optimization over $\mathbf{A}^{(t)}$s for $t = 1, \cdots, K-1$ is conducted separately and sequentially.

According to (3.7) and (3.8), the relay beamforming matrix for the t-th BC time slot, $\mathbf{G}^{(t)}$, directly affects the transmission rates $R_{k,i(k)}$ during this BC time slot and if we assume ideal source coding and detection it does not affect the transmission rates of previous or later BC time slots. Therefore, we propose to optimize $\mathbf{G}^{(t)}$, or equivalently $\mathbf{c}^{(t)}$ by maximizing the sum-rate in the t-th BC time slot, given by

$$R_{\text{sum}}^{(t)} = \sum_{i=1}^{K} \log_2 \left(1 + \frac{p_u |\mathbf{h}_k^T \mathbf{G}^{(t)} \mathbf{h}_i|^2}{|\mathbf{h}_k^T \mathbf{G}^{(t)}|^2 + 1} \right). \tag{3.36}$$

Hence, the optimization problem is

$$\max_{\mathbf{c}^{(t)}} R_{\text{sum}}^{(t)} \tag{3.37}$$

$$\text{s.t. } \phi(\mathbf{c}^{(t)}) \leq P_{\text{R}} \text{ for } t = 1, 2, \cdots, K-1. \tag{3.38}$$

Again, modified gradient-ascent method is used to solve the above optimization problem. The number of constraints on $\mathbf{A}^{(t)}$ decreases as t increases, so, we optimize $\mathbf{A}^{(t)}$s sequentially with $\mathbf{A}^{(1)}$ being the first and $\mathbf{A}^{(K-1)}$ being the last. This separate optimization algorithm is clarified in Algorithm 2.

Algorithm 2 Separate optimization scheme.

Initialize α and *tolerance*.
for $t = 1 : K - 1$ **do**
    Initialize $\mathbf{A}^{(t)}$, $\mathbf{c}^{(t)}$ and calculate $D(R_{\text{sum}}^{(t)}, \mathbf{c}^{(t)})$.
    while $norm(D(R_{\text{sum}}^{(t)}, \mathbf{c}^{(t)})) \geq tolerance$ **do**
        Update $\mathbf{c}^{(t)}$: $\mathbf{c}^{(t)} = \mathbf{c}^{(t)} + \alpha D(R_{\text{sum}}^{(t)}, \mathbf{c}^{(t)})$.
        Construct $\mathbf{A}^{(t)}$ from $\mathbf{c}^{(t)}$.
        Scale $\mathbf{A}^{(t)}$ and construct $\mathbf{c}^{(t)}$.
        Calculate $D(R_{\text{sum}}^{(t)}, \mathbf{c}^{(t)})$.
    end while
end for
Calculate $\mathbf{G}^{(1)}, \cdots, \mathbf{G}^{(K-1)}$ using (3.18).

In the following, we discuss the convergence behavior of our proposed algorithms and analyze their computational complexity in comparison with the existing schemes.

The number of iterations needed for the proposed optimizations depends on the step size α, and there is a natural trade-off between the convergence rate and the achieved sum-rate. Here, we simply choose $\alpha = 0.03$ for the separate optimization and $\alpha = 0.01$ for the joint optimization based on experience. By stopping the iterations when less than 5% improvement is observed over one iteration, the separate optimization algorithm converges after around 75 iterations and the joint one converges after about 100 iterations.

Next, we analyze the computational complexities of the proposed joint and separate optimization algorithms for our PZF beamforming design, and compare them with those of ZF, MMSE, RZF, and MF beamforming schemes [35, 68], as well as the scheme proposed in [60]. The order of complexity with respect to the number of relay antennas N, number of users K, and the iteration number *iter* are used for the analysis.

The mathematical operations in the beamforming matrix optimization include summation, multiplication, division, square root, sorting, taking logarithm, and comparison. Among these operations, division and multiplication have the highest computational complexity and the highest numbers of happening, while other operations lead to much lower computational complexity. So, our analysis focuses on division and multiplication. Considering different beamforming schemes, the required numbers for each of the two operations are

listed in Table 3.1.

Scheme\Operations	$\times$	$\div$
ZF	$K^3(2N)$	$2K^2 + \frac{N^2}{2}$
MMSE and RZF	$K^3(2N)$	N^2
MF	$K^3(2N)$	$\frac{N^2}{2}$
PZF-Joint	$iter \times K^5(6N^2)$	$iter \times 2K^5$
PZF-Separate	$iter \times K^4(6N^2)$	$iter \times 2K^4$

Table 3.1: The numbers of multiplications and divisions in the design of PZF (joint and separate), ZF, MMSE, RZF, and MF schemes.

As can be seen from Table 3.1, the numbers of multiplications for ZF, MMSE, RZF and MF schemes are the same. Actually, this value comes from the calculation of $\mathbf{G}^{(t)} = \mathbf{G}_{\mathrm{TX}}^{(t)}\mathbf{P}^t\mathbf{G}_{\mathrm{RX}}$. For both proposed PZF beamforming methods, the dominant parts for the multiplications are resulted from the calculations of $|\mathbf{h}_k^T\mathbf{G}^{(t)}\mathbf{h}_i|^2$ and $|\mathbf{h}_k^T\mathbf{G}^{(t)}|^2$, while for the divisions they are resulted from the calculations of $p_u|\mathbf{h}_k^T\mathbf{G}^{(t)}\mathbf{h}_i|^2/(|\mathbf{h}_k^T\mathbf{G}^{(t)}|^2 + 1)$.

Table 3.1 also shows that both proposed beamforming optimizations bear higher computational complexity than other common schemes. This higher complexity is due to the iterative feature of our algorithms, and the fact that we optimize more elements in the transformations of the beamforming matrices, i.e., our optimization problems have higher dimensions. For the joint algorithm, the whole beamforming matrices for different time slots are optimized together which leads to higher complexity than the separate algorithm. Despite the high complexity of PZF beamforming algorithms, parallel computing can significantly reduce the computation time. As discussed earlier in this subsection, the iteration number is about 100 for the joint optimization and 75 for the separate optimization. Another insight from this table is that for large number of antennas, i.e., when N is large, the complexity of our PZF beamforming is still tractable, as it is only one order of magnitude higher than the other schemes. On the other hand, for large numbers of antennas and users, i.e., when both N and K are large, the complexity increases 3 orders of magnitude faster than the other schemes which may make it intractable.

Recently, another beamforming strategy is proposed in [60], where the relay beamforming matrices and user decoding are jointly designed and multi-symbol processing is used at each user. While our work targets sum-rate optimization,

[60] studied the SINR max-min optimization. The computational complexity of the scheme in [60] is $O(iter \times K^4 N^6)$, which has a higher order than both of our schemes. Further, it has a higher decoding complexity of $O(K^2)$ at the users due to the multi-symbol processing. In Table I of [60], the authors have shown a comparison between the average CPU processing time for their approach and our approach when $K = 3, N = 3$, and $K = 4, N = 4$ which declares that ours is about 5 times faster.

The proposed algorithms for partial zero-forcing are based on gradient-ascend method and there is no global convergence guarantee. Since ZF relay matrices are selected as the initialization point, the solutions found by our algorithms are guaranteed to achieve higher achievable sum-rates than ZF. To see this, we show that ZF relay matrices are not local optima. We first consider the joint optimization and look at the problem given in (3.19)-(3.21). To show that ZF relay matrices are not local optima of the optimization problem, it is sufficient to prove that $D(f, \mathbf{c}_{\mathrm{ZF}}) = 0$ does not hold where $\mathbf{c}_{\mathrm{ZF}}$ is the corresponding vector of the ZF relay matrices, ${\mathbf{G}_{\mathrm{ZF}}}^{(t)}$ for $t = 1, ..., K-1$. Notice that $D(f, \mathbf{c}_{\mathrm{ZF}})$ is a function of the channel matrix $\mathbf{H}$. From (2.3) and (3.11), it can easily be shown that $\mathbf{A}_{\mathrm{ZF}}^{(t)} = \frac{1}{p_{\mathrm{ZF}}^{(t)}} \mathbf{P}^t$, thus according to (3.14), (3.22), and (3.27), $\mathbf{c}_{\mathrm{ZF}}$ has $(K-1)K$ non-zero coefficients, as there are K non-zero elements in each $\mathbf{P}^t$. For instance, for the case $N = K = 3$, $\mathbf{c}_{\mathrm{ZF}}^{(1)} = \frac{1}{p_{\mathrm{ZF}}^{(1)}}[0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 0]$ and $\mathbf{c}_{\mathrm{ZF}}^{(2)} = \frac{1}{p_{\mathrm{ZF}}^{(2)}}[0 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0]$. From (3.28), the number of equations in $D(f, \mathbf{c}_{\mathrm{ZF}}) = 0$ is $W = (K+2)K(K-1)/2$, which is larger than the number of non-zero parameters in $\mathbf{c}_{\mathrm{ZF}}$. As the $N \times K$ channel matrix $\mathbf{H}$ is random whose entries are i.i.d., with each entry following Rayleigh distribution, the probability that $D(f, \mathbf{c}_{\mathrm{ZF}}) = 0$ holds is zero even when the non-zero coefficients are adjustable. This means that with probability 1, ZF relay matrices are not local optima of the optimization problem. Via similar reasoning, it can be shown that the ZF beamforming matrices cannot be the local optima for the separate optimization method as well. Simulations have also supported this conclusion, and we always obtain higher sum-rates than the ZF beamforming.

3.4.1 Simulation Results

In this section, we compare the simulation results of the sum-rates of MWRNs considering our PZF design and other existing designs [35, 60]. We set $N =$

$K = 3$, i.e., 3 single-antenna users communicate with each other with the help of a relay equipped with 3 antennas.

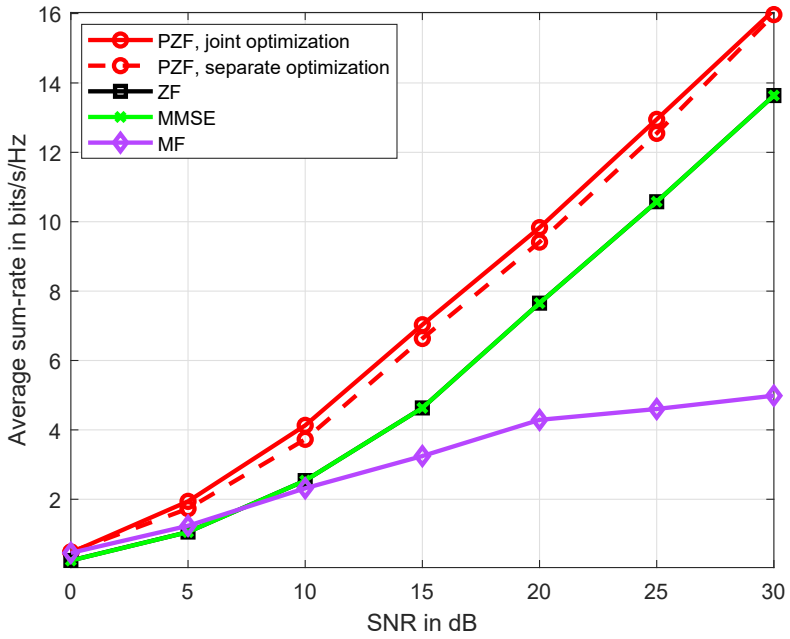


Figure 3.1: Sum-rates for a homogeneous 3-user MWRN with $P_{\mathrm{R}} = 1$.

We assume a homogeneous network where all the channels follow i.i.d. $\mathcal{CN}(0, \beta_h)$, and set $P_{\mathrm{R}} = p_u = 1$. Thus, the SNR of each user will be β_h. Figure 3.1, which is regenerated based on the data available from [91], shows the sum-rates versus different SNR values. It can be seen that the PZF design has the highest sum-rate performance for the all SNR values. Further, this figure shows that the separate and joint optimization methods for the PZF scheme give very close sum-rate performance with the latter slightly better. Simulation results on the sum-rate comparison between our proposed scheme and the one in [60] is available in Figures 13 and 14 of [60] for the cases of $K = 3, N = 3$ and $K = 4, N = 4$. It can be observed that our proposed joint design has slightly lower sum-rate performance and the gap shrinks as the SNR increases.

Moreover, sum-rates of MWRNs with different numbers of users are simulated. We consider that every MWRN has an 8-antenna relay, and the number of users changes from 3 to 8. We set $P_{\mathrm{R}} = p_u = 1$ and SNR = 20 dB at the

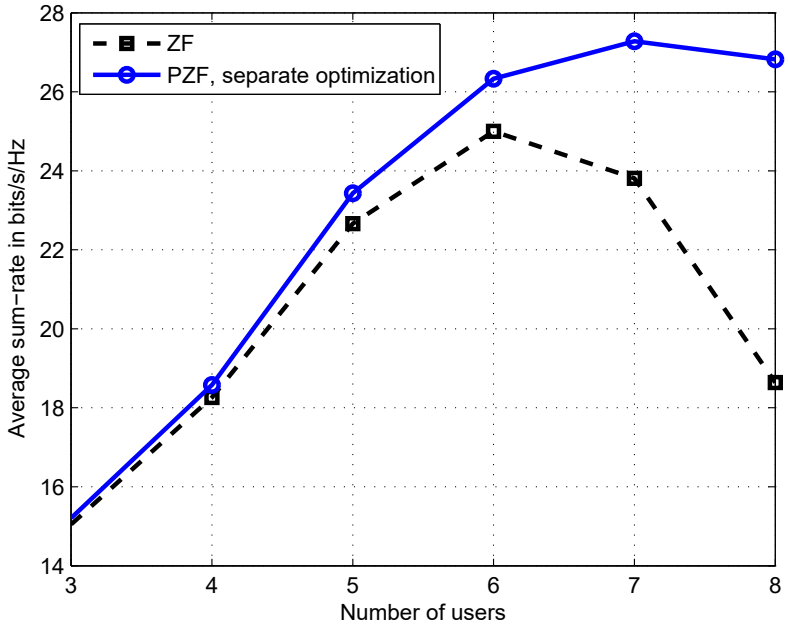


Figure 3.2: Sum-rates of PZF and ZF with unicasting strategy for different numbers of users, $N = 8$, SNR=20 dB.

relay. The channels follow i.i.d. $\mathcal{CN}(0, \beta_h)$. Figure 3.2 shows the relationship between the number of users and the sum-rate for the ZF and PZF schemes. From the figure, we can conclude that the network sum-rate first increases and then decreases as the number of users increases. Also, it can be seen that the advantage of PZF over the ZF design enlarges with more number of users. The reasons for this are two-fold. First, compared to ZF, PZF beamforming allows extra $K(t+1)$ degrees-of-freedom in the design of $\mathbf{G}^{(t)}$. So, as the number of users K increases, there are more extra degrees-of-freedom in the PZF design compared to ZF. Another contributing factor is the ZF beamforming coefficient, $\frac{1}{p_{\text{ZF}}^{(t)}}$, which tends to decrease when K increases. This leads to a lower SNR at the users and thus a lower achievable sum-rate.

Next, we consider a 3-user MWRN with non-identical fading channels due to different path-loss, which is named as a heterogeneous network in this work. Denote d_i as the distance from an arbitrary user, u_i, to the relay. The channels between u_i and the N relay antennas, $h_{n,i}$s, are assumed to follow $\mathcal{CN}(0, \beta_i)$, where $\beta_i = (\psi/d_i)^\nu$ with ψ being a constant. We set $d_3 = 2d_2 = 4d_1$ and assume $\nu = 2$. Due to the heterogeneous setup, the order of decoding may

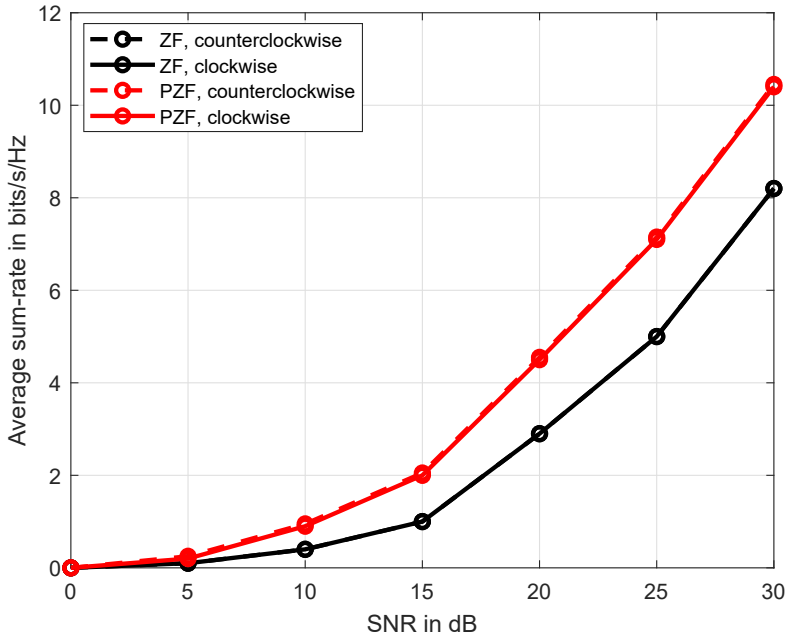


Figure 3.3: Sum-rates for a heterogeneous 3-user MWRN with $P_{\mathrm{R}} = 1$, through separate optimization.

affect the sum-rate, thus, we consider two orderings of detection clockwise as defined in (2.2) and counter clockwise, defined as $i(k,t) = \mathrm{mod}_K(k-t-1)+1$. In Figure 3.3, which is regenerated based on the data available from [91], the x-axis, denoted as SNR, shows u_1's SNR at the relay, thus, SNR $= \beta_1 = 4\beta_2 = 16\beta_3$. It can be seen that the proposed PZF design achieves significantly higher sum-rates than the ZF design. For both clockwise and counter clockwise orderings, ZF provides exactly the same sum-rate performances, while PZF provides slightly different performances. For systems with more users or relay antennas, the advantage of adopting a better decoding order may become larger as the difference between the channel qualities of different users will become larger on average. So, in this case the issue of finding the optimal decoding order becomes more important. Since our focus is the new PZF relay beamforming design not decoding order, we refer further investigations on how the decoding order affects the PZF scheme to future work.

Next, we compare the symbol error rate (SER) of PZF beamforming with the ones of ZF, MMSE, and MF schemes. The SER results for a homogeneous MWRN with $N = K = 3$ are shown in Figure 3.4.Quadrature amplitude

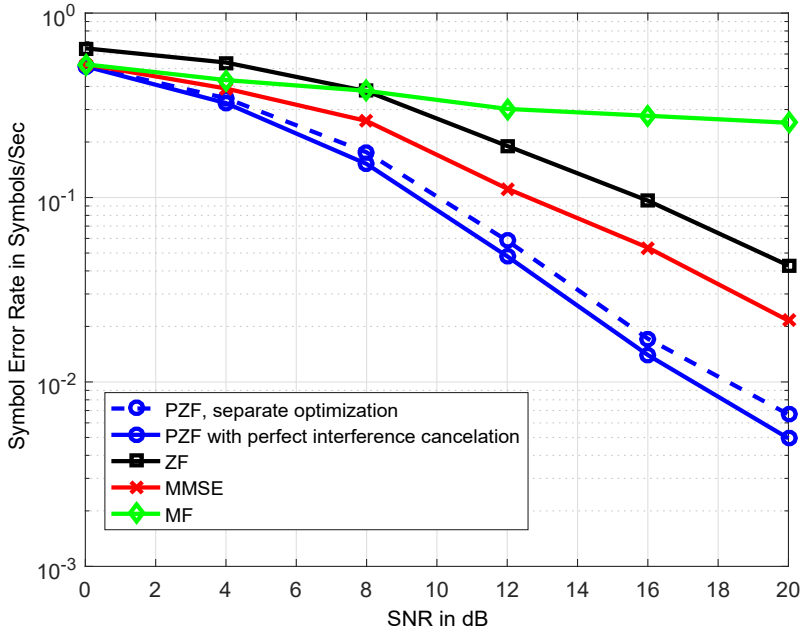


Figure 3.4: SERs for a homogeneous 3-user MWRN with $P_{\mathrm{R}} = 1$.

modulation (QAM) is used for all user symbols. It can be seen that our PZF design provides a far lower SER than ZF, MF, and MMSE schemes. Furthermore, to see the effect of error propagation, which is the detection error of a symbol caused by the symbol detection errors in previous time slots, we have presented the simulation results for the ideal case of perfect interference cancellation for the PZF scheme. In this ideal scheme, for every BC time slot, we cancel the interference caused by the previously decoded signals using the correct and error-free symbols, instead of using the decoding results from the previous time slots. This way, no decoding errors in previous BC time slots can propagate to the coming BC time slots. The simulations show that the effect of error propagation is negligible for our proposed PZF scheme.

Moreover, in order to see the behavior of error propagation when the number of users increases in homogeneous networks, we have presented the simulation results for SER versus the number of users, where $N = K$ changes from 3 to 8. As it is shown in Figure 3.5, the effect of error propagation slightly increases as the user number increases, but it is still very small in comparison to the performance enhancement that our proposed beamforming has brought compared to the ZF scheme.

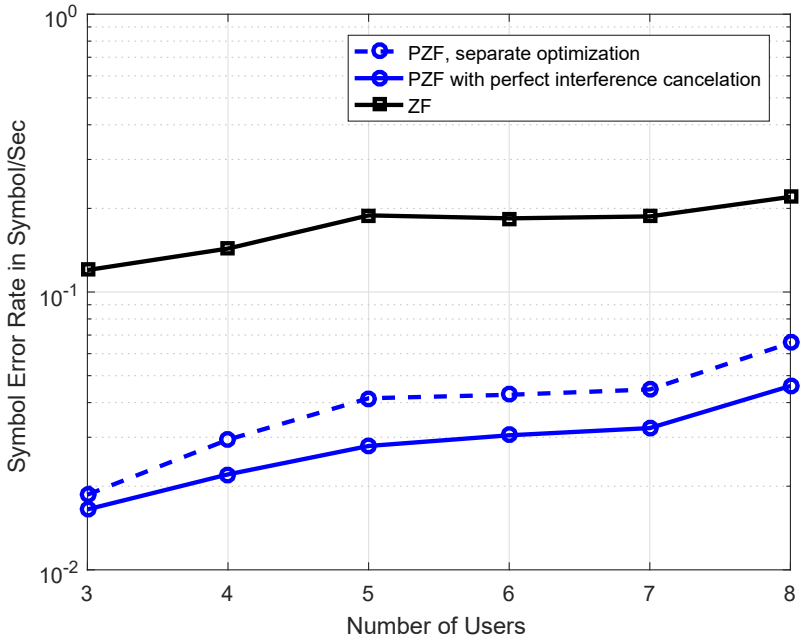


Figure 3.5: SER versus the number of users for homogeneous networks, where $N = K$ and SNR=15 dB.

In addition to homogeneous networks, in Figure 3.6, we have presented the results of SER versus SNR of heterogeneous networks for $K = 4$ and $K = 6$ cases when $N = 32$ antennas are available at the relay. Recall that the channels between an arbitrary user, u_i, and the relay antennas, $h_{n,i}$s, follow $\mathcal{CN}(0, \beta_i)$, where $\beta_i = (\psi/d_i)^\nu$ with ψ being a constant and d_i being the distance from u_i to the relay. In this simulation, we have set $d_i = 2^{(i-1)}d_1$ for $i = 1, 2, ..., K$ and $\nu = 2$. The SNR in Figure 3.6 represents the SNR of the first user, u_1. This figure shows the results for the clockwise order of decoding. It can be seen from Figure 3.6 that the effect of error propagation diminishes as SNR grows, also we have more error propagations when there are higher number of users. Moreover, it can be observed that PZF gives better SER performances than ZF for $K = 6$, while when $K = 4$ this may not be the case in higher SNR ranges. The reason for this is two-fold. 1) Our optimization targets at sum-rate maximization not SER optimization, which may lead to degraded SER performance. 2) The number of degrees-of-freedom increases as the number of users increase, so we can achieve better SER performances in comparison to ZF when higher number of users are involved.

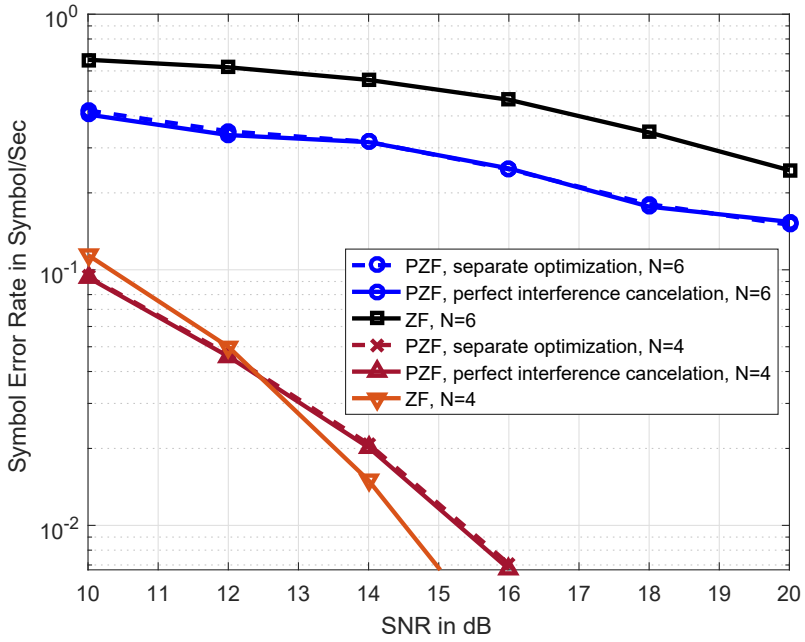


Figure 3.6: SER performance of heterogeneous networks for $K = 4$, and $K = 6$ cases when $N = 32$.

3.5 Extension to MWRNs with Hybrid Uni/Multicasting

In Section 3.3, we considered that in each BC time slot, the relay transmits uniquely different information symbols to different users, which is called transmission via unicasting. However, in this section, the hybrid uni/multicasting strategy is considered. It is shown that when the relay uses hybrid uni/multicasting strategy, PZF is still able to improve the sum-rate performance of MWRNs.

Both the unicasting and hybrid uni/multicasting strategy are proposed in [92]. In hybrid uni/multicasting strategy, in each BC time slot, one information symbol is exclusively transmitted to one user, the unicasted symbol, and another information symbol is transmitted to the other $K - 1$ users, the multicasted symbol. The unicasted information symbol is fixed for all BC time slots and transmitted to a different user in different BC time slots. While the multicasted information symbols are changed in different BC time slot. With this hybrid uni/multicasting scheme each user receives all other users' information symbols within the BC phase. Detection scheduling of this transmission strategy will be discussed by the end of the section. A 4-user example of hy-

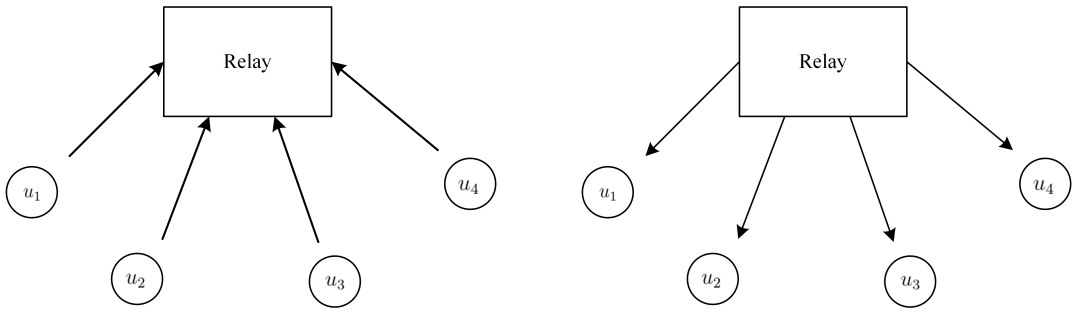
brid uni/multicasting strategy is shown in Figure 3.7. In the BC phase, u_1's signal is chosen as the unicasting symbol, while u_2, u_3, and u_4's signals are chosen as the multicasting symbols for the first, second, and third BC time slots, respectively. It takes the following steps for a full mutual communication.

1. In time slot 1 (MAC Phase), all users transmit their signals simultaneously to the relay.

2. In time slot 2 (first BC time slot), the relay transmits u_1's signal to u_2, and multicasts u_2's signal to u_1, u_3 and u_4.

3. In time slot 3 (second BC time slot), the relay transmits u_1's signal to u_3, and multicasts u_3's signal to u_1, u_2 and u_4.

4. In time slot 4 (third BC time slot), the relay transmits u_1's signal to u_4, and multicasts u_4's signal to u_1, u_2 and u_3.



(a) The MAC phase. (b) A BC time slot.

Figure 3.7: Transceiver protocol of a 4-user MWRN.

After the MAC and the BC phases, each user decodes the information symbols from all other users.

PZF can consistently be extended to the hybrid uni/multicasting transmission strategy. The only necessary modification in the problem formulation of sum-rate maximization is the structures of $\mathbf{A}^{(t)}$ matrices which should be adjusted based on the hybrid uni/multicasting strategy. For example, for the

aforementioned 4-user network, $\mathbf{A}^{(t)}$s should have the following forms

$$\mathbf{A}_{\mathrm{PZF}}^{(1)} = \begin{pmatrix} * & * & 0 & 0 \\ * & * & 0 & 0 \\ 0 & * & * & 0 \\ 0 & * & 0 & * \end{pmatrix}, \mathbf{A}_{\mathrm{PZF}}^{(2)} = \begin{pmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ 0 & * & * & * \end{pmatrix}, \mathbf{A}_{\mathrm{PZF}}^{(3)} = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix}, \tag{3.39}$$

So, this problem can be solved by the modified gradient-ascent method proposed in Section 3.3.

In the following, the simulation results on the sum-rate of MWRNs when hybrid uni/multicasting is the transmission strategy at the relay are presented and compared with the results for unicasting strategy. Results are shown for $N = K = 3$ considering our proposed PZF design and the ZF beamforming design. We consider a homogeneous network where all channels follow i.i.d. $\mathcal{CN}(0, \beta_h)$, and set $P_{\mathrm{R}} = p_u = 1$.

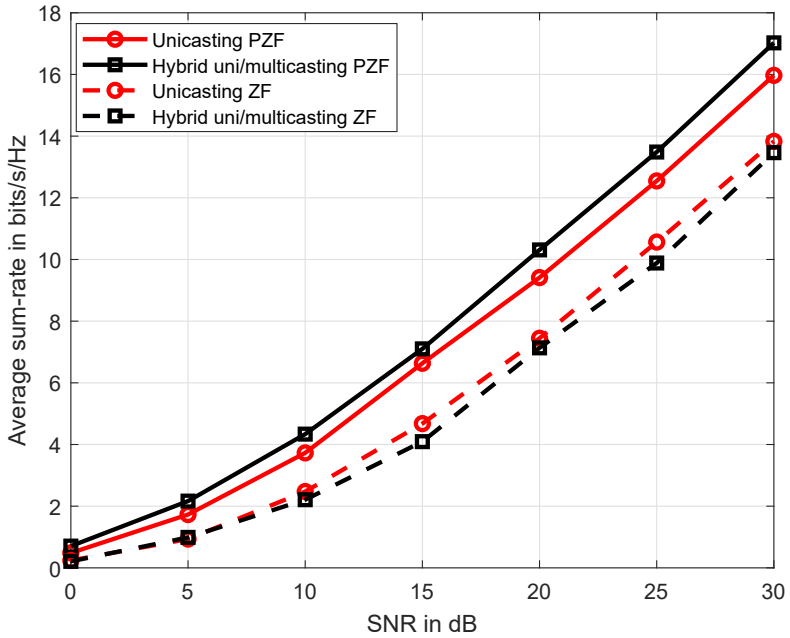


Figure 3.8: Sum-rates of homogeneous 3-user MWRNs with unicasting and hybrid uni/multicasting transmission strategies. For the PZF scheme separate optimization is done.

Figure 3.8, which is regenerated based on the data available from [91], presents the sum-rates versus different SNR values. It can be seen that the

PZF design with the hybrid uni/multicasting strategy has a better sum-rate performance than its ZF counterpart for all SNR values. Further, it can be seen that the sum-rate performance gap between PZF and ZF designs increases when hybrid uni/multicasting is used.

Opposite to the unicast strategy, hybrid uni/multicast strategy brings asymmetry in the transmission of different users' signals. For MWRNs with asymmetric channel conditions, the choices of unicasting and multicasting signals in different BC time slots, or the orderings of detection in the BC phase, may affect the sum-rate performance. Commonly, the channel condition is used to decide which users' signals should be unicasted and the order of other users' signals to be multicasted in the BC time slots. It is beneficial to multicast the signals of users with good channel conditions during earlier BC time slots and the ones with poor channel conditions during later BC time slots. The choice of users' signals to be unicasted is complicated and needs further study. For instance, in a 3-user network, we may choose u_1's signal as the unicast signal, u_2's signal as the mulicast signal for the first BC time slot, and u_3' signal as the multicast signal for the second one. Consequently, u_1 first decodes u_2's signal and then u_3's. We name this detection scheduling hybrid uni/multicasting-1. A different ordering of detection for the same 3-user network happens when u_3's signal is multicasted in the first BC time slot, and u_2's signal is multicasted in the second one. Then, u_1 first decodes u_3's signal and then u_2's. We name this detection scheduling hybrid uni/multicasting-2.

In Figure 3.9, which is regenerated based on the data available from [91], simulation results on different orderings of detection are presented. A 3-user MWRN is considered with non-identical fading channels due to different path-loss. The same as Subsection 3.4.1, we denote d_i as the distance from an arbitrary user, u_i, to the relay. The channels between u_i and the N relay antennas, $h_{n,i}$s, are assumed to follow $\mathcal{CN}(0, \beta_i)$, where $\beta_i = (\psi/d_i)^\nu$ with ψ being a constant. We set $d_3 = 2d_2 = 2d_1$, $\nu = 2$. With this heterogeneous setup, two orderings of detection are considered: the hybrid uni/multicasting-1, and the hybrid uni/multicasting-2. In Figure 3.9, SNR axis shows u_1's SNR at the relay, thus SNR $= \beta_1 = \beta_2 = 4\beta_3$. From this figure, it can be seen that hybrid uni/multicasting-1 results in higher sum-rates than hybrid uni/multicasting-2 when PZF is applied. A reason for this observation is that the channels

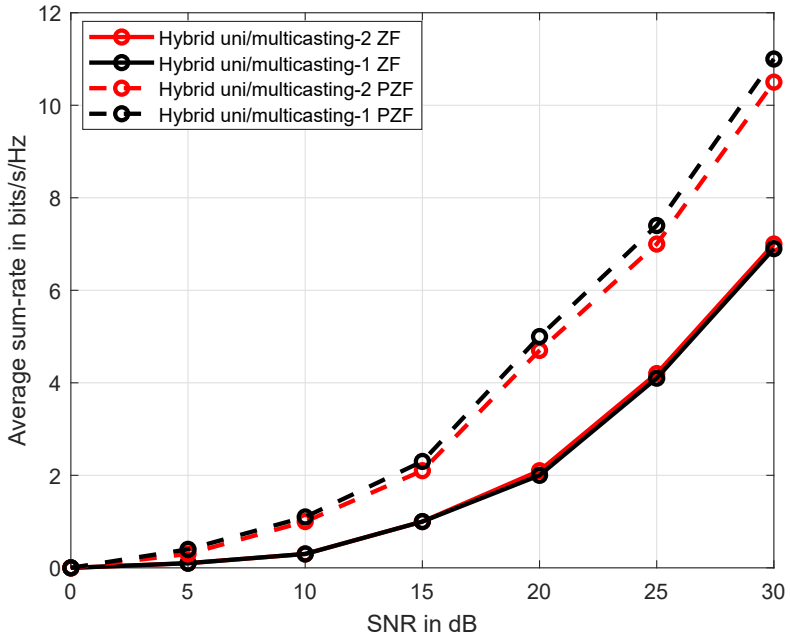


Figure 3.9: Sum-rates of a 3-user MWRN considering hybrid uni/multicasting strategy and different orderings of detection. For PZF scheme separate optimization is performed.

between u_3 and the relay are weaker than the channels between the other two users and the relay. Therefore, as hybrid uni/multicasting-1 chooses to decode the weakest signal from u_3 in the last BC time slot, it leads to higher sum-rates.

3.6 Extension to MWRNs with $N = K - 1$

To provide enough degrees-of-freedom to remove users' interference in ZF beamforming, the number of relay antennas must be larger than or at least equal to the number of users, $N \geq K$, [93]. However, in PZF, due to the fact that the interference do not need to be fully canceled, the number of antennas at the relay can be reduced by one, $N \geq K-1$. The case of $N \geq K$ is considered in Sections 3.3 to 3.5. In this section, we consider MWRNs where the number of relay antennas is one less than the number of users, i.e, $N = K - 1$. The transceiver protocol is the same as Sections 3.3 to 3.5. So, there are 1 MAC time slot and $K - 1$ BC time slots. Also, as PZF design suggests, in each BC time slot for each user only partial interference, interference other than self-interference and interference from previously decoded signals, needs to be

canceled. However, the problem formulation of PZF design for the $N = K - 1$ case, is different from the one in Section 3.3. In fact, as the number of relay antennas is less than the number of users, the dimension of $\mathbf{G}^{(t)}$ is smaller than the dimension of $\mathbf{A}^{(t)}$, and thus, the map from $\mathbf{A}^{(t)}$ to $\mathbf{G}^{(t)}$, in (3.18), does not work. As a result, the sum-rate optimization needs to be done directly with respect to $\mathbf{G}^{(t)}$. This way the optimization problem formulation will be

$$\max_{\mathbf{G}^{(1)},\cdots,\mathbf{G}^{(K-1)}} \sum_{i=1}^{K} \min_{k\neq i} \left\{ \log_2 \left(1+\frac{p_u |\mathbf{h}_k^T \mathbf{G}^{(t)} \mathbf{h}_i|^2}{|\mathbf{h}_k^T \mathbf{G}^{(t)}|^2 + 1} \right) \right\} \tag{3.40}$$

$$\text{s.t.} \quad \text{tr}\left\{ \mathbf{G}^{(t)} \left(p_u \mathbf{H}\mathbf{H}^H + \mathbf{I} \right) (\mathbf{G}^{(t)})^H \right\} \leq P_{\text{R}}, \tag{3.41}$$

$$\text{and} \quad \mathbf{H}^T \mathbf{G}^{(t)} \mathbf{H} = \mathbf{A}^{(t)}, \text{ for } t = 1, 2, \cdots, K-1. \tag{3.42}$$

According to (3.23), there exist $(K-t-1)K$ zero entries in $\mathbf{A}^{(t)}$ and the rest of the entries can take any complex value. To simplify the optimization problem, (3.42) can be written as $(K-t-1)K$ linear homogeneous equations. So, we define $\mathbf{g}^{(t)}$ that contains all entries in $\mathbf{G}^{(t)}$, as

$$\mathbf{g}^{(t)} = [g_{11}^{(t)} \quad g_{12}^{(t)} \quad \cdots \quad g_{1,K-1}^{(t)} \quad g_{21}^{(t)} \quad \cdots \quad g_{K-1,K-1}^{(t)}]. \tag{3.43}$$

$\mathbf{g}^{(t)}$ can be divided into two vectors, $\mathbf{y}^{(t)}$ and $\mathbf{r}^{(t)}$, where $\mathbf{y}^{(t)}$ contains the first $(K-1)^2 - (K-t-1)K$ entries, and $\mathbf{r}^{(t)}$ contains the rest of $(K-t-1)K$ entries. Since the number of entries in $\mathbf{r}^{(t)}$ is equal to the number of linear equations in (3.42), $\mathbf{r}^{(t)}$ can be uniquely represented by $\mathbf{y}^{(t)}$ from (3.42). This way the constraints in (3.42) will be eliminated. Based on this discussion, the sum-rate maximization problem can be transformed into an optimization over $\mathbf{y}^{(t)}$, with the only constraint in (3.41). Thus, the proposed modified gradient-ascent method can be applied. The algorithm is given in Algorithm 3, where the separate optimization of relay beamforming matrices is considered for the matter of complexity .

In the following, we present the simulation results for the sum-rates of MWRNs, where $N = K - 1$ and our PZF design with unicasting is applied. $P_{\text{R}} = p_u = 1$ is set and homogeneous channels following i.i.d. $\mathcal{CN}(0, \beta_h)$ are considered. Figure 3.10 compares the sum-rates of PZF design for $N = 2, K = 3$, $N = 3, K = 4$, $N = K = 3$, and $N = K = 4$ cases. It can be seen from this figure that 1) the average sum-rates for $N = K = 4$ are the highest,

Algorithm 3 Separate optimization scheme for MWRNs where $N = K - 1$.

Initialize α and *tolerance*.
for $t = 1 : K - 1$ **do**
    Initialize $\mathbf{y}^{(t)}$ and construct $\mathbf{G}^{(t)}$ by solving (3.42).
    Scale $\mathbf{G}^{(t)}$ to satisfy (3.41) and construct $\mathbf{y}^{(t)}$.
    Calculate $D(R_{\text{sum}}^{(t)}, \mathbf{y}^{(t)})$.
    while $norm(D(R_{\text{sum}}^{(t)}, \mathbf{y}^{(t)})) \geq tolerance$ **do**
        Update $\mathbf{y}^{(t)}$: $\mathbf{y}^{(t)} = \mathbf{y}^{(t)} + \alpha D(R_{\text{sum}}^{(t)}, \mathbf{y}^{(t)})$.
        Construct $\mathbf{G}^{(t)}$ from $\mathbf{y}^{(t)}$ by solving (3.42).
        Scale $\mathbf{G}^{(t)}$ to satisfy (3.41) and construct $\mathbf{y}^{(t)}$.
    end while
end for

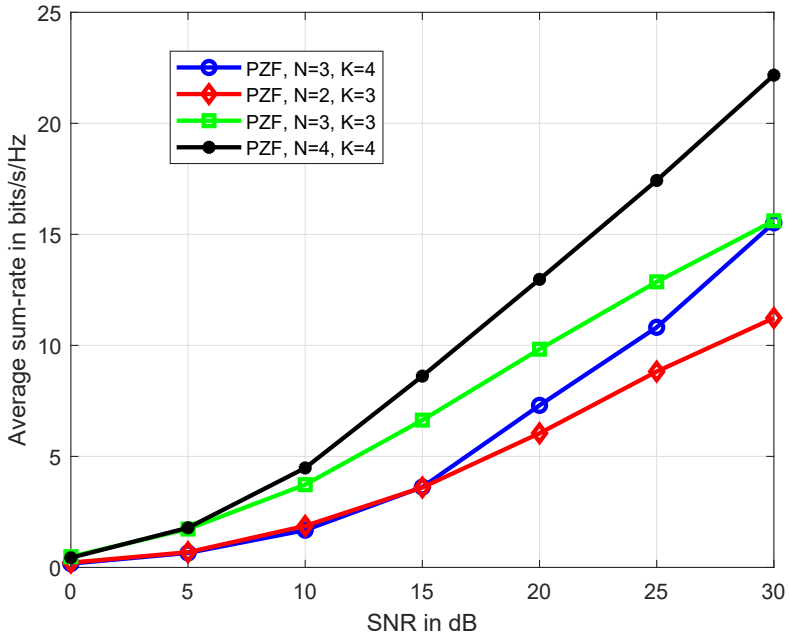


Figure 3.10: Sum-rates of MWRNs considering PZF design with unicasting strategy and separate optimization. The results for four network settings are compared.

2) the $N = K = 3$ case achieves higher average sum-rates than the case of $N = 2, K = 3$, and 3) compared to the case of $N = 3, K = 4$, the average sum-rates of the $N = K = 3$ case are higher for the SNR range of [0 dB, 30 dB], but the advantage decreases with the SNR and the curves indicate that the case of $N = 3, K = 4$ outperforms the case of $N = K = 3$ when the SNR is higher than 30 dB. Further, we can observe that in the high SNR regime, the

case of $N = 3, K = 4$ gives higher sum-rates than the case of $N = 2, K = 3$. However, in low SNR regime, the sum-rates are similar.

3.7 Conclusion

In this chapter, we have proposed a new PZF relay beamforming design for multi-user MWRNs where single-antenna users communicate with each other with the help of a multi-antenna relay. With the help of self-interference cancellation and successive interference cancellation at the users, the proposed scheme allows more degrees-of-freedom in the beamforming design compared to the ZF design. Thus, it can improve the sum-rate.

For the case where the number of users is no larger than the number of relay antennas, the design of the PZF relay beamforming matrices is firstly transformed into the design of the equivalent relay matrices. Then, a modified gradient-ascent method is proposed to solve the optimization problem both jointly and separately. Also, the convergence behavior of the proposed algorithms is studied and computational complexity comparison is provided between the proposed methods and the existing ones. Simulations on the achievable sum-rate and symbol error rate have shown that significant performance improvement is obtained with the proposed new designs. Further, extensions of the proposed schemes are made to MWRNs with hybrid uni/multicasting transmission strategy and MWRNs where the number of users is one more than the number of relay antennas. Similar advantages have been achieved in these cases with the proposed PZF idea.

Chapter 4

Performance Analysis of Massive MIMO Multi-Way Relay Networks with Low-Resolution ADCs

In this chapter, we analyze a general mMIMO multi-way relaying system with a multi-level mixed-ADC architecture. Closed-form asymptotic approximations for the average achievable rates of ZF relaying for both perfect and imperfect CSI are derived by leveraging on Bussgang's decomposition theorem and Lloyd-Max algorithm for quantization. To handle such a challenging setup, we develop a novel method for the achievable rate analysis using distributions of the singular values of Gaussian matrices and properties of Haar matrices. Discussions on asymptotic and special cases are aslso provided. Simulation results are provided to verify our theoretical results and present important insights on the usage of low-resolution mixed-ADC architecture in a mMIMO MWRN.

4.1 Introduction and Literature Review

As mentioned in Subsection 1.2, mMIMO MWRNs, where the relays are equipped with large-scale antenna arrays, benefit from the advantages of both multi-way relaying and mMIMO structures. For mMIMO MWRNs with ZF processing, [52, 65] have obtained closed-form approximations for the spectral and energy efficiencies. It is concluded that the transmit power of each user and the relay can be made inversely proportional to the number of relay antennas while maintaining the required quality of service.

Typically, each receive and transmit antenna at a communication device is connected to a pair of ADCs and a pair of digital-to-analog converters (DACs) in the radio frequency (RF) chain, respectively. Compared to mMIMO systems with all high-resolution ADCs and DACs (e.g., 8-12 bits), it is less costly and

more energy efficient to employ low-cost, low-power, low-resolution ADCs and DACs (e.g., 1-4 bits) [71, 94–96]. Especially, the hardware cost and power consumption of ADCs grow exponentially with the number of quantization bits [97]. Naturally, due to the nonlinear characteristic of coarse quantization, signal processing challenges and complex front-end designs occur [98, 99]. In what follows, we review the literature on mMIMO systems with low-resolution ADCs.

4.1.1 Relevant Prior Work

The primary works on this topic have assumed that all ADCs have the same resolution, also referred to as uniform-ADC [71, 96, 98–101]. For instance, considering frequency-selective channels, uplink performance of a mMIMO system with uniform-ADC that deploys orthogonal frequency-division multiplexing (OFDM) is investigated in [101]. Later, a two-level mixed-ADC architecture is proposed [102, 103], in which part of the antennas are connected to low-resolution ADCs with the same resolution (usually 1-bit), while the remaining antennas are connected to high-resolution (usually the ideal infinite-resolution) ADCs. In [102], the achievable uplink spectral efficiency of a mMIMO system with two-level mixed-ADC receiver assuming perfect CSI is investigated for the MRC detector in the multi-cell scenario and the ZF detector in the single-cell system. Further, for the two-level ADC structure, the CSI obtainment schemes are proposed in [103–107], for example, by using the high-resolution ADCs in a round-robin manner [103]. Also, the mutual information for the uplink of two-level mixed-ADC systems is investigated in [103]. Recently, a general multi-level mixed-ADC structure is proposed in [108] that allows multiple ADC levels and an arbitrary ADC resolution profile for the large-scale antenna arrays. It provides more degrees-of-freedom compared to the uniform-ADC and two-level mixed-ADC architectures in achieving the desirable balance between performance and hardware and energy cost.

There have been many recent works on single-hop mMIMO systems with a mixed-ADC architecture. For the uplink of mMIMO systems with multi-level mixed-ADC architecture and MRC processing, closed-form approximations for the spectral efficiency, receive energy efficiency, and outage probability are derived in [108] and [109]. Also, these works investigate the optimization of ADC

resolutions with certain goals on the achievable sum-rate, outage probability, and receive energy efficiency. These contributions show that the power consumption and hardware cost of the single-hop mMIMO system with a mixed-ADC architecture can be considerably reduced while maintaining most of the gains in the achievable rate.

There are a few research results on the two-hop mMIMO relaying systems with low-resolution ADCs. Among them, [110–113] have investigated the performance of multi-pair mMIMO one-way relaying systems. In [110], the relay and all the users are assumed to have uniform-ADC where closed-form expressions for the achievable sum-rate are derived considering imperfect CSI and MRC/MRT processing at the relay. It is shown that with only low-resolution ADCs at the relay, increasing the number of relay antennas is effective to compensate for the rate loss caused by coarse quantization. However, it becomes ineffective to handle the detrimental effect of low-resolution ADCs at the users. Further, for mMIMO one-way relay systems with two-level mixed-ADC and MRC detection, the achievable rate is investigated in [111], where it is shown that the performance loss due to the low-resolution ADCs can be compensated by increasing the number of relay antennas. The work in [112] and [113] are on mMIMO one-way relay systems with both low-resolution ADCs and low-resolution DACs under CSI error and maximum ratio (MR) processing. In [112], for the case of uniform 1-bit ADCs and DACs, a closed-form asymptotic approximation for the achievable rate is derived. For the two-level mixed-ADCs and mixed-DACs, the work in [113] has derived exact and approximate closed-form expressions for the achievable rate. The trade-off between the achievable rate and power consumption for different numbers of low-resolution ADCs/DACs is also investigated.

4.1.2 Contributions

In this work, for the first time, we derive the average achievable rate for mMIMO multi-way relay systems with a general multi-level mixed-ADC receiver considering ZF beamforming. Both perfect and imperfect CSI cases are investigated. Our work is different from the existing literature [76, 102, 103, 108, 109, 111], and [113], which consider uplink communications or one-way relaying with a two-level mixed-ADC or a general mixed-ADC profile, in two

major ways: 1) ZF beamforming, and 2) multi-way relaying.

Firstly, all the aforementioned literature assume MR processing. While with ZF beamforming, due to the more complicated processing matrix, the theoretical derivations involve calculation of the average of an expression with the number of appearances of the Gaussian channel elements doubled compared to that of MR processing. For example, according to Equations (4) and (5) in [108], considering uplink communications or one-way relaying, the number of appearances of the Gaussian channel in the signal power term of the achievable rate analysis is 4 for MR processing, while it is 8 for ZF. Secondly, compared to the uplink communications and one-way relaying studied in the aforementioned literature, multi-way relaying further complicates the performance analysis through the following aspects. 1) It has multiple broadcast time slots, each with a distinct ZF beamforming matrix, and 2) the channel matrix of the broadcast phase is the transpose of the channel matrix of the multiple access phase, causing more contamination among different time slots. As a result, multi-way relaying further doubles the number of appearances of the Gaussian channel in some terms of the achievable rate formula. Thus, ZF along with multi-way communications leads to dealing with expressions that have up to 16 appearances of Gaussian elements compared to the existing works with up to 4 appearances. Because of these extra complexity, existing methods for the achievable rate analysis of MR processing do not apply. A new approach is proposed in this work and a brief explanation of the approach is given in the contributions summary that follows.

It is worth noting that compared to uniform-ADC and two-level mixed-ADC profiles, the multi-level mixed-ADC structure is more general and provides the system designers with extra degrees-of-freedom for the design and optimization of the system. For example, it enables the achievement of many more optimal points on the trade-off between the achievable rate and energy consumption, as shown in Figure 2 in [108] and Figure 5 in [109]. On the other hand, this general assumption imposes extra complication compared to the uniform-ADC case in performance analysis where existing methods cannot be applied directly. Furthermore, our work uses the mean-squared error-optimal sets of quantization labels and thresholds obtained from Lloyd-Max algorithm for the quantization. Also, to model the ADC quantization error, Bussgang's decomposition theorem

[73] is adopted to find the uncorrelated quantization noise to the quantization input. This is different from the frequently used AQNM that roughly models the quantization noise as an independent signal with Gaussian distribution to the quantization input. The main contributions of this work are summarized as the following:

- While existing derivation methods for mMIMO systems cannot be applied directly to the multi-way relay network with ZF in this work, we develop a new method by firstly using SVD for Gaussian matrices to simplify the expressions, then applying properties of Wishart distributed matrices, Haar matrices, and the distribution of singular values for the Gaussian matrices with i.i.d. entries. This novel method enables us to find a tight approximation for the average achievable rates which is also shown to be asymptotically tight for large number of antennas. This method can be applied in other scenarios where complicated expressions with Gaussian matrix appears in the analysis. The proposed method is fundamentally different from the truncation-based approximation in [102] and truncation error is avoided.

- In addition to the tight closed-form approximation for the average achievable rate, for two asymptotic cases, simplified expressions are derived. One case is when the number of relay antennas approaches infinity while the number of users is fixed. The result in this case reveals a linear relationship between the average achievable rate and the number of antennas at the relay. The other case is when the numbers of users and relay antennas increase toward infinity with a fixed ratio, referred to as the loading factor. The result proves an inverse linear relationship between the average achievable rate and the loading factor. In addition, we provide the result for the special case of uniform-ADC which shows that the average achievable rate has a linear relation with the square of the quantization coefficient pertaining to the ADC resolution. Also, it shows that the square of the quantization coefficient and the user power always appear together and can compensate for each other.

- The achievable rate analysis is extended to the imperfect CSI case where a closed-form asymptotic approximation is derived. It is shown that the gap

between the achievable rates for perfect and imperfect CSI cases shrinks as the average of the ADC resolutions decreases. This inspires that in practical systems with limited CSI quality, using lower ADC resolutions can gain significantly better energy efficiency and hardware cost, while maintaining most of the rate performance.

- The obtained results provide many other interesting observations about the effects of user and relay transmit power, number of relay antennas and users, channel estimation error, and most importantly ADC resolution profile on the achievable rate. For example, it is shown that in the medium to high SNR region, the ADC resolutions have more significant effects on the rate performance compared to the number of relay antennas.

4.2 System Model

Similar to Chapter 3, this work considers a MWRN consisting of K single-antenna users that exchange their information via a multi-antenna relay with N antennas, where $N \geq K$ and $N \gg 1$. Further, similarly, frequency-flat narrowband channels are assumed. Let $\mathbf{H} = \tilde{\mathbf{H}}\mathbf{D}^{\frac{1}{2}}$ be the $N \times K$ channel matrix between the users and the relay, where $\tilde{\mathbf{H}} \in \mathbb{C}^{N\times K}$ is the fast fading channel matrix whose entries are i.i.d. circularly symmetric complex Gaussian with zero mean and unit variance, i.e., $\mathcal{CN}(0,1)$. Moreover, $\mathbf{D} \in \mathbb{R}^{K\times K}$ is a diagonal matrix whose kth diagonal element denoted as β_k, stands for the large-scale fading of the channels from u_k to the relay. We define

$$\beta_{\text{sum}} \triangleq \text{tr}\{\mathbf{D}\} = \sum_{k=1}^{K} \beta_k, \quad \beta_{\setminus i} = \sum_{k=1, k\neq i}^{K} \beta_k.$$

Also, denote the kth columns of $\tilde{\mathbf{H}}$ and $\mathbf{H}$ as $\tilde{\mathbf{h}}_k$ and $\mathbf{h}_k$ which are the fast fading and overall channel vectors from u_k to the relay, respectively. It is assumed that the relay has perfect CSI, and the imperfect CSI case is considered in Section 4.5.

4.2.1 The MAC Phase and ADC Quantization

As mentioned in Section 3.2.2, in the MAC phase, all users transmit their information signals simultaneously to the relay. Hence, similar to (3.1), the

baseband representation of the received discrete-time analog-valued signal[1] at the relay, denoted as $\mathbf{r}_{\mathrm{a}} \in \mathbb{C}^{N\times 1}$ can be written as

$$\mathbf{r}_{\mathrm{a}} = \sqrt{p_u}\mathbf{H}\mathbf{x} + \mathbf{z}_{\mathrm{R}} = \sum_{i=1}^{K} \sqrt{p_u}\mathbf{h}_i x_i + \mathbf{z}_{\mathrm{R}}, \tag{4.1}$$

Denote the nth element of $\mathbf{r}_{\mathrm{a}}$ as $r_{n,\mathrm{a}}$.

We assume that each relay receive antenna is equipped with a radio-frequency chain including a pair of low-resolution ADCs for the in-phase and quadrature components. A generic mixed-ADC structure in which the ADC pairs of different antennas can have arbitrary resolutions is considered. The ADC resolution of the nth antenna is denoted by b_n bits which is a positive integer value between $b_{\min}$ and $b_{\max}$. Then, let $\mathbf{b} = [b_1, \cdots, b_N]$ which is the resolution profile of the relay antennas. The ADC quantization corresponding to the nth antenna can be characterized by a set of $2^{b_n}+1$ quantization thresholds $\mathcal{T}_{b_n} = \{\tau_{n,0}, \tau_{n,1}, \cdots, \tau_{n,2^{b_n}}\}$, where $-\infty = \tau_{n,0} < \tau_{n,1} < \cdots < \tau_{n,2^{b_n}} = \infty$ and a set of 2^{b_n} quantization labels $\mathcal{L}_{b_n} = \{l_{n,0}, l_{n,1}, \cdots, l_{n,2^{b_n}-1}\}$, where $l_{n,i} \in (\tau_{n,i}, \tau_{n,i+1}]$. We describe the joint operation of the nth ADC pair at the relay by the function $\mathcal{Q}_{b_n}(\cdot) : \mathbb{C} \to \mathcal{R}_{b_n}$, where $\mathcal{R}_{b_n} \triangleq \mathcal{L}_{b_n} \times \mathcal{L}_{b_n}$. The quantized signal vector at the relay is denoted by $\hat{\mathbf{r}}$ with $\hat{r}_n$ being its nth entry corresponding to the nth antenna. The quantization function $\mathcal{Q}_{b_n}(\cdot)$ maps the analog received signal, $r_{n,\mathrm{a}}$, to the quantized signal, $\hat{r}_n$, in a way that

$$\begin{aligned} &\hat{r}_n = \mathcal{Q}_{b_n}(r_{n,\mathrm{a}}) = l_{n,k} + j l_{n,p}, \\ &\text{if } \Re\{r_{n,\mathrm{a}}\} \in (\tau_{n,k}, \tau_{n,k+1}] \text{ and } \Im\{r_{n,\mathrm{a}}\} \in (\tau_{n,p}, \tau_{n,p+1}]. \end{aligned}$$

Therefore, the quantized vector at the relay is

$$\hat{\mathbf{r}} = \mathcal{Q}(\mathbf{r}_{\mathrm{a}}) = \mathcal{Q}(\sqrt{p_u}\mathbf{H}\mathbf{x} + \mathbf{z}_{\mathrm{R}}), \tag{4.2}$$

where $\mathcal{Q}(\cdot)$ is the function that quantizes the nth entry of its input vector using $\mathcal{Q}_{b_n}(\cdot)$ for $n \in \{1, 2, \cdots, N\}$.

The optimal sets of $\mathcal{L}_{b_n}$ and $\mathcal{T}_{b_n}$ for $n \in \{1, 2, \cdots, N\}$ that minimize the MSE between the non-quantized received vector $\mathbf{r}_{\mathrm{a}}$ and the quantized vector $\hat{\mathbf{r}}$

[1] This is referred to as "analog signal" for short afterwards.

depends on the distribution of $\mathbf{r}_{\mathrm{a}}$ which changes with respect to the channels and the information signals. From a practical point of view [72], we use the set of quantization labels and the set of thresholds that are optimal for Gaussian signals[2]. From (4.1), the variance of each entry of $\mathbf{r}_{\mathrm{a}}$ can be straightforwardly calculated to be

$$v \triangleq 1 + p_u \beta_{\mathrm{sum}}. \tag{4.3}$$

Then, using Lloyd-Max algorithm for quantization [69, 70], we can find the optimal sets of labels and thresholds, $\mathcal{L}_{b_n}^* = \{l_{n,0}^*, l_{n,1}^*, \cdots, l_{n,2^{b_n}-1}^*\}$ and $\mathcal{T}_{b_n}^* = \{\tau_{n,0}^*, \tau_{n,1}^*, \cdots, \tau_{n,2^{b_n}}^*\}$ that minimize the MSE assuming that the analog signal follows $\mathcal{CN}(0, v)$.

With the set of labels $\mathcal{L}_{b_n}^*$, the set of thresholds $\mathcal{T}_{b_n}^*$, and the Gaussian assumption of $r_{n,\mathrm{a}}$, the variance of $\hat{r}_n$ denoted as $C_{n,\hat{r}}$ can be straightforwardly obtained based on the definition of variance, quantization function, and Gaussian distribution as the following,

$$C_{n,\hat{r}} = \sum_{i=0}^{2^{b_n}-1} {l_{n,i}^*}^2 \left[\left(\mathrm{erf}\left(\frac{\tau_{n,i+1}^*}{\sqrt{v}} \right) - \mathrm{erf}\left(\frac{\tau_{n,i}^*}{\sqrt{v}} \right) \right) \right]. \tag{4.4}$$

Thus, the covariance matrix and the average power of the quantized vector at the relay are respectively,

$$\mathbf{C}_{\hat{\mathbf{r}}} = \mathrm{diag}\{C_{1,\hat{r}}, C_{2,\hat{r}}, \cdots, C_{N,\hat{r}}\}, \tag{4.5}$$

$$\hat{c} \triangleq \frac{1}{N} \mathrm{tr}\{\mathbf{C}_{\hat{\mathbf{r}}}\} = \frac{1}{N} \sum_{n=1}^{N} C_{n,\hat{r}}. \tag{4.6}$$

[2]The Gaussian assumption is accurate in the low-SNR regime or when the number of users is sufficiently large [114]. Simulation results have verified the validity of this assumption for normal ranges of user number and SNR.

4.2.2 The BC Phase

The BC phase takes $K-1$ time slots. ZF beamforming is employed at the relay which according to (2.3), has the following matrix for the tth time slot

$$\mathbf{G}^{(t)} = \sqrt{\alpha^{(t)}}\mathbf{H}^{*}(\mathbf{H}^{T}\mathbf{H}^{*})^{-1}\mathbf{P}^{t}(\mathbf{H}^{H}\mathbf{H})^{-1}\mathbf{H}^{H}, \tag{4.7}$$

where $\mathbf{P}$ is the permutation matrix as in (2.1), and $\alpha^{(t)}$ is the ZF power coefficient used to fulfill the relay transmit power constraint. Then, the transmit signal at the relay is

$$\mathbf{r}_{\mathrm{t}}^{(t)} = \mathbf{G}^{(t)}\hat{\mathbf{r}}. \tag{4.8}$$

Let P_R denote the average transmission power of the relay. Then, $\alpha^{(t)}$ must satisfy $P_R = \mathbb{E}\{\|\mathbf{r}_{\mathrm{t}}^{(t)}\|^2\}$. Therefore, similar to (3.2), the received signal vector of all users in the BC time slot t, $\mathbf{r}_{\mathrm{u}}^{(t)}$, is

$$\mathbf{r}_{\mathrm{u}}^{(t)} = \mathbf{H}^{T}\mathbf{r}_{\mathrm{t}}^{(t)} + \mathbf{z}_{\mathrm{u}}^{(t)}. \tag{4.9}$$

In the BC time slot t, user k is supposed to decode user $i(k,t)$'s information symbol, which is defined according to the permutation matrix, as in (2.2). As mentioned before, to help the presentation, the notation $i(k,t)$ is simplified to $i(k)$ when there is no confusion about the time index.

4.3 Average Achievable Rate Analysis

As mentioned earlier, in this section, we consider the perfect CSI case. First, we analyze the ADC quantization process, then, derive the relay power coefficient for the ZF beamforming in each BC time slot, and finally, obtain the average achievable rate of the considered MWRN.

As discussed in Section 2.3.1, generally, the quantization process at low-resolution ADCs leads to a distortion that is correlated to the quantization input. However, according to Bussgang's theorem [73], when the input to the low-resolution ADCs is Gaussian, the output can be written as a linear combination of the input and a distortion that is uncorrelated to the input. Consider $\mathbf{d}$ as the uncorrelated distortion to $\mathbf{r}_{\mathrm{a}}$, and G_{b_n} as the quantization

coefficient of the nth ADC pair where

$$G_{b_n} = \frac{1}{\sqrt{\pi v}} \sum_{i=0}^{2^{b_n}-1} l_{n,i}^* \left[\exp\left(-\frac{{\tau_{n,i}^*}^2}{v} \right) - \exp\left(-\frac{{\tau_{n,i+1}^*}^2}{v} \right) \right], \tag{4.10}$$

$\mathbf{G}_b = \text{diag}\{G_{b_1}, \cdots, G_{b_N}\}$. Then, due to the Gaussian assumption of $\mathbf{r}_\text{a}$, the quantized received signal vector at the relay can be written as,

$$\begin{aligned} \hat{r}_n &= \mathcal{Q}_{b_n}(r_{n,\text{a}}) = G_{b_n} r_{n,\text{a}} + d_n, \text{or} \\ \hat{\mathbf{r}} &= \mathcal{Q}(\mathbf{r}_\text{a}) = \mathbf{G}_b \mathbf{r}_\text{a} + \mathbf{d}. \end{aligned} \tag{4.11}$$

To help the presentation of the performance analysis, we define

$$g_1 \triangleq \frac{1}{N}\text{tr}\{\mathbf{G}_b\} = \frac{1}{N}\sum_{n=1}^{N} G_{b_n} \text{ and } \quad g_2 \triangleq \frac{1}{N}\text{tr}\{\mathbf{G}_b^2\} = \frac{1}{N}\sum_{n=1}^{N} G_{b_n}^2,$$

which represent the average of the quantization coefficients and quantization coefficients squared. Further, having the variance of each entry of $\mathbf{r}_\text{a}$ as v defined in (4.3), the covariance matrix of the quantization distortion, $\mathbf{C_d} = \mathbb{E}[\mathbf{d}\mathbf{d}^H]$, can be found as

$$\mathbf{C_d} = \mathbf{C}_{\hat{\mathbf{r}}} - \mathbb{E}\{\mathbf{G}_b \mathbf{r}_\text{a} \mathbf{r}_\text{a}^H \mathbf{G}_b\} = \mathbf{C}_{\hat{\mathbf{r}}} - v\mathbf{G}_b^2. \tag{4.12}$$

The closed-form expression for the ZF power coefficient of the considered MWRN is presented in the following theorem.

Theorem 1. *For a multiuser mMIMO MWRN with K single-antenna users, N relay antennas, mixed-ADC with resolution profile $\mathbf{b}$ at the relay, relay power constraint P_R, and perfect CSI, the relay power coefficient for ZF beamforming in the BC time slot t is*

$$\alpha^{(t)} \approx \frac{P_R(N-K)}{p_u g_1^2 \sum_{m=1}^{K} \frac{1}{\beta_m} + \frac{NK+N-K^2}{(N-K)^2(K+1)} \left(\hat{c} - p_u \beta_\text{sum} g_2\right) \sum_{m=1}^{K} \frac{1}{\beta_m \beta_{i(m,t)}}}. \tag{4.13}$$

Proof: Please see Appendix A.1.

Next, we calculate the average achievable rate from user $i(k)$ to k where $i(k)$ is given in (2.2).

From (4.9) and Bussgang's decomposition in (4.11), the received signal vector at the users in the BC time slot t can be written as

$$\mathbf{r}_{\mathrm{u}}^{(t)} = \sqrt{p_u}\mathbf{H}^T\mathbf{G}^{(t)}\mathbf{G}_b\mathbf{H}\mathbf{x} + \mathbf{H}^T\mathbf{G}^{(t)}\mathbf{G}_b\mathbf{z}_{\mathrm{R}} + \mathbf{H}^T\mathbf{G}^{(t)}\mathbf{d} + \mathbf{z}_{\mathrm{u}}^{(t)}. \tag{4.14}$$

Hence, the received signal by u_k, denoted as ${r_k}^{(t)}$, is

$$\begin{aligned} {r_k}^{(t)} =& \sqrt{p_u}{\mathbf{h}_k}^T\mathbf{G}^{(t)}\mathbf{G}_b\mathbf{h}_{i(k)}x_{i(k)} + \sqrt{p_u}{\mathbf{h}_k}^T\mathbf{G}^{(t)}\mathbf{G}_b\sum_{j=1,j\neq i(k)}^{N}\mathbf{h}_j x_j \\ &+ {\mathbf{h}_k}^T\mathbf{G}^{(t)}\mathbf{G}_b\mathbf{z}_{\mathrm{R}} + \mathbf{h}_k^T\mathbf{G}^{(t)}\mathbf{d} + {z_k}^{(t)}, \end{aligned} \tag{4.15}$$

where in the right-hand-side of (4.15), the first term represents the desired signal, the second term represents the interference from other users, the third term shows the effect of $\mathbf{z}_{\mathrm{R}}$ (the received noise at the relay in the MAC phase) under the relay beamforming design in (4.7), and the fourth term shows the effect of $\mathbf{d}$ (the quantization distortion caused by the low-resolution ADCs at the relay). Thus, the interference-plus-noise power is

$$I_{k,i(k)} = p_u\sum_{j=1,j\neq i(k)}^{K}|\mathbf{h}_k^T\mathbf{G}^{(t)}\mathbf{G}_b\mathbf{h}_j|^2 + \|{\mathbf{h}_k}^T\mathbf{G}^{(t)}\mathbf{G}_b\|^2 + \|{\mathbf{h}_k}^T\mathbf{G}^{(t)}\mathbf{d}\|^2 + 1. \tag{4.16}$$

It can be seen from the first term in (4.16) that due to the mixed-ADC structure at the relay, the user interference is not fully eliminated by the ZF design in (4.7). For the special case of uniform-ADC structure [50], $\mathbf{G}_b = G_b\mathbf{I}_N$, and thus, the user interference can be fully eliminated by ZF. From (4.15) and (4.16), the average achievable rate from $u_{i(k)}$ to u_k, denoted by $R_{k,i(k)}$, is given as

$$R_{k,i(k)} = \mathbb{E}\left\{\log_2\left(1 + \frac{p_u|{\mathbf{h}_k}^T\mathbf{G}^{(t)}\mathbf{G}_b\mathbf{h}_{i(k)}|^2}{I_{k,i(k)}}\right)\right\}. \tag{4.17}$$

A closed-form expression for the average achievable rate between the two users of the considered MWRN is presented in the following theorem.

Theorem 2. *For a multiuser mMIMO MWRN with K single-antenna users, N relay antennas, ZF beamforming in (4.7), mixed-ADC with resolution profile* $\mathbf{b}$ *at the relay, and perfect CSI, the average achievable rate from user $i(k)$ to*

user k is

$$R_{k,i(k)} \approx \log_2 \left(1 + \frac{p_u \beta_{i(k)} \left[\frac{(NK+N-K^2-2K)g_1^2+Kg_2}{K+1} \right]}{\hat{c} + \frac{(N-K)}{\alpha^{(t)}} \beta_{i(k)} - p_u \beta_{\setminus i(k)} g_1^2 - p_u \beta_{i(k)} g_2} \right). \tag{4.18}$$

Proof: Please see Appendix A.2.

The formulas (4.13) and (4.18) show how quantitatively the average achievable rate is affected by system settings, such as the user and relay power, and number of users and relay antennas. The effect of the ADC resolutions is shown via g_1 and g_2. Furthermore, it can be seen that the average achievable rate increases as p_u or P_R increases.

4.4 Results for Asymptotic Cases and Uniform-ADC

In what follows, we discuss two asymptotic cases and the special uniform-ADC case to gain more insights on the effects of system parameters on the average achievable rate.

4.4.1 Asymptotic Cases

The first asymptotic case that is commonly considered in mMIMO is when $N \to \infty$ with fixed K. The result on the achievable rate can be simplified from (4.13), (4.18) as

$$R_{k,i(k),1} \approx \log_2 \left(1 + \frac{N p_u \beta_{i(k)} g_1^2}{\hat{c} + \frac{p_u}{P_R} \beta_{i(k)} \left(\sum_{m=1}^{K} \frac{1}{\beta_m} \right) g_1^2 - p_u \beta_{\setminus i(k)} g_1^2 - p_u \beta_{i(k)} g_2} \right). \tag{4.19}$$

It shows that the SINR increases linearly in N. The effect of ADC resolutions is shown through $\hat{c}$, g_1, and g_2. Further, the expression directly shows that the sum-rate is monotonically increasing with respect to p_u and P_R, but with a finite ceiling as $p_u \to \infty$ or $P_R \to \infty$. The ceiling depends on the ADC resolution profile, indicating that the performance penalty brought by low-resolution ADCs may not be fully compensated by increasing the user or relay

transmit power.

The second asymptotic case is when $N, K \to \infty$ with a fixed ratio $K/N = c$. The constant c is referred to as the loading factor. Notice that β_{sum} is linear in K, thus v given in (4.3) and $\hat{c}$ given in (4.5) and (4.6) are also linear in K. We assume that as $K \to \infty$, the values $\bar{\beta}_{-1} \triangleq \sum_{m=1}^{K} 1/\beta_m$, $\bar{\beta}_{-2} \triangleq \sum_{m=1}^{K} 1/(\beta_m \beta_{i(m)})$, $\bar{\beta} \triangleq \beta_{\text{sum}}/K$, and $\bar{\hat{c}} \triangleq \hat{c}/K$, all converge to positive constants. It can be shown that $\lim_{K\to\infty} \left(\bar{\beta} - \beta_{\setminus i(k)}/K\right) = 0$. Thus, for this case, the ZF power coefficient is

$$\alpha_2^{(t)} \approx \frac{NP_R}{\frac{1}{1-c} p_u g_1^2 \bar{\beta}_{-1} + \frac{c}{(1-c)^2}\left(\bar{\hat{c}} - p_u \bar{\beta} g_2\right) \bar{\beta}_{-2}},$$

and the average achievable rate is simplified as

$$R_{k,i(k),2} \approx \log_2 \left(1 + \frac{(1-c)}{c} \cdot \frac{p_u \beta_{i(k)} g_1^2}{(\bar{\hat{c}} - p_u \bar{\beta} g_1^2)}\right). \tag{4.20}$$

It shows that when the number of users increases linearly with the number of relay antennas, the average achievable rate becomes independent of N and decreases as c increases. The SINR for this asymptotic case is linear in $(1-c)/c$. The effect of ADC resolution profile is shown through $\bar{\hat{c}}$ and g_1. Further, the expression indicates that the average achievable rate degradation caused by low-resolution ADCs can be compensated by decreasing c. It also shows that the sum-rate is monotonically increasing with respect to p_u, but with a finite ceiling as $p_u \to \infty$. Similar to the previous case, the ceiling depends on the ADC resolution profile, indicating that the penalty brought by low-resolution ADCs may not be fully compensated by increasing the user power. Finally, it indicates that the average achievable rate is independent of P_R in this case.

4.4.2 Uniform-ADC Case

For the special case of uniform-ADC with b-bit resolution, we denote the quantization coefficient as G_b and the variance of the quantization output as $C_{\hat{r}}$. In this case, we have $G_{b_n} = G_b$ and $C_{n,\hat{r}} = C_{\hat{r}}$ for all n. Consequently, the ZF

relay power coefficient and the average achievable rate can be simplified as

$$\alpha_{\text{uni}}^{(t)} \approx \frac{P_R(N-K)}{\sum_{m=1}^{K} \frac{1}{\beta_m}\left[p_u G_b^2 + \frac{1}{\beta_{i(m)}}(C_{\hat{r}} - p_u \beta_{\text{sum}} G_b^2)\frac{(NK+N-K^2)}{(N-K)^2(K+1)}\right]},$$

$$R_{k,i(k),\text{uni}} \approx \log_2\left(1 + \frac{(N-K)\beta_{i(k)}}{(\sqrt{p_u}G_b)^{-2}\left[\frac{N-K}{\alpha_{\text{uni}}^{(t)}}\beta_{i(k)} + C_{\hat{r}}\right] - \beta_{\text{sum}}}\right). \tag{4.21}$$

The results reveal that the user power, quantization coefficient squared, and the relay power have similar effects on the achievable rate. Increasing each of them has a positive effect with a negative acceleration on the achievable rate. Furthermore, G_b^2 and p_u appear together as a product in the formulas meaning that they can be adjusted to compensate each other's contribution to the achievable rate. Also, it can be concluded that the achievable rate linearly decreases with the number of users while it has an increasing relation with the number of antennas. Finally, comparing (4.18) with (4.21) indicates that the average achievable rate has an almost linear relation with the square of the average of quantization coefficients of the ADC resolution profile.

4.5 Extension to the Imperfect CSI Case

In this section, we extend our results to the imperfect CSI case. We use the widely used channel model[3]: $\tilde{\mathbf{H}} = \hat{\tilde{\mathbf{H}}} + \Delta\tilde{\mathbf{H}}$, where $\hat{\tilde{\mathbf{H}}} \sim \mathcal{CN}(\mathbf{0}, (1-\sigma_e^2)\mathbf{I}_N)$ is the estimation of the small-scale fading channel, $\Delta\tilde{\mathbf{H}} \sim \mathcal{CN}(\mathbf{0}, \sigma_e^2\mathbf{I}_N)$ is the CSI error, and σ_e^2 represents the power of the CSI error. Also, $\hat{\tilde{\mathbf{H}}}$ and $\Delta\tilde{\mathbf{H}}$ are assumed to be independent. The ZF relay beamforming matrix considering the channel estimation $\hat{\tilde{\mathbf{H}}}$, is

$$\hat{\mathbf{G}}^{(t)} = \sqrt{\hat{\alpha}^{(t)}}\hat{\mathbf{H}}^*(\hat{\mathbf{H}}^T\hat{\mathbf{H}}^*)^{-1}\mathbf{P}^t(\hat{\mathbf{H}}^H\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}^H, \tag{4.22}$$

[3]Channel estimation schemes for MIMO systems with low- or mixed-resolution ADCs can be found in [103–106].

where $\hat{\mathbf{H}} = \hat{\tilde{\mathbf{H}}}\mathbf{D}^{\frac{1}{2}}$. Therefore, the received signal by user k is

$$\begin{aligned} r_{k,\mathrm{ICSI}}^{(t)} \approx & \sqrt{p_u}\mathbf{h}_k{}^T\hat{\mathbf{G}}^{(t)}\mathbf{G}_b\mathbf{h}_{i(k)}x_{i(k)} + \sqrt{p_u}\mathbf{h}_k{}^T\hat{\mathbf{G}}^{(t)}\mathbf{G}_b\sum_{j=1,j\neq i(k)}^{N}\mathbf{h}_j x_j \\ & + \mathbf{h}_k{}^T\hat{\mathbf{G}}^{(t)}\mathbf{G}_b\mathbf{z}_{\mathrm{R}} + \mathbf{h}_k^T\hat{\mathbf{G}}^{(t)}\mathbf{d} + z_k{}^{(t)}, \end{aligned} \tag{4.23}$$

and the power of the interference-plus-noise terms is

$$I_{k,i(k),\mathrm{ICSI}} = p_u\sum_{j=1,j\neq i(k)}^{K}|\mathbf{h}_k^T\hat{\mathbf{G}}^{(t)}\mathbf{G}_b\mathbf{h}_j|^2 + \|\mathbf{h}_k{}^T\hat{\mathbf{G}}^{(t)}\mathbf{G}_b\|^2 + \|\mathbf{h}_k{}^T\hat{\mathbf{G}}^{(t)}\mathbf{d}\|^2 + 1. \tag{4.24}$$

From (4.23) and (4.24), the average achievable rate from user $i(k)$ to user k can be written as

$$R_{k,i(k),\mathrm{ICSI}} \approx \mathbb{E}\left\{\log_2\left(1 + \frac{p_u|\mathbf{h}_k{}^T\hat{\mathbf{G}}^{(t)}\mathbf{G}_b\mathbf{h}_{i(k)}|^2}{I_{k,i(k),\mathrm{ICSI}}}\right)\right\}. \tag{4.25}$$

The ZF transmit power coefficient is found as the following theorem.

Theorem 3. *For a multiuser mMIMO MWRN with K single-antenna users, N relay antennas, mixed-ADC with resolution profile $\mathbf{b}$ at the relay, relay power constraint P_R, and imperfect CSI with error power σ_e^2, the relay power coefficient for ZF beamforming in the BC time slot t is*

$$\hat{\alpha}^{(t)} \approx \frac{P_R(N-K)(1-\sigma_e^2)}{p_u g_1^2\sum_{m=1}^{K}\frac{1}{\beta_m} + \frac{NK+N-K^2}{(N-K)^2(K+1)(1-\sigma_e^2)}\left(\hat{c} - p_u\beta_{\mathrm{sum}}g_2\left(1-\sigma_e^2\right)\right)\sum_{m=1}^{K}\frac{1}{\beta_m\beta_{i(m)}}}. \tag{4.26}$$

Proof: Please see Appendix A.3.

Further, the average achievable rate can be found as the following theorem.

Theorem 4. *For a multiuser mMIMO MWRN with K single-antenna users, N relay antennas, ZF beamforming in (4.22), mixed-ADC with resolution profile $\mathbf{b}$ at the relay, and imperfect CSI with error power σ_e^2, the average achievable*

rate from $u_{i(k)}$ to u_k is

$$R_{k,i(k),\mathrm{ICSI}} \approx \log_2\left(1+\frac{\hat{F}_1}{\hat{F}_2}\right), \tag{4.27}$$

where

$$\begin{aligned}\hat{F}_1 \triangleq & \frac{p_u\beta_{i(k)}}{K+1}\Bigg[g_1^2(1-\sigma_e^2)^2(NK+N-K^2-2K)\\ & +g_2\left((1-\sigma_e^2)(K+\sigma_e^2)+\frac{(NK+N-K^2)}{(N-K)^2}\beta_k\beta_{i(k)}\sigma_e^4\sum_{m=1}^{K}\frac{1}{\beta_m\beta_{i(m)}}\right)\Bigg],\end{aligned}$$

$$\begin{aligned}\hat{F}_2 \triangleq & \hat{c}+\frac{(N-K)(1-\sigma_e^2)}{\hat{\alpha}^{(t)}}\beta_{i(k)}-p_u\beta_{i(k)}g_2+p_ug_1^2\Bigg(-\beta_{\setminus i(k)}(1-\sigma_e^2)\\ & +\beta_{i(k)}\beta_k\sigma_e^2\sum_{j=1,j\neq k}^{K}\frac{1}{\beta_j}+\frac{\beta_{i(k)}\beta_k^2\sigma_e^4N}{(N-K)K(1-\sigma_e^2)}\sum_{m=1}^{K}\frac{1}{\beta_m\beta_{i(m)}}\Bigg).\end{aligned} \tag{4.28}$$

Proof: Please see Appendix A.4.

$\hat{F}_1$ and $\hat{F}_2$ in Theorem 4 show the dominant terms in the SINR numerator and denominator. This theorem indicates that the effect of the decrease in the average achievable rate due to the channel estimation error, which is shown by $(1-\sigma_e^2)$, gets scaled by g_1^2 and g_2 in the numerator. In other words, it shows that the gap between the rates of perfect and imperfect CSI cases increases as the average ADC resolution gets higher.

4.6 Simulation Results

This section shows the simulation results for the average achievable rates of mMIMO MWRNs with low-resolution ADCs. The closed-form results in Theorems 2 and 4 are compared with the Monte-Carlo simulated ones. Also, the results for asymptotic and special cases in (4.19)-(4.21) are compared with the general theoretical ones. While $R_{1,2}$, the average achievable rate of user 2 at user 1 is shown, similar results can be obtained for other user pairs.

The simulations consist of two steps: the quantizer optimization and the

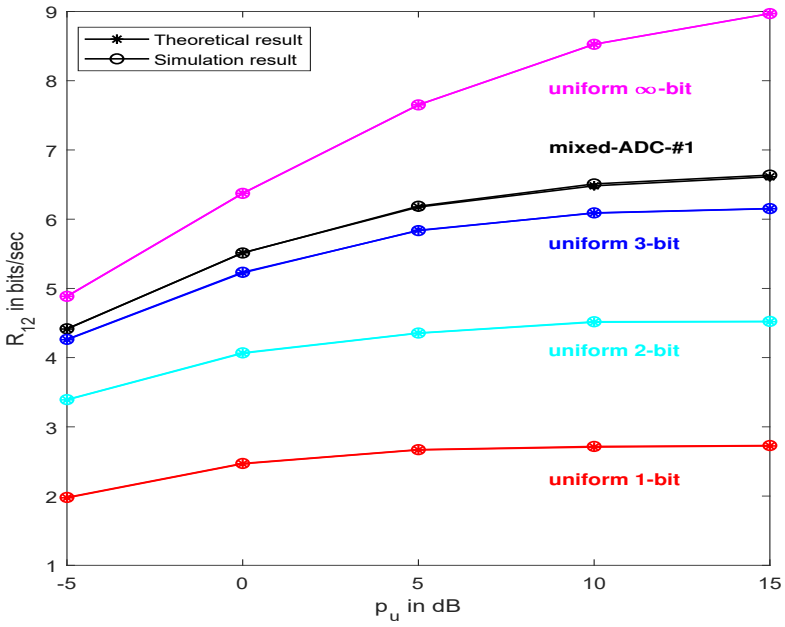


Figure 4.1: Theoretical and simulation rate results versus user power for different ADC profiles at a MWRN with $N = 100$, $K = 5$, and $P_R = 15$ dB.

achievable rate simulation. For the first step, a training set of 10^5 points is generated based on complex Gaussian channels and QAM at the users. Then, Lloyd-Max algorithm is used to find the quantizer for each value in the resolution profile $\mathbf{b}$. For the second step, the quantizers obtained in the previous step are used. A total of 10^3 channel realizations are generated, and for each channel 10^2 information vectors are generated based on QAM. Unless otherwise mentioned, networks with homogeneous channels, i.e., the same large-scale fading for all users, are considered with $\beta_k = 0$ dB for $k \in \{1, 2, \cdots, K\}$.

Table 4.1: The number of antennas with each resolution level in mixed-ADC-#1 profile.

N \ bits	1	2	3	4	5	6	7	8
50	3	6	7	6	7	11	5	5
100	7	12	15	11	14	22	9	10

In Figure 4.1, a MWRN with $N = 100$, $K = 5$, and $P_R = 15$ dB is considered and the simulation and theoretical results obtained by (4.13) and (4.18) are compared when p_u changes from -5 to 15 dB. Five resolution profiles

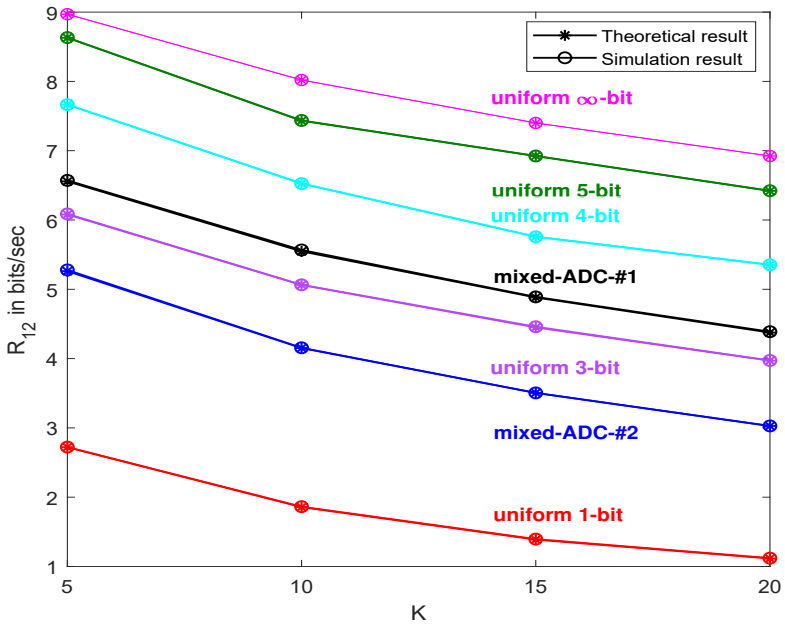


Figure 4.2: Theoretical and simulation rate results versus the number of users for different ADC profiles at MWRNs with $N = 100$, $p_u = P_R = 15$ dB.

are considered: uniform 1-bit, 2-bit, 3-bit, ∞-bit ADCs, and a mixed-ADC one (referred to as mixed-ADC-#1) which is specified in Table 4.1. In mixed-ADC-#1 profile, the ADC resolution for each antenna is randomly and independently generated according to the discrete uniform distribution on $[1, 8]$. It is shown that the simulation and theoretical results perfectly match for all power range and ADC profiles. Also, the figure shows the rate degradation due to low-resolution ADCs, especially in the high SNR region. For instance, when $p_u = 15$ dB, the achievable rate for mixed-ADC-#1 is about 74% of the full precision case, implying the importance of the ADC resolutions on the rate performance of mMIMO MWRNs. Finally, it can be observed that for all low-resolution ADC cases, as p_u increases, the achievable rate quickly saturates.

Figure 4.2 shows the rate results when $N = 100$, $p_u = P_R = 15$ dB, and $K = 5, 10, 15, 20$. Five uniform-ADC profiles with the resolution levels of $1, 3, 4, 5, \infty$ are tested. Also, two mixed-ADC profiles are examined: the mixed-ADC-#1 explained in Table 4.1 and the mixed-ADC-#2 for which the resolutions are 1 to 4 bits and the numbers of antennas are 21, 27, 22, and 30, respectively, for the 4 resolution levels. This figure confirms the perfect match between the

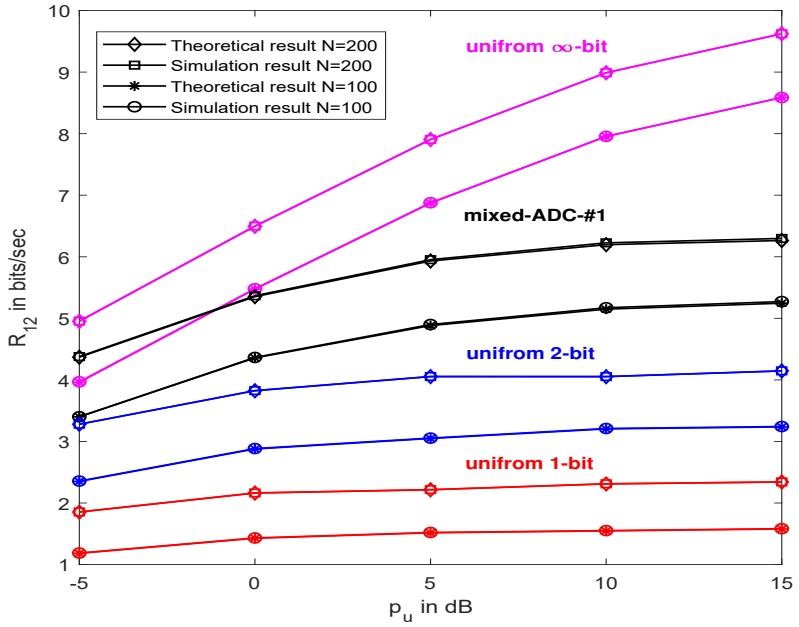


Figure 4.3: Theoretical and simulation rate results versus user power for MWRNs with heterogeneous channels when $K = 5$, $\beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5, \beta_4 = 2, \beta_5 = 3$ and $P_R = 15$ dB for $N = 100, 200$.

simulation and theoretical results for all numbers of users and ADC profiles. It also reveals the degradation in the achievable rate with the increase in the number of users. This is due to the decrease in the ZF power coefficient, $\alpha^{(t)}$, that causes loss in the SINR. Both Figures 1 and 2 show that the higher the average of ADC resolution profile is, the higher the average achievable rate is.

Next, a MWRN with 5 users and heterogeneous channels, i.e., each channel may have a different large-scale fading, is considered where $\beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5, \beta_4 = 2, \beta_5 = 3$, and $P_R = 15$ dB. Resolution profiles of uniform 1-bit, 2-bit, and ∞-bit along with the mixed-ADC-#1 are considered. For $N = 100$, the mixed-ADC-#1 profile is shown in Table 4.1. For $N = 200$ the number of ADC pairs for each resolution level is twice the number for $N = 100$. Figure 4.3 approves the prefect match between the derivation and simulation results for all cases when p_u changes from -5 to 15 dB. An important observation is that in the medium to high SNR region, increasing the ADC resolutions has a higher impact on the rate performance compared to increasing the number of antennas. As an example, we consider System 1 and System 2 with

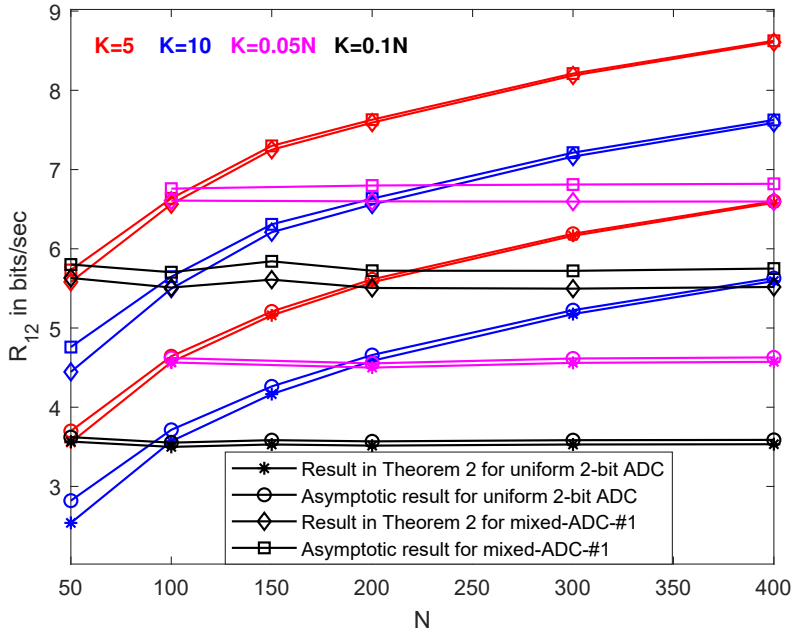


Figure 4.4: Result in Theorem 2 is compared with the asymptotic result in (4.19) when $K = 5, 10$ (in red and blue), and the asymptotic result in (4.20) when $c = 0.05, 0.1$ (in pink and black), in all simulations $p_u = P_{\mathrm{R}} = 15$ dB.

uniform 1-bit ADCs where $N = 100$ and $N = 200$, respectively. System 3 is with uniform 2-bit ADCs where $N = 100$. When $p_u = -5$ dB, the sum-rate improvement of System 2 over System 1 is 56.2%, while System 3 has a higher rate improvement of 98.56% over System 1.

Figure 4.4 compares the results in Theorem 2 with the two asymptotic results in (4.19) and (4.20). For the first asymptotic case, $K = 5, 10$ are tested and for the second one, $c = 0.05, 0.1$ are tested. The number of antennas changes from 50 to 400 and when $K = 0.05N$, N takes multiples of 100. The results are shown for two ADC profiles: uniform 2-bit and mixed-ADC-#1. For mixed-ADC-#1, the ADC profiles for $N = 50$ and 100 are specified in Table 4.1. For $N = 150, 200, 300$, and 400 the ADC profiles are found by scaling the ADC profile for $N = 50$, three times and the ADC profile for $N = 100$, two, three, and four times, respectively. Figure 4.4 confirms that our asymptotic analysis for case 1 perfectly matches the general results for $N \geq 200$, while it is a tight upper bound for $N < 200$. Also, the gap between the results from (4.19) and Theorem 2 shrinks as the number of users decreases. In addition,

Figure 4.4 indicates that the asymptotic result for case 2 works well for the uniform-ADC case while for mixed-ADC case, it is an upper bound with a small gap that shrinks as c decreases. The reason for this gap is that unlike the uniform-ADC case, in the mixed-ADC case, the user interference is not fully canceled by ZF beamforming. Therefore, the ignorance of user interference due to the assumption of large number of users makes a gap between the simulation results and asymptotic analysis. This figure also confirms that for asymptotic case 1, the rate linearly increases with N, while for case 2 it only increases if c decreases.

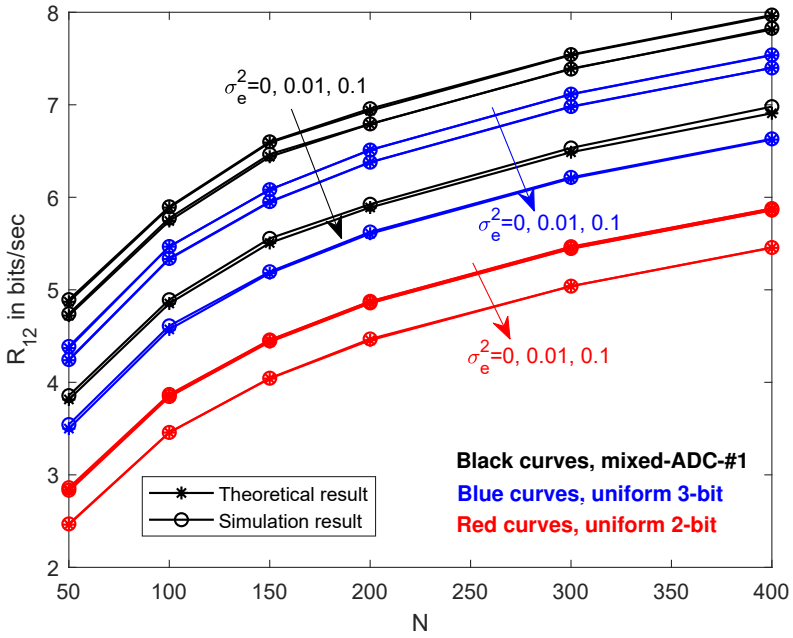


Figure 4.5: Theoretical and simulation rate results versus the number of relay antennas at the MWRN for the imperfect CSI cases, $\sigma_e^2 = 0, 0.01, 0.1$, when $K = 8$ and $p_u = P_R = 15$ dB.

Finally, Figure 4.5 illustrates the rate results versus the number of relay antennas for the imperfect CSI cases $\sigma_e^2 = 0, 0.01, 0.1$, when $K = 8$, and $p_u = P_R = 15$ dB. This figure shows that the results in (4.26)-(4.28) are accurate and higher estimation errors lead to lower rates. In addition, it shows that the higher the average of ADC resolutions is, the larger the gap between the rates of perfect and imperfect CSI cases is, implying that systems with lower ADC resolutions are less sensitive to the CSI error.

4.7 Conclusion

In this chapter, we have presented our work on the multi-level mixed-ADC receive architecture in mMIMO MWRNs. We derive a tight closed-form asymptotic approximation for the average achievable rate considering ZF relay beamforming and both perfect and imperfect CSI. We develope a novel analytical method using SVD for Gaussian matrices, distributions of the singular values of Gaussian matrices, and properties of Haar matrices. Our proposed method is general and can be applied to other beamforming schemes and similar scenarios. The result characterizes the achievable rate in terms of different system parameters and specifically quantifies the performance degradation caused by low-resolution ADCs and channel estimation error. The result indicates that the higher the average of ADC resolutions is, the more sensitive to the CSI quality the system is. Therefore, under channel estimation error, low-resolution ADCs are able to maintain the rate performance, while at the same time save significant hardware cost and energy. Also, it indicates that the average achievable rate has an almost linear relation with the square of the average of quantization coefficients of the ADC resolution profile. Further, in the medium to high SNR ranges the ADC resolution profile has a more significant effect on the rate performance compared to the number of antennas. Insightful asymptotic expressions are derived when the number of antennas grows towards infinity. Also, the special case of uniform-ADC is studied for comparison with the mixed-ADC case.

Chapter 5

Performance Analysis and Optimization of 3D Massive MIMO Multi-Pair Relaying with SWIPT

This chapter presents our contributions on mMIMO multi-user relay networks where the relay is equipped with a three-directional antenna array and SWIPT is performed by the users. In the first section, we briefly review the literature on SWIPT-enabled relay networks. Next, the system model is elaborated which is followed by performance analysis on the average achievable sum-rate. The performance analysis is performed considering two different beamforming methods of MRC/MRT and MF. An optimization problem over the relay antenna array tilt and SWIPT PS parameter to maximize the sum-rate is formulated and solved. Finally, the optimization results are compared for the two beamforming methods through simulations.

5.1 Introduction and Literature Review

Cooperative communication systems with mMIMO relay stations can provide high performance in spectral and energy efficiency, coverage and reliability [2, 14]. As mentioned in Chapter 1, an emerging solution for stretching the durability of energy constrained mMIMO relay networks is SWIPT, which is especially beneficial for applications in IoT and sensor networks [24]. There have been many works on MIMO and mMIMO relay networks with SWIPT [27, 115–120]. For a single-user one-way relay network, both PS and TS protocols are proposed in [27] for AF relay, where the outage probability and ergodic capacity are derived. In [115], for a similar network, the optimum splitting ratios are obtained to maximize the rate when DF is applied. For a multi-user mMIMO relay network where the relay harvests energy from both the source and destination users, the power scaling law is analyzed in [116] for

the MRC/MRT beamforming and DF. It is shown that the harvested energy is independent of the fast fading effect, and that the transmission power of each source and destination can be scaled inversely proportional to the number of relay antennas. For two-way relaying, the work in [117] is on a single-user AF relay network with ZF beamforming, where the sum-rate is analyzed and the relay PS ratio is optimized. For the same setup, but with a DF relay optimum PS and TS ratios are found in [118]. For the multi-user case, the PS ratios of energy harvesting users are optimized in [119] to maximize the achievable rates considering both ZF and MRC. For a multi-way relay network where the relay harvests energy, asymptotic sum-rate analysis is conducted in [120] under ZF.

On the other hand, as mentioned in Chapter 1, a planar antenna array can provide 3D beam pattern, which makes it possible to employ the elevation angle and accommodate massive number of antennas [18, 121]. Further, it is shown that 3D beamforming increases the focus of signal power and reduces the interference to other users [2]. In [122], for the first time, a planar sectorized antenna array is assumed for a SWIPT 3D ad-hoc network and the performance is analyzed. In SWIPT systems, due to the short ranges, the transmission-distance-to-base-station-height ratio is usually less than that of traditional macrocells [123]. Thus, the vertical domain beam design is as important as the horizontal one. Array tilt is the parameter for controlling the beam pattern in the vertical domain providing the degree-of-freedom to bring notable performance gains. In [123], aiming to minimize the BS transmit power, the BS antenna tilt and users' PS ratios are optimized for the downlink communications.

In this work, we study a SWIPT-enabled mMIMO 3D multi-user one-way relay network based on DF [124], where the relay harvests energy from the source users with a PS receiver. Two beamforming methods, MRC/MRT and MF, are considered at the relay and the performance of the corresponding networks are analyzed. Using the results for Haar matrices, we derive closed-form expressions that serve as lower-bounds on the average SINRs. Based on that, asymptotic average achievable sum-rate expressions are obtained. Further, joint optimization of the relay PS ratio and array tilt is formulated to maximize the average achievable sum-rates. Monte-Carlo simulation results are presented to verify the theoretical analysis and the gains that the opti-

mized setups bring. Simulations show that both MRC/MRT and MF schemes lead to a logarithmic relation between the achievable sum-rate and the number of antennas at the relay. Also, the results for the optimized settings show significant performance improvements compared to a conventional system with $\pi/4$ tilt and PS ratio of 0.5. Finally, the results for the two beamforming methods are compared and insightful conclusions are made. It is shown that in most cases, the MF scheme considerably excels the MRC/MRT scheme in terms of the achievable sum-rate.

5.2 System Model

There are K pairs of single-antenna users that are separated into two groups: source users u_{Sk}, and destination users u_{Dk}, for $k = 1, ..., K$, communicating through a common mMIMO one-way relay. u_{Sk} sends information to u_{Dk}. The number of antennas at the relay is denoted by N. In the multiple access phase, the source users perform SWIPT and the relay uses the PS scheme, meaning that part of the received signal at the relay is used to decode the source information, while the other part is used to harvest energy for the coming broadcast phase. In the broadcast phase, the relay uses the harvested energy and transmits the decoded information to the destination users. The relay either employs MRC/MRT or MF beamforming in the DF process.

A simplified presentation of the relay 3D antenna pattern is depicted in Figure 2.6. According to Section 2.4, the observed gain from the relay antenna array at the kth source or destination user, for $i \in \{S, D\}$, can be written as the following in dBi scale:

$$A_{ik}^{\text{dBi}}(\theta_{\text{tilt}}) = -\Big(\min\Big[12\Big(\frac{\phi_{ik}}{\phi_{\text{3dB}}}\Big)^2, \text{SLL}_{\text{az}}\Big] + \min\Big[12\Big(\frac{\theta_{ik} - \theta_{\text{tilt}}}{\theta_{\text{3dB}}}\Big)^2, \text{SLL}_{\text{el}}\Big]\Big), \tag{5.1}$$

where $0 < \theta_{\text{tilt}} < \pi/2$ is the tilt angle between the horizontal line and the antenna array beam, θ_{ik} is the angle between the horizon and the line connecting the user to the relay antenna array, ϕ_{ik} is the angle between the X-axis and the line in the horizontal plane that connects the user to the projection point of the relay on the horizontal plane. Further, the SLLs of the antenna array and the half beamwidth in the horizontal and vertical planes are set similar to

Section 2.4. It is assumed that the relay beam peak is fixed on $\phi = 0$ relative to the X-axis.

5.2.1 Channel Model

Due to the severe path loss, we assume that a direct link between each pair of source and destination does not exist. Let $\mathbf{H}_i = \tilde{\mathbf{H}}_i \mathbf{D}_i^{\frac{1}{2}}$ be the channel matrix from users to the relay for $i \in \{S, D\}$. $\tilde{\mathbf{H}}_i \in \mathbb{C}^{N \times K}$ represents the small-scale Rayleigh fading matrix with i.i.d. elements following $\mathcal{CN}(0, 1)$. Further, the $K \times K$ diagonal matrix $\mathbf{D}_i = \text{diag}\{\beta_{ik}(\theta_{\text{tilt}})\}$ for $k \in \{1, \cdots K\}$, $i \in \{S, D\}$, represents the large scale fading coefficients accounting for path loss and antenna gain. The kth element of $\mathbf{D}_i$ is given by,

$$\beta_{ik}(\theta_{\text{tilt}}) = d_{ik}^{-\nu} A_{ik}(\theta_{\text{tilt}}), \tag{5.2}$$

where d_{ik} denotes the distance between the kth user and the relay, and ν is the path loss exponent. The channel fading keeps invariant in each relaying time block, but changes independently from one block to another. It is assumed that the relay has perfect CSI[1].

5.3 Average Achievable Sum-rate Analysis

In this section, we derive the average achievable sum-rate of the multi-user one-way relay network with SWIPT considering MRC/MRT and MF beamforming schemes. First, the MAC phase of the network is explained and the harvested energy at the relay is derived. Then, the beamforming design, the corresponding BC phase, and the achievable sum-rate results are presented for MRC/MRT and MF relaying schemes, respectively. Discussions are made to compare the achievable rate results for the two beamforming schemes and to analyze the sum-rate performance in terms of different system parameters.

5.3.1 The MAC Phase and Harvested Energy

In the MAC phase, the source users transmit their information signals simultaneously to the relay. The source users perform SWIPT and the relay

[1]The analytical method can be extended to the imperfect CSI case.

both harvests the energy and decodes the information. Let x_k be the transmit signal of u_{Sk}, which has zero-mean and is normalized to have unit-power, i.e., $\mathbb{E}_{x_k \in \mathcal{S}_k}\{|x_k|^2\} = 1$, where $\mathcal{S}_k$ is the modulation set for u_{Sk}. Signals of different users are independent. The vector of transmit signal is denoted as $\mathbf{x} = [x_1 \ \cdots, x_K]^T$. The average transmit power of each source user is denoted by p_s.

The received signal at the relay can be written as

$$\mathbf{y}_{\mathrm{R}} = \sqrt{p_s}\mathbf{H}_S\mathbf{x} + \mathbf{n}_{\mathrm{R}}, \tag{5.3}$$

where $\mathbf{n}_{\mathrm{R}}$ is the AWGN at the relay with entries following i.i.d. $\mathcal{CN}(0, \sigma_{\mathrm{R}}^2)$ distribution. The relay uses PS for both energy harvesting and information reception, where the PS ratio is denoted by $\rho \in (0, 1)$. Thus, ρ-portion of the received signal power is for information decoding and the remaining $(1-\rho)$-portion is for energy harvesting. The signal for relay information decoding is

$$\begin{aligned} \mathbf{y}_{\mathrm{ID}} &= \sqrt{\rho}(\sqrt{p_s}\mathbf{H}_S\mathbf{x} + \mathbf{n}_{\mathrm{R}}) + \mathbf{n}'_{\mathrm{R}}, \\ &= \sqrt{\rho p_s}\mathbf{h}_{Sk}x_k + \sqrt{\rho p_s}\sum_{j\neq k}^{K}\mathbf{h}_{Sj}x_j + \tilde{\mathbf{n}}_{\mathrm{R}}, \end{aligned} \tag{5.4}$$

where $\mathbf{n}'_{\mathrm{R}}$ is the AWGN after the relay PS receiver whose entries are i.i.d. following $\mathcal{CN}(0, {\sigma'}_{\mathrm{R}}^2)$. Further, $\tilde{\mathbf{n}}_{\mathrm{R}}$ defined as $\tilde{\mathbf{n}}_{\mathrm{R}} = \sqrt{\rho}\mathbf{n}_{\mathrm{R}} + \mathbf{n}'_{\mathrm{R}}$ is the effective additive noise at the relay for signal reception. It can be shown via straightforward calculations that entries of $\tilde{\mathbf{n}}_{\mathrm{R}}$ are i.i.d. each following $\mathcal{CN}(0, \rho\sigma_{\mathrm{R}}^2 + {\sigma'}_{\mathrm{R}}^2)$. The energy harvesting portion of the relay received signal is $\sqrt{1-\rho}\mathbf{y}_{\mathrm{R}}$. Notice that $\mathbf{n}_{\mathrm{R}}$ cannot be harvested and the noise power at the relay EH receiver is assumed to be small and thus ignored. So, the EH portion of the received signal at the relay is approximated as

$$\sqrt{(1-\rho)p_s}\mathbf{H}_S\mathbf{x}. \tag{5.5}$$

A closed-form expression for the harvested energy at the relay is provided in the following theorem.

Theorem 5. *The harvested energy per unit time at the relay is*

$$E_{\mathrm{H}} = \eta(1-\rho)p_s N \sum_{i=1}^{K} \beta_{Si}(\theta_{\mathrm{tilt}}), \tag{5.6}$$

where η is the RF-to-direct current (DC) conversion efficiency.

Proof: Please see Appendix B.1.

In the above theorem, for the sake of simplicity, we have used the linear energy harvesting model [26], which models the harvested DC power as a linear function of the incident RF power. Recently, in some other works, in order to capture the non-linearity of energy harvester circuit elements, such as diodes, nonlinear models are proposed [125].

When the relay receives the source users' information, it performs beamforming on the received signal, and then, adopts the DF protocol. So, for the rest of the transceiver protocol, considering MRC/MRT, and then, MF beamforming method, we present the DF process. Further, for each beamforming scheme we derive the average achievable sum-rate.

5.3.2 MRC/MRT Beamforming and Sum-Rate Results

The MRC beamforming matrix for the relay is designed as $\mathbf{A}_S = \mathbf{H}_S^H$. The signal vector at the relay after MRC is denoted by $\mathbf{r}_{\mathrm{R}}^{\mathrm{MRC}}$, where the kth entry pertains to the information of the kth user and can be written as the following,

$$r_{\mathrm{R},k}^{\mathrm{MRC}} = \sqrt{\rho p_s}\|\mathbf{h}_{Sk}\|^2 x_k + \sqrt{\rho p_s}\mathbf{h}_{Sk}^H \sum_{j\neq k}^{K} \mathbf{h}_{Sj} x_j + \mathbf{h}_{Sk}^H \tilde{\mathbf{n}}_{\mathrm{R}}. \tag{5.7}$$

For the achievable sum-rate analysis, we use the method in [126] for mMIMO systems to find a lower bound, which is shown to be tight when N approaches infinity. The key is to write the channel vector norm in the received signal, as the summation of a known mean and the difference part. Hence,

$$\begin{aligned} r_{\mathrm{R},k}^{\mathrm{MRC}} = {} & \sqrt{\rho p_s}\mathbb{E}\{\|\mathbf{h}_{Sk}\|^2\}x_k + \sqrt{\rho p_s}(\|\mathbf{h}_{Sk}\|^2 - \mathbb{E}\{\|\mathbf{h}_{Sk}\|^2\})x_k \\ & + \sqrt{\rho p_s}\sum_{j\neq k}^{K} \mathbf{h}_{Sk}^H \mathbf{h}_{Sj} x_j + \mathbf{h}_{Sk}^H \tilde{\mathbf{n}}_{\mathrm{R}}. \end{aligned} \tag{5.8}$$

By treating the second term as noise, the average SINR of u_{Sk} at the relay in the MAC phase, denoted by $\gamma_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}}$, has the following lower bound

$$\gamma_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}} \geq \bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}} = \frac{N\beta_{Sk}(\theta_{\mathrm{tilt}})}{\beta_{Sk}(\theta_{\mathrm{tilt}}) + \sum_{j\neq k}^{K}\beta_{Sj}(\theta_{\mathrm{tilt}}) + \frac{\rho\sigma_{\mathrm{R}}^2+\sigma'^2_{\mathrm{R}}}{\rho p_s}}. \tag{5.9}$$

The calculations of $\bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}}$ are provided in Appendix B.2. By following the analysis in [126], it can also be shown that $\gamma_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}} \xrightarrow[N\to\infty]{\text{a.s.}} \bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}}$, where $\xrightarrow[N\to\infty]{\text{a.s.}}$ means almost sure convergence when $N \to \infty$. Further, when $N \to \infty$, the normalized effective noise term in (5.8), the sum of the 2nd, 3rd and 4th terms, converges in distribution to a Gaussian random variable. Thus, for the average achievable rate of u_{Sk} at the relay, we have

$$R_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}} = \log_2(1+\gamma_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}}) \xrightarrow[N\to\infty]{\text{a.s.}} \bar{R}_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}} = \log_2(1+\bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}}). \tag{5.10}$$

After the relay beamforming, the relay decodes $\mathbf{x}$ from (5.7), then sends the decoded information in the BC phase. In this phase, the relay uses MRT on the decoded information signals to design its transmit signal vector as the following

$$\mathbf{x}_{\mathrm{R}}^{\mathrm{MRT}} = \sqrt{P_{\mathrm{R}}}\alpha^{\mathrm{MRT}}\mathbf{H}_D^*\mathbf{x}, \tag{5.11}$$

where α^{MRT} is the power coefficient and $P_R = E_{\mathrm{H}}$ is the relay transmission power as in (5.6). The power at the relay is limited according to $\mathbb{E}\{|\mathbf{x}_{\mathrm{R}}^{\mathrm{MRT}}|^2\} = E_{\mathrm{H}}$, and we have

$$\alpha^{\mathrm{MRT}} = \frac{1}{\sqrt{\mathbb{E}\left\{\mathrm{tr}\left\{\mathbf{H}_D^T\mathbf{H}_D^*\right\}\right\}}} = \frac{1}{\sqrt{N\sum_{i=1}^{K}\beta_{Di}(\theta_{\mathrm{tilt}})}}. \tag{5.12}$$

By noticing that the channel matrix from the relay to the destination users is $\mathbf{H}_D^T$, the received signal at the kth destination user is

$$r_{\mathrm{D},k}^{\mathrm{MRT}} = \sqrt{P_{\mathrm{R}}}\alpha^{\mathrm{MRT}}\|\mathbf{h}_{Dk}\|^2 x_k + \sqrt{P_{\mathrm{R}}}\alpha^{\mathrm{MRT}}\mathbf{h}_{Dk}^*\sum_{j\neq k}^{K}\mathbf{h}_{Dj}^T x_j + n_{\mathrm{D},k}, \tag{5.13}$$

where $n_{\mathrm{D},k}$ is the AWGN at the kth destination user with $\mathcal{CN}(0,\sigma_{\mathrm{D}}^2)$ distribution. By following the same steps as in (5.8) and Appendix B.2, a lower-bound expression for the average SINR at u_{Dk} in the BC phase is as the following

$$\begin{aligned}\gamma_{k,\mathrm{BC}}^{\mathrm{MRC/MRT}} &\geq \bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MRC/MRT}},\\ \bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MRC/MRT}} &= \frac{N\beta_{Dk}(\theta_{\mathrm{tilt}})}{\beta_{Dk}(\theta_{\mathrm{tilt}})+\sum_{j\neq k}^{K}\beta_{Dj}(\theta_{\mathrm{tilt}})+\frac{\sigma_{\mathrm{D}}^2\sum_{i=1}^{K}\beta_{Di}(\theta_{\mathrm{tilt}})}{\eta(1-\rho)p_s N\beta_{Dk}(\theta_{\mathrm{tilt}})\sum_{i=1}^{K}\beta_{Si}(\theta_{\mathrm{tilt}})}},\end{aligned} \tag{5.14}$$

where $\gamma_{k,\mathrm{BC}}^{\mathrm{MRC/MRT}} \xrightarrow[N\to\infty]{\text{a.s.}} \bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MRC/MRT}}$.

Then, for the average achievable rate from the relay to u_{Dk}, we have

$$R_{k,\mathrm{BC}}^{\mathrm{MRC/MRT}} = \log_2(1+\gamma_{k,\mathrm{BC}}^{\mathrm{MRC/MRT}}) \xrightarrow[N\to\infty]{\text{a.s.}} \bar{R}_{k,\mathrm{BC}}^{\mathrm{MRC/MRT}} = \log_2(1+\bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MRC/MRT}}). \tag{5.15}$$

The average achievable rate of the kth source and destination pair is the smaller of the achievable rates of the two phases given as

$$\bar{R}_{k}^{\mathrm{MRC/MRT}} = \min\{\bar{R}_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}}, \bar{R}_{k,\mathrm{BC}}^{\mathrm{MRC/MRT}}\}. \tag{5.16}$$

Finally, the average achievable sum-rate of all user pairs in the relay network is defined as

$$\bar{R}_{\mathrm{sum}}^{\mathrm{MRC/MRT}} = \sum_{i=1}^{K}\bar{R}_{k}^{\mathrm{MRC/MRT}}. \tag{5.17}$$

5.3.3 MF Beamforming and Sum-Rate Results

In the following, we consider MF beamforming which is among the beamforming schemes with the lowest complexity. For the MAC phase, the MF reception beamforming matrix is designed as

$$\mathbf{B}_{\mathrm{S}} = \left[\begin{array}{ccc} \frac{\mathbf{h}_{S1}}{\|\mathbf{h}_{S1}\|} & \cdots & \frac{\mathbf{h}_{SK}}{\|\mathbf{h}_{SK}\|} \end{array}\right]^{H}, \tag{5.18}$$

From (5.3) and (5.4), the processed received vector at the relay can be written as

$$\begin{aligned}\mathbf{r}_{\mathrm{R}}^{\mathrm{MF}} &= \mathbf{B}_{\mathrm{S}}\mathbf{y}_{\mathrm{ID}} \\ &= \sqrt{\rho p_s}\mathbf{B}_{\mathrm{S}}\mathbf{H}_S\mathbf{x} + \mathbf{B}_{\mathrm{S}}\tilde{\mathbf{n}}_{\mathrm{R}}.\end{aligned} \tag{5.19}$$

The kth entry of $\mathbf{r}_{\mathrm{R}}^{\mathrm{MF}}$ containing the information of the kth source user can be written as the following,

$$r_{k,\mathrm{R}}^{\mathrm{MF}} = \sqrt{\rho p_s}\|\mathbf{h}_{Sk}\|x_k + \sqrt{\rho p_s}\frac{\mathbf{h}_{Sk}^H}{\|\mathbf{h}_{Sk}\|}\sum_{j\neq k}^{K}\mathbf{h}_{Sj}x_j + \frac{\mathbf{h}_{Sk}^H}{\|\mathbf{h}_{Sk}\|}\tilde{\mathbf{n}}_{\mathrm{R}}.$$

Again, for the achievable sum-rate analysis, we use the method for mMIMO systems in [126] to find a lower bound. Thus, we write the channel vector norm in the received signal in terms of the summation of a known mean and the uncorrelated difference part as follows

$$\begin{aligned}r_{k,\mathrm{R}}^{\mathrm{MF}} &= \sqrt{\rho p_s}\mathbb{E}\{\|\mathbf{h}_{Sk}\|\}x_k + \sqrt{\rho p_s}(\|\mathbf{h}_{Sk}\| - \mathbb{E}\{\|\mathbf{h}_{Sk}\|\})x_k \\ &\quad + \sqrt{\rho p_s}\frac{\mathbf{h}_{Sk}^H}{\|\mathbf{h}_{Sk}\|}\sum_{j\neq k}^{K}\mathbf{h}_{Sj}x_j + \frac{\mathbf{h}_{Sk}^H}{\|\mathbf{h}_{Sk}\|}\tilde{\mathbf{n}}_{\mathrm{R}}.\end{aligned} \tag{5.20}$$

By treating the second term as noise, the average SINR of u_{Sk} at the relay denoted by γ_k^{MAC} has the following lower bound

$$\gamma_{k,\mathrm{MAC}}^{\mathrm{MF}} \geq \bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MF}} = \frac{\beta_{Sk}(\theta_{\mathrm{tilt}})\left(\frac{\Gamma(N+\frac{1}{2})}{\Gamma(N)}\right)^2}{\beta_{Sk}(\theta_{\mathrm{tilt}})\left(N - \left(\frac{\Gamma(N+\frac{1}{2})}{\Gamma(N)}\right)^2\right) + \sum_{j\neq k}^{K}\beta_{Sj}(\theta_{\mathrm{tilt}}) + \frac{\rho\sigma_{\mathrm{R}}^2 + \sigma'^2_{\mathrm{R}}}{\rho p_s}}. \tag{5.21}$$

Please refer to Appendix B.3 for the derivation of $\bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MF}}$. By following the analysis in [126], it can also be shown that $\gamma_{k,\mathrm{MAC}}^{\mathrm{MF}} \xrightarrow[N\to\infty]{\text{a.s.}} \bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MF}}$. Further, similar to the derivations for the MRC/MRT scheme, when $N \to \infty$, the normalized effective noise term in (5.20), the sum of the 2nd, 3rd and 4th terms, converges in distribution to a Gaussian random variable. Thus, the

average achievable rate of u_{Sk} at the relay has the following result

$$R_{k,\mathrm{MAC}}^{\mathrm{MF}} = \log_2(1+\gamma_{k,\mathrm{MAC}}^{\mathrm{MF}}) \xrightarrow[N\to\infty]{\text{a.s.}} \bar{R}_{k,\mathrm{MAC}}^{\mathrm{MF}} = \log_2(1+\bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MF}}). \tag{5.22}$$

For the BC phase, the relay uses the MF beamforming to precode the decoded source signals. The MF transmit beamforming matrix is designed as

$$\mathbf{B}_{\mathrm{D}} = \left[\frac{\mathbf{h}_{D_1}^*}{\|\mathbf{h}_{D_1}\|} \cdots \frac{\mathbf{h}_{D_K}^*}{\|\mathbf{h}_{D_K}\|} \right]. \tag{5.23}$$

The transmit signal vector of the relay is thus,

$$\begin{aligned} \mathbf{x}_{\mathrm{R}}^{\mathrm{MF}} &= \sqrt{P_{\mathrm{R}}}\alpha^{\mathrm{MF}}\mathbf{B}_{\mathrm{D}}\mathbf{x}, \\ &= \sqrt{P_{\mathrm{R}}}\alpha^{\mathrm{MF}} \sum_{k=1}^{K} \frac{\mathbf{h}_{Dk}^*}{\|\mathbf{h}_{Dk}\|} x_k. \end{aligned} \tag{5.24}$$

The relay transmission power is $P_R = E_{\mathrm{H}}$, which is given in (5.6) and α^{MF} is the MF power coefficient. Due to the power constraint at the relay, $\mathbb{E}\{|\mathbf{x}_{\mathrm{R}}^{\mathrm{MF}}|^2\} = E_{\mathrm{H}}$, we have

$$\alpha^{\mathrm{MF}} = \frac{1}{\sqrt{\sum_{k=1}^{K} \mathbb{E}\left\{ \frac{\mathbf{h}_{Dk}^T \mathbf{h}_{Dk}^*}{\|\mathbf{h}_{Dk}\|^2} \right\}}} = \frac{1}{\sqrt{K}}. \tag{5.25}$$

The received signal at the kth destination user will be

$$r_{\mathrm{D},k}^{\mathrm{MF}} = \sqrt{P_{\mathrm{R}}}\alpha^{\mathrm{MF}}\|\mathbf{h}_{Dk}\|x_k + \sqrt{P_{\mathrm{R}}}\alpha^{\mathrm{MF}} \frac{\mathbf{h}_{Dk}^*}{\|\mathbf{h}_{Dk}\|} \sum_{j\neq k}^{K} \mathbf{h}_{Dj}^T x_j + n_{\mathrm{D},k}. \tag{5.26}$$

To obtain achievable rate of the BC phase, the received signal at u_{Dk} can be written as

$$\begin{aligned} r_{\mathrm{D},k}^{\mathrm{MF}} = {} & \sqrt{P_R}\alpha^{\mathrm{MF}}\mathbb{E}\{\|\mathbf{h}_{Dk}\|\}x_k + \sqrt{P_R}\alpha^{\mathrm{MF}}(\|\mathbf{h}_{Dk}\| - \mathbb{E}\{\|\mathbf{h}_{Dk}\|\})x_k \\ & + \sqrt{P_R}\alpha^{\mathrm{MF}} \frac{\mathbf{h}_{Dk}^*}{\|\mathbf{h}_{Dk}\|} \sum_{j\neq k}^{K} \mathbf{h}_{Dj}^T x_j + n_{\mathrm{D},k}. \end{aligned} \tag{5.27}$$

So, a lower-bound expression for the average SINR at u_{Dk} in the BC phase,

denoted by $\gamma_{k,\mathrm{BC}}^{\mathrm{MF}}$, will be

$$\begin{aligned}
\gamma_{k,\mathrm{BC}}^{\mathrm{MF}} &\geq \bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MF}}, \\
\bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MF}} &= \frac{|\mathbb{E}\{\|\mathbf{h}_{Dk}\|\}|^2}{\mathrm{Var}\left\{\|\mathbf{h}_{Dk}\|\right\} + \sum_{j\neq k}^{K} \mathbb{E}\left\{\left|\frac{\mathbf{h}_{Dk}^{*}\mathbf{h}_{Dj}^{T}}{\|\mathbf{h}_{Dk}\|}\right|^2\right\} + \frac{\sigma_{\mathrm{D}}^2}{\alpha^2 E_H}} \\
&= \frac{\beta_{Dk}(\theta_{\mathrm{tilt}})\left(\frac{\Gamma(N+\frac{1}{2})}{\Gamma(N)}\right)^2}{\beta_{Dk}(\theta_{\mathrm{tilt}})\left(N - \left(\frac{\Gamma(N+\frac{1}{2})}{\Gamma(N)}\right)^2\right) + \sum_{j\neq k}^{K}\beta_{Dj}(\theta_{\mathrm{tilt}}) + \frac{\sigma_{\mathrm{D}}^2 K}{\eta(1-\rho)p_s N \sum_{i=1}^{K}\beta_{Si}(\theta_{\mathrm{tilt}})}},
\end{aligned} \tag{5.28}$$

where $\gamma_{k,\mathrm{BC}}^{\mathrm{MF}} \xrightarrow[N\to\infty]{\text{a.s.}} \bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MF}}$. Notice that $\mathbf{h}_{Sk}$ and $\mathbf{h}_{Dk}$ have the same distribution, thus, by applying (5.6), (5.25) and the equations (B.7)-(B.9) from Appendix B.3, the value of $\bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MF}}$ is obtained as (5.28). Then, for the average achievable rate from the relay to u_{Dk}, we have

$$R_{k,\mathrm{BC}}^{\mathrm{MF}} = \log_2(1+\gamma_{k,\mathrm{BC}}^{\mathrm{MF}}) \xrightarrow[N\to\infty]{\text{a.s.}} \bar{R}_{k,\mathrm{BC}}^{\mathrm{MF}} = \log_2(1+\bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MF}}). \tag{5.29}$$

The average achievable rate of the kth source and destination pair will be

$$\bar{R}_k^{\mathrm{MF}} = \min\{\bar{R}_{k,\mathrm{MAC}}^{\mathrm{MF}}, \bar{R}_{k,\mathrm{BC}}^{\mathrm{MF}}\}. \tag{5.30}$$

Finally, the average achievable sum-rate of all user pairs is

$$\bar{R}_{\mathrm{sum}}^{\mathrm{MF}} = \sum_{i=1}^{K} \bar{R}_k^{\mathrm{MF}}. \tag{5.31}$$

5.3.4 Discussions

Comparing the beamforming matrices for the MRC/MRT scheme, which are designed as $\mathbf{H}_{\mathrm{S}}^{H}$ and $\mathbf{H}_{\mathrm{D}}^{*}$, with the beamforming matrices for the MF scheme given in (5.18) and (5.23), we can see that the MF scheme has a similar structure to MRC/MRT but with row-normalization. Note that every row of the MF beamforming matrices has unit-norm. While for the MRC/MRT scheme, users with strong channel gains (larger channel vector norms) are given higher weights or more transmit power; for the MF scheme, the same weight or power

is allocated to all users. This leads to difference in the achievable rate derivations. Further, by comparing the power coefficient for MF in (5.25) with its counterpart for the MRC/MRT scheme in (5.12), we see that the normalization results in a simpler power coefficient. Our analysis shows that the MF scheme achieves higher average achievable sum-rates than the MRC/MRT scheme.

From the results in (5.9), (5.14), (5.21), and (5.28), it can be seen that the SINRs for both MRC/MRT and MF schemes are increasing functions of the source users' average transmit power, p_s. Further, we can see that higher σ_{R}^2 and ${\sigma'_{\mathrm{R}}}^2$ (noise powers at the relay signal receiver and after the power splitting receiver of the relay) decreases the SINR of the MAC phase, and higher σ_{D}^2 (noise power at the destination user) decreases the SINR of the BC phase. Thus, naturally, an increase in the noise power degrades the system rate performance.

With respect to the number of relay antennas, N, we have

$$\lim_{N\to\infty} \frac{\bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}}}{N} = \frac{\beta_{Sk}(\theta_{\mathrm{tilt}})}{\beta_{Sk}(\theta_{\mathrm{tilt}}) + \sum_{j\neq k}^{K} \beta_{Sj}(\theta_{\mathrm{tilt}}) + \frac{\rho\sigma_{\mathrm{R}}^2 + {\sigma'_{\mathrm{R}}}^2}{\rho p_s}}. \tag{5.32}$$

Similarly,

$$\lim_{N\to\infty} \frac{\bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MRC/MRT}}}{N} = \frac{\beta_{Dk}(\theta_{\mathrm{tilt}})}{\beta_{Dk}(\theta_{\mathrm{tilt}}) + \sum_{j\neq k}^{K} \beta_{Dj}(\theta_{\mathrm{tilt}}) + \frac{\sigma_{\mathrm{D}}^2 \sum_{i=1}^{K} \beta_{Di}(\theta_{\mathrm{tilt}})}{\eta(1-\rho)p_s N \beta_{Dk}(\theta_{\mathrm{tilt}}) \sum_{i=1}^{K} \beta_{Si}(\theta_{\mathrm{tilt}})}}. \tag{5.33}$$

Further, for the MF beamforming scheme, it can be shown that

$$\lim_{N\to\infty} \left[N - \left(\frac{\Gamma(N+\frac{1}{2})}{\Gamma(N)} \right)^2 \right] = \frac{1}{4}.$$

Hence, we have

$$\lim_{N\to\infty} \frac{\bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MF}}}{N} = \frac{\beta_{Sk}(\theta_{\mathrm{tilt}})}{\frac{1}{4}\beta_{Sk}(\theta_{\mathrm{tilt}}) + \sum_{j\neq k}^{K} \beta_{Sj}(\theta_{\mathrm{tilt}}) + \frac{\rho\sigma_{\mathrm{R}}^2 + {\sigma'_{\mathrm{R}}}^2}{\rho p_s}}. \tag{5.34}$$

Similarly,

$$\lim_{N\to\infty} \frac{\bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MF}}}{N} = \frac{\beta_{Dk}(\theta_{\mathrm{tilt}})}{\frac{1}{4}\beta_{Dk}(\theta_{\mathrm{tilt}}) + \sum_{j\neq k}^{K} \beta_{Dj}(\theta_{\mathrm{tilt}}) + \frac{\sigma_{\mathrm{D}}^2 K}{\eta(1-\rho)p_s N \sum_{i=1}^{K} \beta_{Si}(\theta_{\mathrm{tilt}})}}. \tag{5.35}$$

Therefore, for large N, all $\bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}}$, $\bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MRC/MRT}}$, $\bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MF}}$, and $\bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MF}}$, increase linearly with N, leading to a logarithmic increase in the sum-rate with respect to the number of relay antennas. Comparing (5.34) with its counterpart for the MRC/MRT scheme in (5.32), we can see that the MF design has a smaller variance, 1/4, of the equivalent channel, leading to a higher rate for the MAC phase. Similar phenomenon can be observed for the BC phase by comparing (5.33) and (5.35); however, the comparison of the noise terms containing σ_{D}^2 is not straightforward in this case.

Higher power splitting ratio, ρ, increases $\bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MF}}$ and $\bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}}$, thus, improves the achievable rate of the MAC phase. But, it decreases $\bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MF}}$ and $\bar{\gamma}_{k,\mathrm{BC}}^{\mathrm{MRC/MRT}}$, and thus, degrades the achievable rate of the BC phase. The tilt of the relay antenna array, θ_{tilt}, affects the sum-rate through $\beta_{Sk}(\theta_{\mathrm{tilt}})$ and $\beta_{Dk}(\theta_{\mathrm{tilt}})$.

To improve SWIPT efficiency, joint optimization of the tilt and PS ratio can be conducted to maximize the average achievable sum-rate of the network, $\bar{R}_{\mathrm{sum}}$. The optimization can be formulated as

$$\begin{aligned} &\max_{\theta_{\mathrm{tilt}},\rho} \ \bar{R}_{\mathrm{sum}} \\ &\text{s.t.} \ \ 0 < \rho < 1, \\ &\text{and} \ \ 0 < \theta_{\mathrm{tilt}} < \pi/2. \end{aligned} \tag{5.36}$$

For either MF or MRC/MRT beamforming scheme, the objective function is highly non-convex. To find a closed-form expression for the optimum ρ and θ_{tilt} seems far-fetched given the complicated relations of the average achievable rates of the MAC and BC phases. Also, notice that both θ_{tilt} and ρ have small ranges of values. Hence, it is feasible to use a 2-dimensional grid search to find the optimal values for ρ and θ_{tilt}.

5.4 Simulation Results

In this section, simulated average achievable sum-rate results are shown to verify the derived analytical results. Also, the achievable sum-rate results for the MRC/MRT scheme are compared to the results for the MF scheme. The parameters of the relay antenna array gain in (5.1) are defined previously in Section 2.4. For other system parameters, we set $\eta = 0.5$, $\sigma_{\mathrm{R}}^2 = -80$ dB, $\sigma_{\mathrm{R}}'^2 = -70$ dB, $\sigma_{\mathrm{D}}^2 = -50$ dB, and $\nu = 3.76$. The height of the relay station is set as 4m and the users' locations are generated via the uniform distribution on the surface of a 120° sector of a circle with 6m radius. Two tilt and power splitting settings are considered. The first is a typical choice of $\rho = 0.5$, $\theta_{\mathrm{tilt}} = \pi/4$, which are the middle points of the possible ranges for ρ and θ_{tilt}, respectively. The second is the optimal solution of (5.36), denoted as ρ^* and θ_{tilt}^*. Each sum-rate value presented in this section is obtained by averaging over the sum-rates for 100 randomly generated location sets for the K user pairs. For the optimal design, optimization is performed over the theoretical sum-rate results in (5.17) and (5.31) for each location set. The search step sizes for θ_{tilt} and ρ are 0.1047 rad and 0.0667, respectively. The theoretical results are then verified by the Monte-Carlo simulations where for each random location set 10^3 channel realizations are generated.

Figure 5.1 and 5.2 show that the theoretical results perfectly match the Monte-Carlo simulations for all user power and relay antenna number ranges. In Figure 5.1, the average sum-rate results are shown when N changes from 45 to 170, $p_s = 15$ dB, and $K = 5, 7$. We can see that the average sum-rate is an increasing function of N with a logarithmic relation. The use of the optimal ρ^* and θ_{tilt}^* brings significant boost in the sum-rate over the typical choice. For example, considering MRC/MRT scheme, the proposed optimum system can bring the average sum-rate improvements of at least 67.04% for $K = 5$, and 61.81% for $K = 7$, both of which happen at $N = 170$. Also, considering MF scheme, at least 70.10% sum-rate increase is observed when $K = 5$, and 59.07% when $K = 7$, which also happens at $N = 170$. The gap between the optimal setting and the typical one grows as N increases. Further, compared to the MRC/MRT scheme, almost always the MF scheme achieves higher average sum-rates, and the advantage increases for larger N or smaller K.

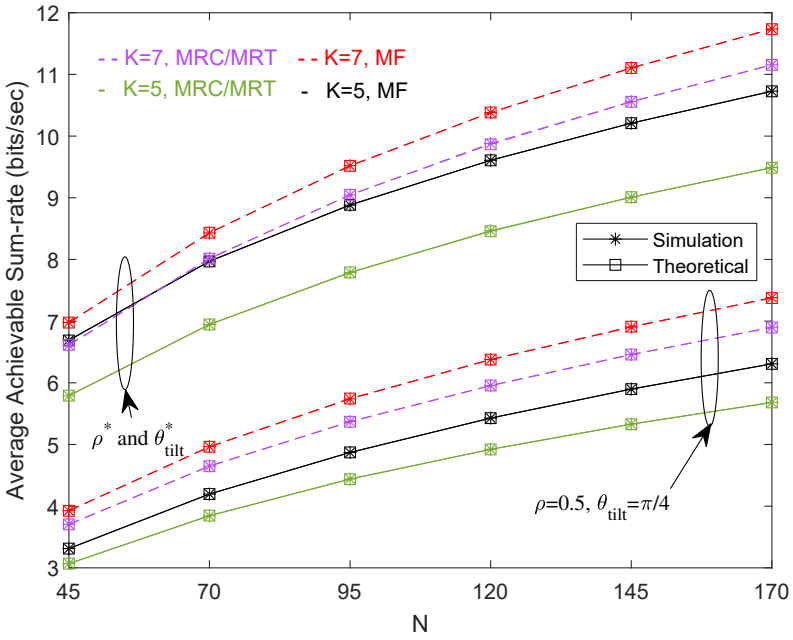


Figure 5.1: Theoretical and simulated average sum-rate results versus the number of relay antennas when $p_s = 15$ dB, and $K = 5, 7$. Two beamforming schemes, MF and MRC/MRT, are presented for both the optimal and typical settings of the antenna array tilt and PS ratio.

Figure 5.2 shows the average sum-rate results when p_s, the average transmit power of the source users, changes form 0 to 25 dB, $N = 100$, and $K = 5, 7$. It can be seen that the sum-rate is an increasing function of p_s with a negative acceleration. The use of the optimal ρ^* and θ^*_{tilt} brings significant sum-rate improvements of at least 72.95% for $K = 5$, and 68.97% for $K = 7$, considering MRC/MRT scheme, and 63.88% for $K = 5$, and 62.43% for $K = 7$, considering MF scheme, all of which happens when $p_s = 25$ dB. The advantage of the optimal setting slightly decreases with p_s. Again, the MF scheme shows performance advantage over the MRC/MRT scheme except for the typical setting with low SNR when $K = 5$. For the setting with the optimal ρ^* and θ^*_{tilt}, the advantage of MF over MRC/MRT decreases with p_s, while for the typical setting, the opposite is observed. Finally, this figure indicates that a self-sufficient energy harvesting relay with 100 antennas can provide an average sum-rate of 6.5 bits/sec considering MRC/MRT and 7 bits/sec considering MF scheme, for 5 pairs of users with only 1 Watt average source users' power.

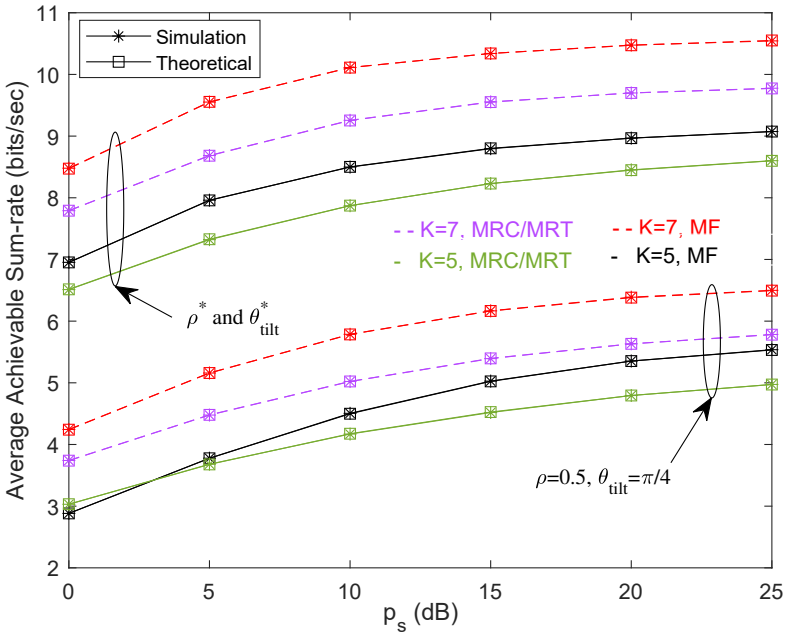


Figure 5.2: Theoretical and simulated average sum-rate results versus average user transmit power when $N = 100$, and $K = 5, 7$. Two beamforming schemes, MF and MRC/MRT, are presented for both the optimal and typical settings of the antenna array tilt and PS ratio.

5.5 Conclusion

In this chapter, we have analyzed the average achievable sum-rate performance of a SWIPT-enabled multi-user mMIMO one-way relay network with a planar antenna array. The planar antenna array makes 3D beamforming possible, which provides significant performance gains if properly designed. The relay harvests energy from the source users using a PS receiver. We provide the analysis considering both MRC/MRT and MF transceive beamforming at the relay. By using the results for Haar matrix and chi-square random vectors, closed-form lower-bound expressions for the average achievable sum-rates are derived, which confirm a logarithmic relation between the achievable sum-rate and the number of antennas at the relay for both MRC/MRT and MF schemes. Based on the derivations, a joint optimization over the PS ratio and the relay antenna array tilt is conducted for the highest sum-rate performance. Simulation results have validated the tightness of the analytical results. Due to the extra degrees-of-freedom offered by the vertical tilt and SWIPT, significant

performance improvement has been observed by using the optimal PS ratio and antenna array tilt. Further, it is shown that the MF beamforming at the relay results in higher sum-rates compared to the MRC/MRT scheme.

Chapter 6

Conclusions and Future Work

In this chapter, the contributions of this dissertation are summarized. Then, new problems are described for future research directions.

6.1 Conclusions

Modern relays to be employed in 5G wireless networks are envisioned to be able to significantly improve the data rate and coverage, while reducing the energy consumption and cost. Thus, there are various new challenges and ideas involved to make modern relays adaptable to the 5G industry. In this thesis, we present three projects related to relay communications networks. We provide a novel beamforming method for multi-way relaying which significantly improves the achievable data rate compared to the state-of-the-art. We develop a novel mathematical solution to derive a closed-form expression for the average achievable rate of the complicated multi-way relay network with low-resolution mixed-level ADCs. Finally, we study 3D beamforming accompanied by SWIPT in a mMIMO one-way relay network. The developed results and designs help us better understand the performance of relay networks with respect to different system parameters, and enable us to efficiently use the extra degrees-of-freedom available through the new technologies for 5G systems. Specific descriptions of the contributions for each project are as follows.

In Chapter 3, we propose a novel beamforming design, PZF, for MWRNs with a MIMO relay. Compared to the conventional ZF scheme, PZF forces only part of the user-interference at the relay to zero, instead of all. Thus, it provides extra degrees-of-freedom in the beamforming matrix for rate performance optimization. The users eliminate the remaining interference through self-interference and successive interference cancellation. A sum-rate maximization problem is formulated to exploit the extra degrees-of-freedom resulted from PZF. Further, modified gradient-ascent algorithms are proposed to solve the

optimization. Simulation results show that the proposed PZF relay beamforming design achieves significantly better network sum-rates than the existing linear beamforming designs.

In Chapter 4, we consider a general mMIMO MWRN with a multi-level mixed-ADC architecture in which each antenna is connected to an ADC pair with an arbitrary resolution. By leveraging on Bussgang's decomposition theorem and Lloyd-Max algorithm for quantization, tight closed-form asymptotic approximations are derived for the average achievable rates considering ZF relaying scheme and both perfect and imperfect CSI. To handle such a challenging setup, we develop a novel method for the achievable rate analysis using distributions of the singular values of Gaussian matrices and properties of Haar matrix. We demonstrate that the average achievable rate has a close-to-linear relation with the square of the average of quantization coefficients pertaining to the ADC resolution profile. In addition, in the medium to high SNR region, the ADC resolutions have a more significant effect on the rate compared to the number of antennas. Our work also reveals that the performance gap between the perfect and imperfect CSI cases is smaller for lower ADC resolutions.

In Chapter 5, we investigate a SWIPT-enabled mMIMO multi-user one-relay network where the relay is equipped with a planar antenna array making 3D beamforming possible. SWIPT is applied by the users and the relay harvests energy based on PS protocol. Two relay beamforming schemes, MRC/MRT and MF, are considered for the analysis and comparison. By using the results for Haar matrix and chi-square random vectors, closed-form lower-bounds are derived for the average achievable sum-rates of the network. The derived theoretical results which tightly match the simulations, show a logarithmic relation between the achievable sum-rate and the relay antenna number for both MRC/MRT and MF. Joint optimization of the relay antenna array tilt and PS ratio is conducted to maximize the sum-rate. It is shown that the tilt and power splitting ratio optimization can bring significant performance improvements. Further, our analyses verify that MF scheme results in a better sum-rate performance than the MRC/MRT scheme.

6.2 Future Work

In the following, we outline some possible future research directions regarding each of our contributions.

In Chapter 3, we assumed perfect CSI and proposed the PZF scheme, while in practice the channel estimation cannot be perfect. The estimation error will cause different levels of CSI error among different channels between the relay and users. Therefore, along with heterogeneous channels in which decoding order may affect the system performance, imperfect CSI may also need further attention in relaying design and user decoding. One possible extension of this project is to further analyze the effect of channel estimation error on the performance of PZF beamforming in MWRNs. Another research direction is to investigate the effect of joint optimization of signal processing at the users and PZF beamforming at the relay on the rate performance. This design may be able to bring a higher rate performance; however, it also leads to a higher computational complexity. Therefore, the complexity-rate trade-off should be investigated. There are other designs available in the literature on the joint relay beamforming and signal processing at the user with which the complexity-rate trade-off of our design can be compared with.

A possible research direction in continuation of our contributions in Chapter 4 is to extend the performance analysis to other beamforming schemes and make a comparison between their performances. As explained in Chapter 4, the mixed-ADC (non-uniform) resolution profile causes the ZF beamforming not to fully force the interference to zero. Therefore, chances are that other well-designed linear beamforming schemes may excel ZF in such a network. Our proposed analytical method makes the analysis of other channel-dependent linear beamforming schemes possible. Another idea in this regard is to design a novel relay beamforming scheme based on the knowledge of mixed-ADC profile to improve the interference cancellation compared to the traditional ZF. This beamforming design may excel the ZF design, but its performance still needs to be compared with other beamforming schemes to find the best one in terms of performance and complexity of design. Another interesting extension to Chapter 4 is the optimization of the mixed-ADC resolution profile for mMIMO MWRNs in order to minimize the relay energy consumption considering certain

common rate requirements for each user, or maximizing the minimum common rate considering certain limitations on the relay energy consumption.

One interesting extension of our work in Chapter 5 is to find the limitation on the cell size given certain rate performance requirements for the users. As it is mentioned in Chapter 5, the relay-to-users distance plays an important role in the efficiency of SWIPT, thus, one should set cell-edge QoS requirements while employing such a network with SWIPT. Further, this work can be extended to two-way and MWRNs where performance analysis should be conducted to see how much improvement can the extra degrees-of-freedom from SWIPT and 3D beamforming bring. For MWRNs in particular, it is possible that the harvested energy at the relay in the MAC phase is not enough for the whole BC phase in a medium to large size MWRN. Thus, there should be a constraint on the number of users in SWIPT-enabled MWRNs. A helpful idea is to perform energy allocation to the relay transmission power in different BC time slots.

Bibliography

[1] Ericsson mobility report. https://www.ericsson.com/4acd7e/assets/local/mobility-report/documents/2019/emr-november-2019.pdf, 2019. [Online; accessed 7-April-2020].

[2] Jeffrey G Andrews, Stefano Buzzi, Wan Choi, Stephen V Hanly, Angel Lozano, Anthony CK Soong, and Jianzhong Charlie Zhang. What will 5G be? *IEEE Journal on Selected Areas in Communications*, 32(6): 1065–1082, 2014.

[3] KJ Ray Liu, Ahmed K Sadek, Weifeng Su, and Andres Kwasinski. *Cooperative communications and networking*. Cambridge university press, 2009.

[4] Cheng-Xiang Wang, Fourat Haider, Xiqi Gao, Xiao-Hu You, Yang Yang, Dongfeng Yuan, Hadi M Aggoune, Harald Haas, Simon Fletcher, and Erol Hepsaydir. Cellular architecture and key technologies for 5G wireless communication networks. *IEEE Communications Magazine*, 52(2):122–130, 2014.

[5] Shanzhi Chen and Jian Zhao. The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication. *IEEE Communications Magazine*, 52(5):36–43, 2014.

[6] Monowar Hasan, Ekram Hossain, and Dong In Kim. Resource allocation under channel uncertainties for relay-aided device-to-device communication underlaying LTE-A cellular networks. *IEEE Transactions on Wireless Communications*, 13(4):2322–2338, 2014.

[7] Khoa T Phan, Tho Le-Ngoc, Sergiy A Vorobyov, and Chintha Tellambura. Power allocation in wireless multi-user relay networks. *IEEE Transactions on Wireless Communications*, 8(5):2535–2545, 2009.

[8] Siavash Fazeli-Dehkordy, Shahram Shahbazpanahi, and Saeed Gazor. Multiple peer-to-peer communications using a network of relays. *IEEE Transactions on Signal Processing*, 57(8):3053–3062, 2009.

[9] Michael Gastpar, Gerhard Kramer, and Piyush Gupta. The multiple-relay channel: Coding and antenna-clustering capacity. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, number LCAV-CONF-2002-008, page 136, 2002.

[10] Alain Sibille, Claude Oestges, and Alberto Zanella. *MIMO: from theory to implementation*. Academic Press, 2010.

[11] Jerry R Hampton. *Introduction to MIMO communications*. Cambridge university press, 2013.

[12] Fredrik Rusek, Daniel Persson, Buon Kiong Lau, Erik G Larsson, Thomas L Marzetta, Ove Edfors, and Fredrik Tufvesson. Scaling up MIMO: Opportunities and challenges with very large arrays. *IEEE Signal Procesing Magazine*, 30(1):40–60, 2013.

[13] Thomas L Marzetta and Hong Yang. *Fundamentals of massive MIMO*. Cambridge University Press, 2016.

[14] Erik G Larsson, Ove Edfors, Fredrik Tufvesson, and Thomas L Marzetta. Massive MIMO for next generation wireless systems. *IEEE Communications Magazine*, 52(2):186–195, 2014.

[15] Bin Le, Thomas W Rondeau, Jeffrey H Reed, and Charles W Bostian. Analog-to-digital converters. *IEEE Signal Processing Magazine*, 22(6): 69–77, 2005.

[16] Robert H Walden. Analog-to-digital converter survey and analysis. *IEEE Journal on selected areas in communications*, 17(4):539–550, 1999.

[17] Christer Svensson, Stefan Andersson, and Peter Bogner. On the power consumption of analog to digital converters. In *2006 NORCHIP*, pages 49–52. IEEE, 2006.

[18] Young-Han Nam, Boon Loong Ng, Krishna Sayana, Yang Li, Jianzhong Zhang, Younsun Kim, and Juho Lee. Full-dimension MIMO (FD-MIMO) for next generation cellular technology. *IEEE Communications Magazine*, 51(6):172–179, 2013.

[19] Johannes Koppenborg, Hardy Halbauer, Stephan Saur, and Cornelis Hoek. 3D beamforming trials with an active antenna array. In *2012 International ITG Workshop on Smart Antennas (WSA)*, pages 110–114. IEEE, 2012.

[20] Hardy Halbauer, Stephan Saur, Johannes Koppenborg, and Cornelis Hoek. 3D beamforming: Performance improvement for cellular networks. *Bell Labs Technical Journal*, 18(2):37–56, 2013.

[21] Xiaojia Lu, Antti Tolli, Olli Piirainen, Markku Juntti, and Wei Li. Comparison of antenna arrays in a 3-d multiuser multicell network. In *2011 IEEE international Conference on Communications (ICC)*, pages 1–6. IEEE, 2011.

[22] Shangbin Wu, Cheng-Xiang Wang, Mohammed M Alwakeel, Yejun He, et al. A non-stationary 3-d wideband twin-cluster model for 5G massive MIMO channels. *IEEE Journal on Selected Areas in Communications*, 32(6):1207–1218, 2014.

[23] Young-Han Nam, Yang Li, and Jianzhong Charlie Zhang. 3D channel models for elevation beamforming and FD-MIMO in LTE-A and 5G. In *2014 48th Asilomar Conference on Signals, Systems and Computers*, pages 805–809. IEEE, 2014.

[24] Erol Gelenbe and Yves Caseau. The impact of information technology on energy consumption and carbon emissions. *Ubiquity*, 2015(June):1, 2015.

[25] Ioannis Krikidis, Stelios Timotheou, Symeon Nikolaou, Gan Zheng, Derrick Wing Kwan Ng, and Robert Schober. Simultaneous wireless information and power transfer in modern communication systems. *IEEE Communications Magazine*, 52(11):104–110, 2014.

[26] Rui Zhang and Chin Keong Ho. MIMO broadcasting for simultaneous wireless information and power transfer. *IEEE Transactions on Wireless Communications*, 12(5):1989–2001, 2013.

[27] Ali A Nasir, Xiangyun Zhou, Salman Durrani, and Rodney A Kennedy. Relaying protocols for wireless energy harvesting and information pro-

cessing. *IEEE Transactions on Wireless Communications*, 12(7):3622–3636, 2013.

[28] Zhouyue Pi and Farooq Khan. An introduction to millimeter-wave mobile broadband systems. *Communications Magazine, IEEE*, 49(6):101–107, 2011.

[29] Joseph Mitola III and Gerald Q Maguire Jr. Cognitive radio: making software radios more personal. *Personal Communications, IEEE*, 6(4): 13–18, 1999.

[30] Yuya Saito, Yoshihisa Kishiyama, Anass Benjebbour, Takehiro Nakamura, Anxin Li, and Kenichi Higuchi. Non-orthogonal multiple access (NOMA) for cellular future radio access. In *2013 IEEE 77th vehicular technology conference (VTC Spring)*, pages 1–5. IEEE, 2013.

[31] Junpei Umehara, Yoshihisa Kishiyama, and Kenichi Higuchi. Enhancing user fairness in non-orthogonal access with successive interference cancellation for cellular downlink. In *2012 IEEE International Conference on Communication Systems (ICCS)*, pages 324–328. IEEE, 2012.

[32] Nagisa Otao, Yoshihisa Kishiyama, and Kenichi Higuchi. Performance of non-orthogonal access with SIC in cellular downlink using proportional fair-based resource allocation. In *2012 international symposium on wireless communication systems (ISWCS)*, pages 476–480. IEEE, 2012.

[33] Jeffrey G Andrews. Seven ways that hetnets are a cellular paradigm shift. *IEEE Communications Magazine*, 51(3):136–144, 2013.

[34] Jeffrey G Andrews, Holger Claussen, Mischa Dohler, Sundeep Rangan, and Mark C Reed. Femtocells: Past, present, and future. *IEEE Journal on Selected Areas in communications*, 30(3):497–508, 2012.

[35] Aditya Umbu Tana Amah and Anja Klein. Non-regenerative multi-antenna multi-group multi-way relaying. *EURASIP Journal on Wireless Communications and Networking*, 2011(1):1–19, 2011.

[36] Aditya Umbu Tana Amah and Anja Klein. Non-regenerative multi-way relaying with linear beamforming. In *Personal, Indoor and Mobile Radio*

Communications, 2009 IEEE 20th International Symposium on, pages 1843–1847. IEEE, 2009.

[37] Deniz Gunduz, Aylin Yener, Andrea Goldsmith, and H Vincent Poor. The multiway relay channel. *IEEE Transactions on Information Theory*, 59(1):51–63, 2013.

[38] J Nicholas Laneman, David NC Tse, and Gregory W Wornell. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Transactions on Information Theory*, 50(12):3062–3080, 2004.

[39] Peter Larsson and Hu Rong. Large-scale cooperative relaying network with optimal coherent combining under aggregate relay power constraints. In *WWRF 2004*, 2004.

[40] Erdem Koyuncu, Yindi Jing, and Hamid Jafarkhani. Distributed beamforming in wireless relay networks with quantized feedback. *IEEE Journal on Selected Areas in Communications*, 26(8):1429–1439, 2008.

[41] Qian Cao, H Vicky Zhao, and Yindi Jing. Power allocation and pricing in multiuser relay networks using stackelberg and bargaining games. *IEEE Transactions on Vehicular Technology*, 61(7):3177–3190, 2012.

[42] Boris Rankov and Armin Wittneben. Spectral efficient protocols for half-duplex fading relay channels. *Selected Areas in Communications, IEEE Journal on*, 25(2):379–389, 2007.

[43] Brian P Day, Adam R Margetts, Daniel W Bliss, and Philip Schniter. Full-duplex MIMO relaying: Achievable rates under limited dynamic range. *Selected Areas in Communications, IEEE Journal on*, 30(8):1541–1553, 2012.

[44] Rudolf Ahlswede, Ning Cai, S-YR Li, and Raymond W Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46(4):1204–1216, 2000.

[45] Hyun Jong Yang and Joohwan Chun. Generalized Schur decomposition-based two-way relaying for wireless MIMO systems. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008*, pages 1–6, 2008.

[46] Min Chen and Aylin Yener. Multiuser two-way relaying: detection and interference management strategies. *IEEE Transactions on Wireless Communications*, 8(8):4296–4305, 2009.

[47] Jingon Joung and Ali H Sayed. Multiuser two-way amplify-and-forward relay processing and power control methods for beamforming systems. *IEEE Transactions on Signal Processing*, 58(3):1833–1846, 2009.

[48] Jiayi Zhang, Xipeng Xue, Emil Björnson, Bo Ai, and Shi Jin. Spectral efficiency of multipair massive MIMO two-way relaying with hardware impairments. *IEEE Wireless Communications Letter*, 7(1):14–17, 2018.

[49] Lawrence Ong, Sarah J Johnson, and Christopher M Kellett. The capacity region of multiway relay channels over finite fields with full data exchange. *IEEE Transactions on Information Theory*, 57(5):3016–3031, 2011.

[50] Samira Rahimian, Wuhua Zhang, Moslem Noori, Yindi Jing, and Masoud Ardakani. Partial zero-forcing for multi-way relay networks. *IEEE Transactions on Communications*, 66(10):4444–4456, 2018.

[51] Lawrence Ong, Sarah J Johnson, and Christopher M Kellett. An optimal coding strategy for the binary multi-way relay channel. *IEEE Communications Letters*, 14(4):330–332, 2010.

[52] Gayan Amarasuriya and H Vincent Poor. Multi-way amplify-and-forward relay networks with massive MIMO. In *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, pages 595–600. IEEE, 2014.

[53] Moslem Noori and Masoud Ardakani. On the capacity gap of gaussian multi-way relay channels. In *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pages 1–4. IEEE, 2012.

[54] Moslem Noori and Masoud Ardakani. Optimal user pairing for asymmetric multi-way relay channels with pairwise relaying. *IEEE Communications Letters*, 16(11):1852–1855, 2012.

[55] Aditya Umbu Tana Amah and Anja Klein. Non-regenerative multi-way relaying: Space-time analog network coding and repetition. *IEEE Communications Letters*, 15(12):1362–1364, 2011.

[56] Gayan Amarasuriya, Chintha Tellambura, and Masoud Ardakani. Multi-way MIMO amplify-and-forward relay networks with zero-forcing transmission. *IEEE Transactions on Communications*, 61(12):4847–4863, 2013.

[57] Hui Gao, Yuan Ren, Chau Yuen, and Tiejun Lv. A distributed user scheduling scheme for MIMO multi-way relay channel. In *Communications (ICC), 2014 IEEE International Conference on*, pages 4826–4831. IEEE, 2014.

[58] Holger Degenhardt, Yue Rong, and Anja Klein. Non-regenerative multi-way relaying: Combining the gains of network coding and joint processing. *IEEE Transactions on Wireless Communications*, 12(11):5692–5703, 2013.

[59] Wen Li and Min Dong. Joint relay beamforming and receiver processing for multi-way multi-antenna relay networks. *IEEE Transactions on Communications*, 66(2):576–588, 2018.

[60] Wen Li and Min Dong. Joint relay beamforming and receiver processing for multi-way multi-antenna relay networks. *IEEE Transactions on Communications*, 2017.

[61] Yindi Jing and Hamid Jafarkhani. Network beamforming using relays with perfect channel information. *IEEE Transactions on Information Theory*, 55(6):2499–2517, 2009.

[62] J Nicholas Laneman, Gregory W Wornell, and David NC Tse. An efficient protocol for realizing cooperative diversity in wireless networks. In *Information Theory, 2001. Proceedings. 2001 IEEE International Symposium on*, page 294. IEEE, 2001.

[63] Xiaojun Tang and Yingbo Hua. Optimal design of non-regenerative MIMO wireless relays. *IEEE Transactions on Wireless Communications*, 6(4):1398–1407, 2007.

[64] Olga Muñoz-Medina, Josep Vidal, and Adrián Agustín. Linear transceiver design in nonregenerative relays with channel state information. *IEEE Transactions on Signal Processing.*, 55(6):2593–2604, 2007.

[65] Chung Duc Ho, Hien Quoc Ngo, Michail Matthaiou, and Trung Q Duong. On the performance of zero-forcing processing in multi-way massive MIMO relay networks. *IEEE Communications Letters*, 21(4):849–852, 2017.

[66] Aditya Umbu Tana Amah and Anja Klein. Beamforming-based physical layer network coding for non-regenerative multi-way relaying. *EURASIP Journal on Wireless Communications and Networking*, 2010:7, 2010.

[67] Michael Joham, Wolfgang Utschick, Josef Nossek, et al. Linear transmit processing in MIMO communications systems. *IEEE Transactions on Signal Processing.*, 53(8):2700–2712, 2005.

[68] Christian B Peel, Bertrand M Hochwald, and A Lee Swindlehurst. A vector-perturbation technique for near-capacity multiantenna multiuser communication-part i: channel inversion and regularization. *IEEE Transactions on Communications*, 53(1):195–202, 2005.

[69] Joel Max. Quantizing for minimum distortion. *IRE Transactions on Information Theory*, 6(1):7–12, 1960.

[70] Stuart Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

[71] Li Fan, Shi Jin, Chao-Kai Wen, and Haixia Zhang. Uplink achievable rate for massive MIMO systems with low-resolution ADC. *IEEE Communications Letters*, 19(12):2186–2189, 2015.

[72] Sven Jacobsson, Giuseppe Durisi, Mikael Coldrey, Ulf Gustavsson, and Christoph Studer. Throughput analysis of massive MIMO uplink with low-resolution ADCs. *IEEE Transactions on Wireless Communications*, 16(6):4038–4051, 2017.

[73] Julian Jakob Bussgang. Crosscorrelation functions of amplitude-distorted Gaussian signals. 1952.

[74] Wenyi Zhang. A general framework for transmission with transceiver distortion and some applications. *IEEE Transactions on Communications*, 60(2):384–399, 2012.

[75] Ning Liang and Wenyi Zhang. A mixed-ADC receiver architecture for massive MIMO systems. In *Information Theory Workshop-Fall (ITW), 2015 IEEE*, pages 229–233. IEEE, 2015.

[76] Weiqiang Tan, Shi Jin, Chao-Kai Wen, and Yindi Jing. Spectral efficiency of mixed-ADC receivers for massive MIMO systems. *IEEE Access*, 4: 7841–7846, 2016.

[77] EUTR Access. Further advancements for E-UTRA physical layer aspects. *3GPP Technical Specification TR*, 36:V2, 2010.

[78] Antonia M Tulino, Sergio Verdú, et al. Random matrix theory and wireless communications. *Foundations and Trends® in Communications and Information Theory*, 1(1):1–182, 2004.

[79] Thomas L Marzetta. Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE transactions on wireless communications*, 9(11):3590–3600, 2010.

[80] Ansuman Adhikary, Junyoung Nam, Jae-Young Ahn, and Giuseppe Caire. Joint spatial division and multiplexing—the large-scale array regime. *IEEE transactions on information theory*, 59(10):6441–6463, 2013.

[81] Emil Björnson, Jakob Hoydis, Marios Kountouris, and Merouane Debbah. Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits. *IEEE Transactions on Information Theory*, 60(11):7112–7139, 2014.

[82] Jakob Hoydis, Mérouane Debbah, and Mari Kobayashi. Asymptotic moments for interference mitigation in correlated fading channels. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2796–2800. IEEE, 2011.

[83] Abla Kammoun, Axel Müller, Emil Björnson, and Mérouane Debbah. Linear precoding based on polynomial expansion: Large-scale multi-cell

MIMO systems. *IEEE Journal of Selected Topics in Signal Processing*, 8 (5):861–875, 2014.

[84] Axel Mueller, Abla Kammoun, Emil Björnson, and Mérouane Debbah. Linear precoding based on polynomial expansion: Reducing complexity in massive MIMO. *EURASIP journal on wireless communications and networking*, 2016(1):63, 2016.

[85] Fred E. Szabo. U. In Fred E. Szabo, editor, *The Linear Algebra Survival Guide*, pages 385 – 392. Academic Press, Boston, 2015. ISBN 978-0-12-409520-5. doi: https://doi.org/10.1016/B978-0-12-409520-5.50028-X. URL `http://www.sciencedirect.com/science/article/pii/B978012409520550028X`.

[86] Fumio Hiai and Dénes Petz. *Asymptotic freeness almost everywhere for random matrices*. University of Aarhus. Centre for Mathematical Physics and Stochastics (MaPhySto)[MPS], 1999.

[87] Nathaniel R Goodman. Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction). *The Annals of mathematical statistics*, 34(1):152–177, 1963.

[88] Dirk Maiwald and Dieter Kraus. On moments of complex wishart and complex inverse wishart distributed matrices. In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, pages 3817–3820, 1997.

[89] Chung Duc Ho, Hien Quoc Ngo, Michail Matthaiou, and Trung Q Duong. Multi-way massive MIMO with maximum-ratio processing and imperfect CSI. In *Signal Processing Conference (EUSIPCO), 2017 25th European*, pages 1704–1708. IEEE, 2017.

[90] Himal A Suraweera, Hien Quoc Ngo, Trung Q Duong, Chau Yuen, and Erik G Larsson. Multi-pair amplify-and-forward relaying with very large antenna arrays. In *2013 IEEE International Conference on Communications (ICC)*, pages 4635–4640. IEEE, 2013.

[91] Wuhua Zhang. Partial zero forcing for multi-way relay networks. Master's thesis, University of Alberta, Canada, 2015.

[92] Aditya Umbu Tana Amah and Anja Klein. Regenerative multi-group multi-way relaying. *IEEE Transactions on Vehicular Technology*, 60(7): 3017–3029, 2011.

[93] Syed Ali Jafar and Shlomo Shamai. Degrees of freedom region of the MIMO channel. *IEEE Transactions on Information Theory*, 54(1):151–170, 2008.

[94] Emil Björnson, Jakob Hoydis, Luca Sanguinetti, et al. Massive MIMO networks: Spectral, energy, and hardware efficiency. *Foundations and Trends® in Signal Processing*, 11(3-4):154–655, 2017.

[95] Jiayi Zhang, Linglong Dai, Xu Li, Ying Liu, and Lajos Hanzo. On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems. *IEEE Communications Magazine*, 56(7):205–211, 2018.

[96] Jiayi Zhang, Linglong Dai, Shengyang Sun, and Zhaocheng Wang. On the spectral efficiency of massive MIMO systems with low-resolution ADCs. *IEEE Communications Letters*, 20(5):842–845, 2016.

[97] Hae-Seung Lee and Charles G Sodini. Analog-to-digital converters: Digitizing the analog world. *Proc. IEEE*, 96(2):323–334, 2008.

[98] Yongzhi Li, Cheng Tao, Gonzalo Seco-Granados, Amine Mezghani, A Lee Swindlehurst, and Liu Liu. Channel estimation and performance analysis of one-bit massive MIMO systems. *IEEE Transactions on Signal Processing.*, 65(15):4075–4089, 2017.

[99] Ahmet Gokceoglu, Emil Björnson, Erik G Larsson, and Mikko Valkama. Spatio-temporal waveform design for multiuser massive MIMO downlink with 1-bit receivers. *IEEE Journal of Selected Topics in Signal Processing*, 11(2):347–362, 2017.

[100] Peihao Dong, Hua Zhang, Wei Xu, and Xiaohu You. Efficient low-resolution ADC relaying for multiuser massive MIMO system. *IEEE Transactions on Vehicular Technology*, 66(12):11039–11056, 2017.

[101] Christoph Studer and Giuseppe Durisi. Quantized massive MU-MIMO-OFDM uplink. *IEEE Transactions on Communications*, 64(6):2387–2399, 2016.

[102] Mengjiao Zhang, Weiqiang Tan, Junhui Gao, and Shi Jin. Spectral efficiency and power allocation for mixed-adc massive MIMO system. *China Communications*, 15(3):112–127, 2018.

[103] Ning Liang and Wenyi Zhang. Mixed-ADC massive MIMO. *IEEE Journal on Selected Areas in Communications*, 34(4):983–997, 2016.

[104] Sven Jacobsson, Giuseppe Durisi, Mikael Coldrey, Ulf Gustavsson, and Christoph Studer. One-bit massive MIMO: Channel estimation and high-order modulations. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 1304–1309. IEEE, 2015.

[105] Junil Choi, Jianhua Mo, and Robert W Heath. Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs. *IEEE Transactions on Communications*, 64(5): 2005–2018, 2016.

[106] Jianhua Mo, Philip Schniter, and Robert W Heath. Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs. *IEEE Transactions on Signal Processing*, 66(5):1141–1154, 2018.

[107] Hessam Pirzadeh and A Lee Swindlehurst. Spectral effciency of mixed-ADC massive MIMO. *IEEE Transactions on Signal Processing.*, 66(13): 3599–3613, 2018.

[108] Qingfeng Ding and Yindi Jing. Receiver energy efficiency and resolution profile design for massive MIMO uplink with mixed ADC. *IEEE Transactions on Vehicular Technology*, 67(2):1840–1844, 2018.

[109] Qingfeng Ding and Yindi Jing. Outage probability analysis and resolution profile design for massive MIMO uplink with mixed-ADC. *IEEE Transactions on Wireless Communications*, 17(9):6293–6306, 2018.

[110] Chuili Kong, Caijun Zhong, Shi Jin, Sheng Yang, Hai Lin, and Zhaoyang Zhang. Full-duplex massive MIMO relaying systems with low-resolution ADCs. *IEEE Transactions on Wireless Communications*, 16(8):5033–5047, 2017.

[111] Jian Liu, Jindan Xu, Wei Xu, Shi Jin, and Xiaodai Dong. Multiuser massive MIMO relaying with mixed-ADC receiver. *IEEE Signal Process. Lett.*, 24(1):76–80, 2017.

[112] Chuili Kong, Amine Mezghani, Caijun Zhong, A Lee Swindlehurst, and Zhaoyang Zhang. Multipair massive MIMO relaying systems with one-bit ADCs and DACs. *IEEE Transactions on Signal Processing.*, 66(11): 2984–2997, 2018.

[113] Jiayi Zhang, Linglong Dai, Ziyan He, Bo Ai, and Octavia A Dobre. Mixed-ADC/DAC multipair massive MIMO relaying systems: Performance analysis and power optimization. *IEEE Transactions on Communications*, 67(1):140–153, 2019.

[114] Amine Mezghani and Josef A Nossek. Capacity lower bound of MIMO channels with output quantization and correlated noise. In *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2012.

[115] MinChul Ju, Kyu-Min Kang, Kyu-Sung Hwang, and Cheol Jeong. Maximum transmission rate of PSR/TSR protocols in wireless energy harvesting DF-based relay networks. *IEEE Journal on Selected Areas in Communications*, 33(12):2701–2717, 2015.

[116] Hongwu Liu and Kyung Sup Kwak. Multipair massive MIMO relay with simultaneous wireless information and power transfer. In *2016 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 225–230. IEEE, 2016.

[117] Jinlong Wang, Liming Zheng, Ming Ding, Gang Wang, and Zihuai Lin. Performance analysis of massive MIMO two-way relay systems with SWIPT. In *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, pages 1–6. IEEE, 2019.

[118] MinChul Ju and Hong-Chuan Yang. Optimum design of energy harvesting relay for two-way decode-and-forward relay networks under max-min and max-sum criterions. *IEEE Transactions on Communications*, 67(10): 6682–6697, 2019.

[119] Xinhua Wang, Ju Liu, and Chao Zhai. Wireless power transfer-based multi-pair two-way relaying with massive antennas. *IEEE Transactions on Wireless Communications*, 16(11):7672–7684, 2017.

[120] Gayan Amarasuriya, Erik G Larsson, and H Vincent Poor. Wireless information and power transfer in multiway massive MIMO relay networks. *IEEE Transactions on Wireless Communications*, 15(6):3837–3855, 2016.

[121] Xiang Cheng, Bo Yu, Liuqing Yang, Jianhua Zhang, Guangyi Liu, Yong Wu, and Lei Wan. Communicating in the real world: 3D MIMO. *IEEE Wireless Communications*, 21(4):136–144, 2014.

[122] Ioannis Krikidis. SWIPT in 3-D bipolar ad hoc networks with sectorized antennas. *IEEE Communications Letters*, 20(6):1267–1270, 2016.

[123] Lixing Fan, Haiyang Zhang, Yongming Huang, and Luxi Yang. Exploiting BS antenna tilt for SWIPT in 3-D massive MIMO systems. *IEEE Wireless Communications Letter*, 6(5):666–669, 2017.

[124] Samira Rahimian, Yindi Jing, and Masoud Ardakani. Performance analysis and optimization of 3D massive MIMO multi-pair relaying with SWIPT. *arXiv preprint arXiv:2001.11803*, 2020.

[125] Elena Boshkovska, Derrick Wing Kwan Ng, Nikola Zlatanov, and Robert Schober. Practical non-linear energy harvesting model and resource allocation for SWIPT systems. *IEEE Communications Letters*, 19(12): 2082–2085, 2015.

[126] Babak Hassibi and Bertrand M Hochwald. How much training is needed in multiple-antenna wireless links? *IEEE Transactions on Information Theory*, 49(4):951–963, 2003.

[127] Cristoff Martin and Björn Ottersten. Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading. *IEEE Transactions on Wireless Communications*, 3(4):1350–1359, 2004.

[128] Qi Zhang, Shi Jin, Kai-Kit Wong, Hongbo Zhu, and Michail Matthaiou. Power scaling of uplink massive MIMO systems with arbitrary-rank channel means. *IEEE Journal on Selected Topics in Signal Processing*, 8(5): 966–981, 2014.

Appendices

Appendix A

Proofs for Chapter 4

A.1 Proof of Theorem 1

The coefficient $\alpha^{(t)}$ is the solution of $P_R = \mathbb{E}\{\|\mathbf{r}_{\mathrm{t}}^{(t)}\|^2\}$, where the right-hand-side can be written as the sum of several terms. In calculating each term, we first simplify the expression by the SVD of the channel matrix. Then, we use properties of Wishart matrices, Haar matrices, and singular values for Gaussian matrices to calculate the value. From (3.1), (4.8), and (4.11) we can write

$$\mathbf{r}_{\mathrm{t}}^{(t)} = \mathbf{G}^{(t)}\mathbf{G}_b(\sqrt{p_u}\mathbf{H}\mathbf{x} + \mathbf{z}_{\mathrm{R}}) + \mathbf{G}^{(t)}\mathbf{d}.$$

As $\mathbf{r}_{\mathrm{a}}$ and $\mathbf{d}$ are assumed to be uncorrelated, we can write $\mathbb{E}\{\|\mathbf{r}_{\mathrm{t}}^{(t)}\|^2\} = c_1 + c_2 + c_3$ where

$$\begin{aligned} c_1 &\triangleq p_u\mathbb{E}[\mathrm{tr}\{\mathbf{G}^{(t)}\mathbf{G}_b\mathbf{H}\mathbf{H}^H\mathbf{G}_b(\mathbf{G}^{(t)})^H\}], \\ c_2 &\triangleq \mathbb{E}[\mathrm{tr}\{\mathbf{G}^{(t)}\mathbf{G}_b^2(\mathbf{G}^{(t)})^H\}], c_3 \triangleq \mathbb{E}[\mathrm{tr}\{\mathbf{G}^{(t)}\mathbf{d}\mathbf{d}^H(\mathbf{G}^{(t)})^H\}]. \end{aligned}$$

For the calculation of c_1, we use the following approximation[1]

$$\mathbb{E}[\mathrm{tr}\{\mathbf{G}^{(t)}\mathbf{G}_b\mathbf{H}\mathbf{H}^H\mathbf{G}_b(\mathbf{G}^{(t)})^H\} \approx \left(\frac{1}{N}\sum_{n=1}^{N} G_{b_n}\right)^2 \mathbb{E}[\mathrm{tr}\{\mathbf{G}^{(t)}\mathbf{H}\mathbf{H}^H(\mathbf{G}^{(t)})^H\}].$$

Then, by using the $\mathbf{G}^{(t)}$ expression in (4.7), $c_1 \approx \alpha^{(t)}p_u g_1^2\mathbb{E}[\mathrm{tr}\{(\mathbf{H}^T\mathbf{H}^*)^{-1}\}]$. Since $\mathbf{H}$ has i.i.d. rows following $\mathcal{CN}(0, \mathbf{D})$, where $\mathbf{D} = \mathrm{diag}\{\beta_1, \beta_2, \cdots, \beta_K\}$, $\mathbf{H}^T\mathbf{H}^*$ is a $K \times K$ Wishart matrix of N degrees of freedom. Therefore, according

[1]This approximation is obtained by replacing the $\mathbf{G}_b$ matrices on the left-hand side by $\frac{1}{N}\sum_{n=1}^{N} G_{b_n}\mathbf{I}_N$. Our simulation results show that this approximation is tight for large N.

to Lemma 2, we have $\mathbb{E}\left[\left((\mathbf{H}^T\mathbf{H}^*)^{-1}\right)_{ii}\right] = \frac{1}{(N-K)\beta_i}$. Finally,

$$c_1 \approx \frac{p_u \alpha^{(t)}}{N-K} g_1^2 \sum_{i=1}^{K} \frac{1}{\beta_i}. \tag{A.1}$$

The next is to calculate c_2. After using the $\mathbf{G}^{(t)}$ expression in (4.7),

$$c_2 = \alpha^{(t)} \mathrm{tr}\{\mathbf{G}_b^2 \mathbb{E}[\mathbf{B}]\}, \tag{A.2}$$

where

$$\mathbf{B} \triangleq \tilde{\mathbf{H}}(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}})^{-1}\mathbf{D}^{-\frac{1}{2}}(\mathbf{P}^t)^T\mathbf{D}^{-\frac{1}{2}}(\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}^*)^{-1}\mathbf{D}^{-\frac{1}{2}}\mathbf{P}^t\mathbf{D}^{-\frac{1}{2}}(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}})^{-1}\tilde{\mathbf{H}}^H.$$

Consider the SVD

$$\tilde{\mathbf{H}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \tag{A.3}$$

where $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{\Sigma}$ are $N \times K$, $K \times K$, and $K \times K$ matrices. $\mathbf{U}$ and $\mathbf{V}$ contain the left and right singular vectors of $\tilde{\mathbf{H}}$, respectively, and $\mathbf{\Sigma} = \mathrm{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_K\}$ contains the singular values of $\tilde{\mathbf{H}}$. Further, as mentioned in Section 2.5.1, according to Definition 2.5 in [78], $\mathbf{U}$, and $\mathbf{V}$ are Haar matrices. Then, we have

$$\mathbf{B} = \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^H\mathbf{D}^{-\frac{1}{2}}(\mathbf{P}^t)^T\mathbf{D}^{-\frac{1}{2}}\mathbf{V}^*\mathbf{\Sigma}^{-2}\mathbf{V}^T\mathbf{D}^{-\frac{1}{2}}\mathbf{P}^t\mathbf{D}^{-\frac{1}{2}}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^H.$$

Let $\mathbf{P}_{ij}$ be the unitary permutation matrix that changes the positions of the i-th and the j-th rows of a matrix if it is multiplied from the left side. Then, using the fact that $\mathbf{U}$ and $\mathbf{P}_{ij}\mathbf{U}$ have the same distribution, we conclude that

$$\mathbf{B}' \triangleq \mathbf{P}_{ij}\mathbf{B}\mathbf{P}_{ij}^H = \mathbf{P}_{ij}\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^H\mathbf{D}^{-\frac{1}{2}}(\mathbf{P}^t)^T\mathbf{D}^{-\frac{1}{2}}\mathbf{V}^*\mathbf{\Sigma}^{-2}\mathbf{V}^T\mathbf{D}^{-\frac{1}{2}}\mathbf{P}^t\mathbf{D}^{-\frac{1}{2}}\mathbf{V}\mathbf{\Sigma}^{-1}(\mathbf{P}_{ij}\mathbf{U})^H,$$

has the same distribution as $\mathbf{B}$. From the construction of $\mathbf{B}'$, we have $b_{ii} = b'_{jj}$. It can thus be concluded that $\mathbb{E}\{b_{ii}\}$ is the same for all $i \in \{1, 2, \cdots, N\}$ and (A.2) can be written as

$$c_2 = \alpha^{(t)} \left(\frac{1}{N}\sum_{n=1}^{N} G_{b_n}^2\right) \mathbb{E}[\mathrm{tr}\{\mathbf{B}\}]. \tag{A.4}$$

Next, we calculate $\mathbb{E}[\text{tr}\{\mathbf{B}\}] = \text{tr}\{\mathbb{E}[\mathbf{B}]\}$,

$$\text{tr}\{\mathbb{E}[\mathbf{B}]\} = \sum_{m=1}^{K} \frac{1}{\beta_m \beta_{i(m)}} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \mathbb{E}\left[\frac{|v_{i(m)k_1}|^2 |v_{mk_2}|^2}{\sigma_{k_1}^2 \sigma_{k_2}^2}\right]. \tag{A.5}$$

According to Lemma 1, for $k_1 \neq k_2$ and any m we have

$$\mathbb{E}[|v_{i(m)k_1}|^2 |v_{mk_2}|^2] = \frac{1}{K^2 - 1}. \tag{A.6}$$

Further, according to Theorem 1 in [127], the eigenvalues of the Wishart matrix $\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}$, $\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_K^2\}$, become independent as $N \to \infty$. Thus, when $k_1 \neq k_2$ and $N \gg 1$,

$$\mathbb{E}\left[\frac{1}{\sigma_{k_1}^2 \sigma_{k_2}^2}\right] \approx \left(\mathbb{E}\left[\frac{1}{\sigma_{k_1}^2}\right]\right)^2 = \frac{1}{(N-K)^2}, \tag{A.7}$$

where for the last step we have used the equality $\mathbb{E}\left[\frac{1}{\sigma_k^2}\right] = \frac{1}{K}\mathbb{E}[\text{tr}\{(\tilde{\mathbf{H}}^H \tilde{\mathbf{H}})^{-1}\}] = \frac{1}{N-K}$, for any k. Also, since the entries of $\tilde{\mathbf{H}}$ follow i.i.d. $\mathcal{CN}(0,1)$, $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{\Sigma}$ are independent. Then, by using (A.6) and (A.7) in (A.5), for $k_1 \neq k_2$,

$$\mathbb{E}\left[\frac{|v_{i(m)k_1}|^2 |v_{mk_2}|^2}{\sigma_{k_1}^2 \sigma_{k_2}^2}\right] = \mathbb{E}[|v_{i(m)k_1}|^2 |v_{mk_2}|^2]\mathbb{E}\left[\frac{1}{\sigma_{k_2}^2 \sigma_{k_1}^2}\right] \approx \frac{1}{(K^2-1)(N-K)^2}. \tag{A.8}$$

Moreover, according to Lemma 1, for $k_1 = k_2 = k$ and any m and k we have

$$\mathbb{E}[|v_{i(m)k}|^2 |v_{mk}|^2] = \frac{1}{K(K+1)}. \tag{A.9}$$

Also, for any k,

$$\begin{aligned}\mathbb{E}\left[\frac{1}{\sigma_k^4}\right] &= \frac{1}{K}\mathbb{E}\left[\text{tr}\left\{(\tilde{\mathbf{H}}^H \tilde{\mathbf{H}})^{-1}\left((\tilde{\mathbf{H}}^H \tilde{\mathbf{H}})^{-1}\right)^H\right\}\right] = \frac{1}{K}\mathbb{E}[\text{tr}\{(\tilde{\mathbf{H}}^H \tilde{\mathbf{H}})^{-2}\}] \\ &= \frac{1}{(N-K)(N-K-1)},\end{aligned} \tag{A.10}$$

where the last step comes from the results of the second order statistics for inverse Wishart matrices in Lemmas 2. By combining (A.9) and (A.10), for

$k_1 = k_2 = k$,

$$\mathbb{E}\left[\frac{|v_{i(m)k}|^2|v_{mk}|^2}{\sigma_k^4}\right] = \frac{1}{K(K+1)(N-K)(N-K-1)}. \tag{A.11}$$

By using (A.8) and (A.11) in (A.5) and then (A.4), we have

$$c_2 \approx \frac{\alpha^{(t)} g_2 (NK + N - 2K - K^2)}{(N-K)^2(N-K-1)(K+1)} \sum_{m=1}^{K} \frac{1}{\beta_m \beta_{i(m)}}. \tag{A.12}$$

For c_3, with similar arguments as c_2, we can show that

$$\begin{aligned} c_3 =& \mathrm{tr}\{\mathbb{E}[\mathbf{d}\mathbf{d}^H]\mathbb{E}[(\mathbf{G}^{(t)})^H\mathbf{G}^{(t)}]\} = \mathrm{tr}\{\mathbf{C}_\mathbf{d}\mathbb{E}[(\mathbf{G}^{(t)})^H\mathbf{G}^{(t)}]\} \\ \approx& \frac{\alpha^{(t)}(\hat{c} - v g_2)(NK + N - 2K - K^2)}{(N-K)^2(N-K-1)(K+1)} \sum_{m=1}^{K} \frac{1}{\beta_m \beta_{i(m)}}. \end{aligned} \tag{A.13}$$

By combining (A.1), (A.12), and (A.13) and also ignoring lower order terms of N when N is large, i.e., we consider that $NK \gg K$ when $N \gg 1$, (4.13) is obtained.

A.2 Proof of Theorem 2

We use the common approximation $\mathbb{E}\{\log_2(1 + \frac{X}{Y})\} \approx \log_2(1 + \frac{\mathbb{E}\{X\}}{\mathbb{E}\{Y\}})$ for mMIMO systems. It is tight when $N \to \infty$ and X and Y are both sums of nonnegative random variables which converge to their means due to the law of large numbers [128]. Therefore,

$$R_{k,i(k)} \approx \log_2\left(1 + \frac{A_4}{A_1 + A_2 + A_3 + 1}\right),$$

where

$$\begin{aligned} A_1 \triangleq& p_u \mathbb{E}\left[\sum_{j=1, j\neq i(k)}^{K} |\mathbf{h}_k^T \mathbf{G}^{(t)} \mathbf{G}_b \mathbf{h}_j|^2\right], A_2 \triangleq \mathbb{E}\left[\|\mathbf{h}_k{}^T \mathbf{G}^{(t)} \mathbf{G}_b\|^2\right], \\ A_3 \triangleq& \mathbb{E}\left[\|\mathbf{h}_k{}^T \mathbf{G}^{(t)} \mathbf{d}\|^2\right], A_4 \triangleq p_u \mathbb{E}\left[|\mathbf{h}_k{}^T \mathbf{G}^{(t)} \mathbf{G}_b \mathbf{h}_{i(k)}|^2\right]. \end{aligned}$$

Similar to the proof in Appendix A.1, for the calculation of A_1, A_2, A_3, and A_4, we first simplify the expressions by the SVD of channel matrix. Then, we

use the properties of Wishart matrices, Haar matrices, and singular values for Gaussian matrices to calculate each value.

Let $\mathbf{e}_k$ be the kth canonical basis. Substituting $\mathbf{G}^{(t)}$ from (4.7) in A_1, we have

$$\begin{aligned} A_1 &= p_u \alpha^{(t)} \sum_{j=1, j\neq i(k)}^{K} \mathbb{E}\left[\mathbf{e}_{i(k)}^T (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{G}_b\mathbf{h}_j\mathbf{h}_j^H\mathbf{G}_b^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{e}_{i(k)}\right] \\ &= \frac{p_u\alpha^{(t)}}{\beta_{i(k)}} \sum_{j=1, j\neq i(k)}^{K} \beta_j \mathbb{E}\left[\left|((\tilde{\mathbf{H}}^H\tilde{\mathbf{H}})^{-1}\tilde{\mathbf{H}}^H\mathbf{G}_b\tilde{\mathbf{H}})_{i(k)j}\right|^2\right]. \end{aligned}$$

By using the SVD in (A.3),

$$\begin{aligned} A_1 &= \frac{p_u\alpha^{(t)}}{\beta_{i(k)}} \sum_{j=1, j\neq i(k)}^{K} \beta_j \mathbb{E}\left[\left|(\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^H\mathbf{G}_b\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H)_{i(k)j}\right|^2\right] \\ &= \frac{p_u\alpha^{(t)}}{\beta_{i(k)}} \sum_{j=1, j\neq i(k)}^{K} \beta_j \mathbb{E}\left[\left|\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\left(\frac{\sigma_{k_2}}{\sigma_{k_1}}\mathbf{v}_{k_1}\mathbf{u}_{k_1}^H\mathbf{G}_b\mathbf{u}_{k_2}\mathbf{v}_{k_2}^H\right)_{i(k)j}\right|^2\right] \\ &= \frac{p_u\alpha^{(t)}}{\beta_{i(k)}} \sum_{j=1, j\neq i(k)}^{K} \beta_j \sum_{k_1,k_2}^{K}\sum_{k_1',k_2'}^{K} \mathbb{E}\left[\frac{\sigma_{k_2}\sigma_{k_2'}}{\sigma_{k_1}\sigma_{k_1'}} v_{i(k),k_1}v_{j,k_2}^* v_{i(k),k_1'}^* v_{j,k_2'}\mathbf{u}_{k_1}^H\mathbf{G}_b\mathbf{u}_{k_2}\mathbf{u}_{k_2'}^H\mathbf{G}_b\mathbf{u}_{k_1'}\right]. \end{aligned} \tag{A.14}$$

As mentioned before, $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{\Sigma}$ are independent. Thus, from Lemma 1, if at least one of k_1, k_2, k_1', k_2' is different from the others, the corresponding expectation term in (A.14) is 0. The remaining terms in the summations in (A.14), are considered in the following four cases:

1. If $k_1 = k_2$, $k_1' = k_2'$, and $k_1 \neq k_1'$,

$$\begin{aligned} b_1 &\triangleq \sum_{k_1=1}^{K}\sum_{k_1'=1,\neq k_1}^{K} \mathbb{E}\left[v_{i(k),k_1}v_{j,k_1}^* v_{i(k),k_1'}^* v_{j,k_1'}\mathbf{u}_{k_1}^H\mathbf{G}_b\mathbf{u}_{k_1}\mathbf{u}_{k_1'}^H\mathbf{G}_b\mathbf{u}_{k_1'}\right] \\ &= -\frac{1}{(K+1)}\left[\frac{\sum_{n=1}^{N} G_{b_n}^2}{N(N+1)} + \frac{\sum_{n_1=1}^{N}\sum_{n_2=1,\neq n_1}^{N} G_{b_{n_1}}G_{b_{n_2}}}{N^2-1}\right]. \end{aligned}$$

2. If $k_1 = k_1'$, $k_2 = k_2'$, and $k_1 \neq k_2$,

$$b_2 \triangleq \sum_{k_1=1}^{K} \sum_{k_2=1,\neq k_1}^{K} \mathbb{E}\left[\frac{\sigma_{k_2}^2}{\sigma_{k_1}^2}|v_{i(k),k_1}|^2|v_{j,k_2}|^2|\mathbf{u}_{k_1}^H \mathbf{G}_b \mathbf{u}_{k_2}|^2\right]$$
$$= \sum_{k_1=1}^{K} \sum_{k_2=1,\neq k_1}^{K} \mathbb{E}\left[\frac{\sigma_{k_2}^2}{\sigma_{k_1}^2}\right] \frac{1}{(K^2-1)} \left[\frac{\sum_{n=1}^{N} G_{b_n}^2}{N(N+1)} - \frac{\sum_{n_1=1}^{N}\sum_{n_2=1,\neq n_1}^{N} G_{b_{n_1}} G_{b_{n_2}}}{N(N^2-1)}\right].$$

As mentioned earlier in the proof of Theorem 1, for $k_1 \neq k_2$, $\sigma_{k_1}^2$ and $\sigma_{k_2}^2$ are unordered eigenvalues of Wishart matrix which become independent as $N \to \infty$. Thus, for $N \gg 1$,

$$\sum_{k_1=1}^{K} \sum_{k_2=1,\neq k_1}^{K} \mathbb{E}\left[\frac{\sigma_{k_2}^2}{\sigma_{k_1}^2}\right] \approx \sum_{k_1=1}^{K} \sum_{k_2=1,\neq k_1}^{K} \mathbb{E}[\sigma_{k_2}^2]\mathbb{E}\left[\frac{1}{\sigma_{k_1}^2}\right].$$

For any $k_1 \neq k_2$, using the properties of Wishart and inverse Wishart matrices in Lemma 2, we have $\mathbb{E}\left[\frac{1}{\sigma_{k_1}^2}\right] = \frac{1}{N-K}$, and $\mathbb{E}[\sigma_{k_2}^2] = \frac{1}{K}\mathbb{E}[\mathrm{tr}\{\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\}] = N$. Therefore,

$$b_2 \approx \frac{NK}{(N-K)(K+1)} \left[\frac{\sum_{n=1}^{N} G_{b_n}^2}{N(N+1)} - \frac{\sum_{n_1=1}^{N}\sum_{n_2=1,\neq n_1}^{N} G_{b_{n_1}} G_{b_{n_2}}}{N(N^2-1)}\right].$$

3. If $k_1 = k_2'$, $k_2 = k_1'$, and $k_1 \neq k_2$,

$$b_3 \triangleq \sum_{k_1=1}^{K} \sum_{k_1'=1,\neq k_1}^{K} \mathbb{E}\left[v_{i(k),k_1} v_{j,k_1'}^* v_{i(k),k_1'}^* v_{j,k_1} \mathbf{u}_{k_1}^H \mathbf{G}_b \mathbf{u}_{k_1'} \mathbf{u}_{k_1}^H \mathbf{G}_b \mathbf{u}_{k_1'}\right] = 0.$$

4. If $k_1 = k_2 = k_1' = k_2'$,

$$b_4 \triangleq \sum_{k=1}^{K} \mathbb{E}\left[|v_{i(k),k}|^2|v_{j,k}|^2|\mathbf{u}_k^H \mathbf{G}_b \mathbf{u}_k|^2\right] = \left[\frac{\sum_{n=1}^{N} G_{b_n}^2 + \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} G_{b_{n_1}} G_{b_{n_2}}}{(K+1)N(N+1)}\right].$$

By using the above results on b_1, b_2, b_3, and b_4 in (A.14), for $N \gg 1$,

$$A_1 \approx \frac{p_u \alpha^{(t)}(N-K+NK)}{\beta_{i(k)}(K+1)(N-K)N(N^2-1)} \left[N\sum_{n=1}^{N} G_{b_n}^2 - \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} G_{b_{n_1}} G_{b_{n_2}}\right] \sum_{j=1,j\neq i(k)}^{K} \beta_j$$
$$\approx \frac{p_u}{\beta_{i(k)}} \frac{\alpha^{(t)}}{(N-K)}(g_2 - g_1^2)\beta_{\setminus i(k)}. \qquad \text{(A.15)}$$

Next, we calculate A_2. After substituting $\mathbf{G}^{(t)}$ from (4.7),

$$\begin{aligned}
A_2 =&\alpha^{(t)}\mathbb{E}\left[\mathbf{e}_{i(k)}^T\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{G}_b^2\mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{e}_{i(k)}\right]\\
=&\frac{\alpha^{(t)}}{\beta_{i(k)}}\mathbb{E}\left[\left(\left(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\right)^{-1}\tilde{\mathbf{H}}^H\mathbf{G}_b^2\tilde{\mathbf{H}}\left(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\right)^{-1}\right)_{i(k)i(k)}\right]\\
=&\frac{\alpha^{(t)}}{\beta_{i(k)}}\mathbb{E}\left[\left(\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^H\mathbf{G}_b^2\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^H\right)_{i(k)i(k)}\right].
\end{aligned}$$

Following similar reasoning as the one which led to (A.4), we have

$$A_2 = \frac{\alpha^{(t)}}{\beta_{i(k)}K}\mathbb{E}\left[\mathrm{tr}(\mathbf{G}_b^2\mathbf{U}\mathbf{\Sigma}^{-2}\mathbf{U}^H)\right] = \frac{\alpha^{(t)}}{\beta_{i(k)}K}\mathbb{E}\left[\mathrm{tr}(\mathbf{G}_b^2\sum_{k=1}^{K}\sigma_k^{-2}\mathbf{u}_k\mathbf{u}_k^H)\right]. \quad \text{(A.16)}$$

From Lemma 1, $\mathbb{E}[u_{ik}u_{jk}^*] = 0$ for $i \neq j$ and $\mathbb{E}[|u_{nk}|^2] = \frac{1}{N}$ for all n, k. So, (A.16) will be,

$$A_2 = \frac{\alpha^{(t)}}{\beta_{i(k)}K}\mathbb{E}\left[\sum_{k=1}^{K}\sigma_k^{-2}\sum_{n=1}^{N}G_{b_n}^2|u_{nk}|^2\right] = \frac{\alpha^{(t)}}{\beta_{i(k)}K}\mathbb{E}\left[\sum_{k=1}^{K}\sigma_k^{-2}\right]\left(\frac{1}{N}\sum_{n=1}^{N}G_{b_n}^2\right).$$

As mentioned earlier, for the inverse Wishart matrix $(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}})^{-1}$, we have $\mathbb{E}\left[\sum_{k=1}^{K}\frac{1}{\sigma_k^2}\right] =$ $\mathbb{E}\left[\mathrm{tr}\{(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}})^{-1}\}\right] = \frac{K}{N-K}$. Thus,

$$A_2 = \frac{1}{\beta_{i(k)}}\frac{\alpha^{(t)}}{N-K}g_2. \quad \text{(A.17)}$$

Similarly, A_3 can be found as,

$$A_3 = \frac{1}{\beta_{i(k)}}\frac{\alpha^{(t)}}{N-K}\left(\hat{c} - vg_2\right). \quad \text{(A.18)}$$

Finally, after substituting $\mathbf{G}^{(t)}$ from (4.7) in A_4, we have

$$A_4 = p_u\alpha^{(t)}\mathbb{E}\left[\left|\left(\left(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\right)^{-1}\tilde{\mathbf{H}}^H\mathbf{G}_b\tilde{\mathbf{H}}\right)_{i(k)i(k)}\right|^2\right]$$

$$=p_u\alpha^{(t)}\sum_{k_1,k_2}\sum_{k_1',k_2'}\mathbb{E}\left[\frac{\sigma_{k_2}\sigma_{k_2'}}{\sigma_{k_1}\sigma_{k_1'}}v_{i(k),k_1}v^*_{i(k),k_2}v^*_{i(k),k_1'}v_{i(k),k_2'}\mathbf{u}_{k_1}^H\mathbf{G}_b\mathbf{u}_{k_2}\mathbf{u}_{k_2'}^H\mathbf{G}_b\mathbf{u}_{k_1'}\right]. \tag{A.19}$$

Similar to the derivations for A_1 and according to Lemma 1, if at least one of k_1, k_2, k_1', k_2' is different from the others, the corresponding expectation term in (A.19) is 0. The remaining terms are considered in the following four cases.

1. If $k_1 = k_2$, $k_1' = k_2'$, and $k_1 \neq k_1'$, the sum of the corresponding terms in (A.19) can be calculated as follows.

$$d_1 = \frac{p_u\alpha^{(t)}(K-1)}{(K+1)}\left[\frac{\sum_{n=1}^N G_{b_n}^2}{N(N+1)} + \frac{\sum_{n_1=1}^N\sum_{n_2=1,\neq n_1}^N G_{b_{n_1}}G_{b_{n_2}}}{N^2-1}\right].$$

2. If $k_1 = k_1'$, $k_2 = k_2'$, and $k_1 \neq k_2$, the sum of the corresponding terms in (A.19) can be calculated as follows.

$$d_2 = \frac{p_u\alpha^{(t)}N(K-1)}{(K+1)(N-K)}\left[\frac{\sum_{n=1}^N G_{b_n}^2}{N(N+1)} - \frac{\sum_{n_1=1}^N\sum_{n_2=1,\neq n_1}^N G_{b_{n_1}}G_{b_{n_2}}}{N(N^2-1)}\right].$$

3. If $k_1 = k_2'$, $k_2 = k_1'$, and $k_1 \neq k_2$, each of the corresponding terms in (A.19) are 0.

4. If $k_1 = k_2 = k_1' = k_2'$, the sum of the corresponding terms in (A.19) can be calculated as follows.

$$d_3 = \frac{2p_u\alpha^{(t)}}{(K+1)}\left[\frac{\sum_{n=1}^N G_{b_n}^2 + \sum_{n_1=1}^N\sum_{n_2=1}^N G_{b_{n_1}}G_{b_{n_2}}}{N(N+1)}\right].$$

Hence, using the above results of d_1, d_2, and d_3 in (A.19), for $N \gg 1$,

$$\begin{aligned}A_4 =& \frac{p_u\alpha^{(t)}}{(K+1)(N-K)(N^2-1)} \times \Big[(N^2K+N^2-N-3NK+K+K^2)g_2 \\ &+ N(N^2K+N^2-NK^2-N-2NK+2K)g_1^2\Big] \\ \approx& p_u\frac{\alpha^{(t)}}{N-K}\left[\frac{Kg_2+(NK+N-K^2-2K)g_1^2}{K+1}\right].\end{aligned} \tag{A.20}$$

By using (A.15), (A.17), (A.18), (A.20), and ignoring lower order terms of N for

large N, i.e., we consider that $NK \gg K$ when $N \gg 1$, the average achievable rate result in Theorem 2 is obtained.

A.3 Proof of Theorem 3

In this section, we follow similar steps as Appendix A.1, while modifying the SVD in (A.3) as $\hat{\tilde{\mathbf{H}}} = \mathbf{U}\hat{\mathbf{\Sigma}}\mathbf{V}^{\mathbf{H}}$, where $\hat{\mathbf{\Sigma}} = \sqrt{1-\sigma_e^2}\mathbf{\Sigma}$. We have $P_R = \mathbb{E}\{\|\mathbf{r}_{\text{ICSI}}^{(t)}\|^2\} = \hat{c}_1 + \hat{c}_2 + \hat{c}_3$, where

$$\begin{aligned}
\hat{c}_1 &\triangleq p_u \mathbb{E}[\text{tr}\{\hat{\mathbf{G}}^{(t)}\mathbf{G}_b\mathbf{H}\mathbf{H}^H\mathbf{G}_b(\hat{\mathbf{G}}^{(t)})^H\}, \\
\hat{c}_2 &\triangleq \mathbb{E}[\text{tr}\{\hat{\mathbf{G}}^{(t)}\mathbf{G}_b^2(\hat{\mathbf{G}}^{(t)})^H\}], \\
\hat{c}_3 &\triangleq \mathbb{E}[\text{tr}\{\hat{\mathbf{G}}^{(t)}\mathbf{d}\mathbf{d}^H(\hat{\mathbf{G}}^{(t)})^H\}].
\end{aligned}$$

$\hat{c}_1$ can be written as $\hat{c}_1 = \hat{c}_{1.a} + \hat{c}_{1.b}$, where

$$\begin{aligned}
\hat{c}_{1.a} &\triangleq p_u \mathbb{E}[\text{tr}\{\hat{\mathbf{G}}^{(t)}\mathbf{G}_b\hat{\mathbf{H}}\hat{\mathbf{H}}^H\mathbf{G}_b(\hat{\mathbf{G}}^{(t)})^H\}, \\
\hat{c}_{1.b} &\triangleq p_u \mathbb{E}[\text{tr}\{\hat{\mathbf{G}}^{(t)}\mathbf{G}_b\Delta\mathbf{H}\Delta\mathbf{H}^H\mathbf{G}_b(\hat{\mathbf{G}}^{(t)})^H\}].
\end{aligned}$$

Following similar procedure as for c_1, and c_2 in the perfect CSI case, $\hat{c}_{1,a}$, and $\hat{c}_{1,b}$ are found as the following, respectively.

$$\begin{aligned}
\hat{c}_{1.a} &\approx \frac{p_u\hat{\alpha}^{(t)}g_1^2}{(N-K)(1-\sigma_e^2)}\sum_{i=1}^{K}\frac{1}{\beta_i}, \\
\hat{c}_{1.b} &\approx \frac{p_u\hat{\alpha}^{(t)}g_2(NK+N-2K-K^2)\beta_{\text{sum}}\sigma_e^2}{(N-K)^2(N-K-1)(K+1)(1-\sigma_e^2)^2}\sum_{m=1}^{K}\frac{1}{\beta_m\beta_{i(m)}}.
\end{aligned}$$

Further, following similar procedure as for c_2 and c_3 for the perfect CSI case, $\hat{c}_2$ and $\hat{c}_3$ are calculated as below, respectively.

$$\begin{aligned}
\hat{c}_2 &\approx \frac{\hat{\alpha}^{(t)}g_2(NK+N-2K-K^2)}{(N-K)^2(N-K-1)(K+1)(1-\sigma_e^2)^2}\sum_{m=1}^{K}\frac{1}{\beta_m\beta_{i(m)}}, \\
\hat{c}_3 &\approx \frac{\hat{\alpha}^{(t)}(\hat{c}-vg_2)(NK+N-2K-K^2)}{(N-K)^2(N-K-1)(K+1)(1-\sigma_e^2)^2}\sum_{m=1}^{K}\frac{1}{\beta_m\beta_{i(m)}}.
\end{aligned}$$

Therefore, after simplifications, and ignoring lower order terms of N for large N, i.e., we consider that $NK \gg K$ when $N \gg 1$, the ZF transmit power

coefficient for the imperfect CSI case will be as in (4.26).

A.4 Proof of Theorem 4

Similar to the perfect CSI case, for the average achievable rate we have:

$$R_{k,i(k)} \approx \log_2 \left(1 + \frac{\hat{A}_4}{\hat{A}_1 + \hat{A}_2 + \hat{A}_3 + 1} \right), \tag{A.21}$$

where the values of $\hat{A}_1, \hat{A}_2, \hat{A}_3$, and $\hat{A}_4$ are calculated as follows.

$$\begin{aligned}
\hat{A}_1 \triangleq & p_u \mathbb{E} \left[\sum_{j=1, j \neq i(k)}^{K} |\mathbf{h}_k^T \hat{\mathbf{G}}^{(t)} \mathbf{G}_b \mathbf{h}_j|^2 \right], \\
\hat{A}_2 \triangleq & \mathbb{E}[\|\mathbf{h}_k{}^T \hat{\mathbf{G}}^{(t)} \mathbf{G}_b\|^2], \\
\hat{A}_3 \triangleq & \mathbb{E}[\|\mathbf{h}_k{}^T \hat{\mathbf{G}}^{(t)} \mathbf{d}\|^2], \\
\hat{A}_4 \triangleq & p_u \mathbb{E}[|\mathbf{h}_k{}^T \hat{\mathbf{G}}^{(t)} \mathbf{G}_b \mathbf{h}_{i(k)}|^2].
\end{aligned}$$

$\hat{A}_1$ can be written as $\hat{A}_1 = \hat{A}_{1,a} + \hat{A}_{1,b} + \hat{A}_{1,c} + \hat{A}_{1,d}$, where

$$\begin{aligned}
\hat{A}_{1,a} \triangleq & \; p_u \mathbb{E} \left[\sum_{j=1, j \neq i(k)}^{K} |\hat{\mathbf{h}}_k^T \hat{\mathbf{G}}^{(t)} \mathbf{G}_b \hat{\mathbf{h}}_j|^2 \right], \\
\hat{A}_{1,b} \triangleq & \; p_u \mathbb{E} \left[\sum_{j=1, j \neq i(k)}^{K} |\Delta \mathbf{h}_k^T \hat{\mathbf{G}}^{(t)} \mathbf{G}_b \Delta \mathbf{h}_j|^2 \right], \\
\hat{A}_{1,c} \triangleq & \; p_u \mathbb{E} \left[\sum_{j=1, j \neq i(k)}^{K} |\hat{\mathbf{h}}_k^T \hat{\mathbf{G}}^{(t)} \mathbf{G}_b \Delta \mathbf{h}_j|^2 \right], \\
\hat{A}_{1,d} \triangleq & \; p_u \mathbb{E} \left[\sum_{j=1, j \neq i(k)}^{K} |\Delta \mathbf{h}_k^T \hat{\mathbf{G}}^{(t)} \mathbf{G}_b \hat{\mathbf{h}}_j|^2 \right].
\end{aligned}$$

Following the same steps as A_1 for the perfect CSI case,

$$\hat{A}_{1,a} \approx \frac{p_u \hat{\alpha}^{(t)}}{\beta_{i(k)}(N-K)} (g_2 - g_1^2) \beta_{\setminus i(k)}.$$

For $\hat{A}_{1,b}$, we have

$$\hat{A}_{1,b} \approx \frac{p_u \hat{\alpha}^{(t)} \beta_k \beta_{\setminus i(k)} \sigma_e^4 g_2 (NK + N - 2K - K^2)}{(N-K)^2 (N-K-1)(K+1)(1-\sigma_e^2)^2} \sum_{m=1}^{K} \frac{1}{\beta_m \beta_{i(m)}} + \frac{p_u \hat{\alpha}^{(t)} \beta_k^2 \sigma_e^4 (N g_1^2 - g_2)(K^2-2)}{(N-K)(N-K-1)K(K^2-1)(1-\sigma_e^2)^2} \sum_{m=1}^{K} \frac{1}{\beta_m \beta_{i(m)}},$$

where the first term is found using the approximation $\mathbb{E}[|\Delta \tilde{h}_{ik}|^4] \approx \mathbb{E}[|\Delta \tilde{h}_{ik}|^2]^2$, $\forall i$ and following similar steps to find $\hat{c}_2$. The second term is added for the case when $j = k$ which is found using Lemma 1. Note that for the special case when K is even and $t = K/2$ there will be a third term, as

$$\frac{p_u \hat{\alpha}^{(t)} \beta_k^2 \sigma_e^4 (N g_1^2 - g_2)}{(N-K)^2 (K^2-1)(1-\sigma_e^2)^2} \sum_{m=1}^{K} \frac{1}{\beta_m \beta_{i(m)}}.$$

Then, following the same steps as the steps to find A_2 for the perfect CSI case, we have

$$\begin{aligned} \hat{A}_{1,c} &= p_u \hat{\alpha}^{(t)} \sigma_e^2 \beta_{\setminus i(k)} \mathbb{E}\left[\left(\left(\hat{\mathbf{H}}^H \hat{\mathbf{H}} \right)^{-1} \hat{\mathbf{H}}^H \mathbf{G}_b^2 \hat{\mathbf{H}} \left(\hat{\mathbf{H}}^{\mathbf{H}} \hat{\mathbf{H}} \right)^{-1} \right)_{i(k),i(k)} \right] \\ &= \frac{p_u \hat{\alpha}^{(t)} \sigma_e^2 \beta_{\setminus i(k)} g_2}{(N-K)(1-\sigma_e^2)\beta_{i(k)}}. \end{aligned}$$

Finding the exact expression for $\hat{A}_{1,d}$ requires the 6th order statistics of the elements of Haar matrix, $\mathbf{V}$. Therefore, we use the approximation $\mathbf{G}_b \approx \frac{1}{N} \sum_{i=1}^{N} G_{b_i} \mathbf{I}_N$ to avoid the high computational complexity, and will have

$$\begin{aligned} \hat{A}_{1,d} &\approx p_u \hat{\alpha}^{(t)} g_1^2 \beta_k \sigma_e^2 \mathbb{E}\left[\sum_{j=1, j\neq i(k)}^{K} \mathbf{e}_j^T (\mathbf{p}^t)^T \left(\hat{\mathbf{H}}^T \hat{\mathbf{H}}^* \right)^{-1} \mathbf{P}^t \mathbf{e}_j \right] \\ &= \frac{p_u \hat{\alpha}^{(t)} g_1^2 \beta_k \sigma_e^2}{(N-K)(1-\sigma_e^2)} \sum_{j=1, j\neq k}^{K} \frac{1}{\beta_j}, \end{aligned}$$

which is found similar to c_1 for the perfect CSI case.

For $\hat{A}_2$ we have,

$$\begin{aligned}\hat{A}_2 =& \hat{A}_{2,a} + \hat{A}_{2,b},\\ \hat{A}_{2,a} \triangleq& \mathbb{E}[\|\hat{\mathbf{h}}_k^T \hat{\mathbf{G}}^{(t)} \mathbf{G}_b\|^2],\\ \hat{A}_{2,b} \triangleq& \mathbb{E}[\|\Delta \mathbf{h}_k{}^T \hat{\mathbf{G}}^{(t)} \mathbf{G}_b\|^2].\end{aligned}$$

Similar to the steps for A_2 in the perfect CSI case, $\hat{A}_{2,a}$ can be found, as

$$\hat{A}_{2,a} = \frac{\hat{\alpha}^{(t)} g_2}{(N-K)(1-\sigma_e^2)\beta_{i(k)}}.$$

Using the results for c_2 from the perfect CSI case, for $\hat{A}_{2,b}$ we have

$$\hat{A}_{2,b} \approx \frac{\beta_k \sigma_e^2 \hat{\alpha}^{(t)} g_2 (NK + N - 2K - K^2)}{(1-\sigma_e^2)^2 (N-K)^2 (N-K-1)(K+1)} \sum_{m=1}^{K} \frac{1}{\beta_m \beta_{i(m)}}.$$

Similarly,

$$\begin{aligned}\hat{A}_3 =& \hat{A}_{3,a} + \hat{A}_{3,b},\\ \hat{A}_{3,a} \triangleq& \mathbb{E}[\|\hat{\mathbf{h}}_k^T \hat{\mathbf{G}}^{(t)} \mathbf{d}\|^2],\\ \hat{A}_{3,b} \triangleq& \mathbb{E}[\|\Delta \mathbf{h}_k{}^T \hat{\mathbf{G}}^{(t)} \mathbf{d}\|^2].\end{aligned}$$

Which can be found as

$$\begin{aligned}\hat{A}_{3,a} &= \frac{\hat{\alpha}^{(t)} (\hat{c} - v g_2)}{(N-K)(1-\sigma_e^2)\beta_{i(k)}},\\ \hat{A}_{3,b} &\approx \frac{\beta_k \sigma_e^2 \hat{\alpha}^{(t)} (\hat{c} - v g_2)(NK + N - 2K - K^2)}{(1-\sigma_e^2)^2 (N-K)^2 (N-K-1)(K+1)} \sum_{m=1}^{K} \frac{1}{\beta_m \beta_{i(m)}}.\end{aligned}$$

Finally,

$$\begin{aligned}\hat{A}_4 &= \hat{A}_{4,a} + \hat{A}_{4,b} + \hat{A}_{4,c} + \hat{A}_{4,d},\\ \hat{A}_{4,a} &\triangleq p_u \mathbb{E}[|\hat{\mathbf{h}}_k^T \hat{\mathbf{G}}^{(t)} \mathbf{G}_b \hat{\mathbf{h}}_{i(k)}|^2],\\ \hat{A}_{4,b} &\triangleq p_u \mathbb{E}\left[|\Delta \mathbf{h}_k^T \hat{\mathbf{G}}^{(t)} \mathbf{G}_b \Delta \mathbf{h}_{i(k)}|^2\right],\\ \hat{A}_{4,c} &\triangleq p_u \mathbb{E}\left[|\hat{\mathbf{h}}_k^T \hat{\mathbf{G}}^{(t)} \mathbf{G}_b \Delta \mathbf{h}_{i(k)}|^2\right],\end{aligned}$$

$$\hat{A}_{4,d} \triangleq p_u \mathbb{E}\left[|\Delta \mathbf{h}_k^T \hat{\mathbf{G}}^{(t)} \mathbf{G}_b \hat{\mathbf{h}}_{i(k)}|^2\right].$$

$\hat{A}_{4,a}$ can be found similar to A_4 in the perfect CSI case, as

$$\hat{A}_{4,a} \approx \frac{p_u \hat{\alpha}^{(t)}}{N-K}\left[\frac{Kg_2 + (NK+N-K^2-2K)g_1^2}{K+1}\right].$$

$\hat{A}_{4,b}$ can be found similar to $\hat{A}_{1,b}$, as

$$\hat{A}_{4,b} \approx \frac{p_u \hat{\alpha}^{(t)} g_2 (NK+N-2K-K^2)\beta_k \beta_{i(k)} \sigma_e^4}{(N-K)^2(N-K-1)(K+1)(1-\sigma_e^2)^2} \sum_{m=1}^{K} \frac{1}{\beta_m \beta_{i(m)}}.$$

$\hat{A}_{4,c}$ can be found similar to $\hat{A}_{1,c}$, as

$$\hat{A}_{4,c} = \frac{p_u \hat{\alpha}^{(t)} \beta_{i(k)} \sigma_e^2 g_2}{(N-K)(1-\sigma_e^2)\beta_{i(k)}}.$$

For $\hat{A}_{4,d}$ similar to $\hat{A}_{1,d}$, in order to avoid excessive complexity, we use the approximation, $\mathbf{G}_b \approx \frac{1}{N}\sum_{i=1}^{N} G_{b_i} \mathbf{I}_N$, and will have

$$\begin{aligned}\hat{A}_{4,d} &\approx p_u \beta_k \sigma_e^2 \hat{\alpha}^{(t)} g_1^2 \mathbb{E}\left[\mathbf{e}_{i(k)}^T (\mathbf{p}^t)^T \left(\hat{\mathbf{H}}^T \hat{\mathbf{H}}^*\right)^{-1} \mathbf{P}^t \mathbf{e}_{i(k)}\right] \\ &= \frac{p_u \hat{\alpha}^{(t)} g_1^2 \beta_k \sigma_e^2}{(N-K)(1-\sigma_e^2)\beta_k}.\end{aligned}$$

Combining all the obtained components in (A.21), and ignoring lower order terms of N for large N, i.e., we consider that $NK \gg K$ when $N \gg 1$, the average achievable rate for the imperfect CSI case can be written as in Theorem 4.

Appendix B

Proofs for Chapter 5

B.1 Proof of Theorem 5

From (5.5), the harvested energy at the relay per unit time is,

$$E_{\mathrm{H}} = \eta(1-\rho)p_s\mathbb{E}\{\mathrm{tr}\{\mathbf{H}_S\mathbf{H}_S^H\}\}. \tag{B.1}$$

We use the following SVD for the channel matrix $\tilde{\mathbf{H}}_S$,

$$\tilde{\mathbf{H}}_S = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H,$$

where $\mathbf{U}$, $\mathbf{V}$, and $\boldsymbol{\Sigma}$ are $N \times K$, $K \times K$, and $K \times K$ matrices. Also, as mentioned in Appendix A, $\mathbf{U}$ and $\mathbf{V}$ are Haar matrices, containing the left and right singular vectors of $\tilde{\mathbf{H}}_S$, respectively. $\boldsymbol{\Sigma} = \mathrm{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_K\}$ contains the singular values of $\tilde{\mathbf{H}}_S$. Further, since the entries of $\tilde{\mathbf{H}}_S$ follow i.i.d. $\mathcal{CN}(0,1)$, $\mathbf{U}$, $\mathbf{V}$, and $\boldsymbol{\Sigma}$ are independent. Therefore,

$$\begin{aligned}
\mathbb{E}\{\mathrm{tr}\{\mathbf{H}_S\mathbf{H}_S^H\}\} &= \mathbb{E}\{\mathrm{tr}\{\tilde{\mathbf{H}}_S\mathbf{D}_S\tilde{\mathbf{H}}_S^H\}\} \\
&= \mathbb{E}\{\mathrm{tr}\{\boldsymbol{\Sigma}^2\mathbf{V}^H\mathbf{D}_S\mathbf{V}\}\} \\
&= \sum_{i_2=1}^{K}\mathbb{E}\{\sigma_{i_2}^2\}\sum_{i_1=1}^{K}\beta_{Si_1}(\theta_{\mathrm{tilt}})\mathbb{E}\{|v_{i_1i_2}|^2\}.
\end{aligned}$$

According to Lemma 1, for any i, j,

$$\mathbb{E}\{|v_{ij}|^2\} = \frac{1}{K}.$$

Further, for any arbitrary i, we have

$$\mathbb{E}\{\sigma_i^2\} = \frac{1}{K}\mathbb{E}\{\mathrm{tr}\{\tilde{\mathbf{H}}_S^H\tilde{\mathbf{H}}_S\}\} = N. \tag{B.2}$$

Therefore, (5.6) will be resulted.

B.2 Proof of Equation (5.9)

According to (5.8),

$$\bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MRC/MRT}} = \frac{|\mathbb{E}\{|\mathbf{h}_{Sk}|^2\}|^2}{\mathrm{Var}\{|\mathbf{h}_{Sk}|^2\} + \sum_{j\neq k}^{K}\mathbb{E}\{\mathbf{h}_{Sk}^H\mathbf{h}_{Sj}\mathbf{h}_{Sj}^H\mathbf{h}_{Sk}\} + \frac{\mathbb{E}\{|\mathbf{h}_{Sk}|^2\}\sigma_{\mathrm{R}}^2}{\rho p_s}}.$$

We have

$$\begin{aligned}\mathbb{E}\{|\mathbf{h}_{Sk}|^2\} &= N\beta_{Sk}(\theta_{\mathrm{tilt}}),\\ \mathrm{Var}\{|\mathbf{h}_{Sk}|^2\} &= \mathbb{E}\{|\mathbf{h}_{Sk}|^4\} - |\mathbb{E}\{|\mathbf{h}_{Sk}|^2\}|^2,\end{aligned} \tag{B.3}$$

and

$$\begin{aligned}\mathbb{E}\{|\mathbf{h}_{Sk}|^4\} &= \beta_{Sk}^2(\theta_{\mathrm{tilt}})\mathbb{E}\left\{|\tilde{\mathbf{h}}_{Sk}|^4\right\}\\ &= \beta_{Sk}^2(\theta_{\mathrm{tilt}})\mathbb{E}\left\{\left(\sum_{i=1}^{N}|\tilde{h}_{Sik}|^2\right)^2\right\}\\ &= \beta_{Sk}^2(\theta_{\mathrm{tilt}})\mathbb{E}\left\{\sum_{i=1}^{N}|\tilde{h}_{Sik}|^4 + \sum_{i\neq j}^{N}\sum_{j=1}^{N}|\tilde{h}_{Sik}|^2|\tilde{h}_{Sjk}|^2\right\}.\end{aligned}$$

For any i,k, we can write $\tilde{h}_{Sik} = \tilde{h}_{\mathrm{R}} + j\tilde{h}_{\mathrm{I}}$, where $\tilde{h}_{\mathrm{R}}$ and $\tilde{h}_{\mathrm{I}}$ are the real and imaginary components that are i.i.d. following $\mathcal{N}(0,1/2)$. Also,

$$\mathbb{E}\{\tilde{h}_{\mathrm{R}}^4\} = \mathbb{E}\{\tilde{h}_{\mathrm{I}}^4\} = 3/4.$$

Thus,

$$\begin{aligned}\mathbb{E}\left\{|\tilde{h}_{Sik}|^4\right\} &= \mathbb{E}\{\tilde{h}_{\mathrm{R}}^4 + \tilde{h}_{\mathrm{I}}^4 + 2\tilde{h}_{\mathrm{R}}^2\tilde{h}_{\mathrm{I}}^2\} = 2,\\ \mathrm{Var}\{|\mathbf{h}_{Sk}|^2\} &= N\beta_{Sk}^2(\theta_{\mathrm{tilt}}).\end{aligned} \tag{B.4}$$

Further, for any k,j,

$$\sum_{j\neq k}^{K}\mathbb{E}\left\{\mathbf{h}_{Sk}^H\mathbf{h}_{Sj}\mathbf{h}_{Sj}^H\mathbf{h}_{Sk}\right\} = \tag{B.5}$$

$$\beta_{Sk}(\theta_{\text{tilt}}) \sum_{j \neq k}^{K} \beta_{Sj}(\theta_{\text{tilt}}) \mathbb{E}\left\{ \tilde{\mathbf{h}}_{Sk}^H \tilde{\mathbf{h}}_{Sj} \tilde{\mathbf{h}}_{Sj}^H \tilde{\mathbf{h}}_{Sk} \right\}$$

and

$$\mathbb{E}\left\{ \tilde{\mathbf{h}}_{Sk}^H \tilde{\mathbf{h}}_{Sj} \tilde{\mathbf{h}}_{Sj}^H \tilde{\mathbf{h}}_{Sk} \right\} = \mathbb{E}\left\{ \tilde{\mathbf{h}}_{Sk}^H \tilde{\mathbf{h}}_{Sk} \right\} = N. \tag{B.6}$$

Hence, by using (B.3), (B.4), and (B.6) in $\bar{\gamma}_{k,\text{MAC}}^{\text{MRC/MRT}}$, (5.9) results.

B.3 Proof of Equation (5.21)

According to (5.20),

$$\bar{\gamma}_{k,\text{MAC}}^{\text{MF}} = \frac{|\mathbb{E}\{\|\mathbf{h}_{Sk}\|\}|^2}{\text{Var}\{\|\mathbf{h}_{Sk}\|\} + \sum_{j \neq k}^{K} \mathbb{E}\left\{ \left| \frac{\mathbf{h}_{Sk}^H \mathbf{h}_{Sj}}{\|\mathbf{h}_{Sk}\|} \right|^2 \right\} + \frac{\rho \sigma_{\text{R}}^2 + \sigma_{\text{R}}'^2}{\rho p_s}}$$

Note that $\|\mathbf{h}_{Sk}\|$ is the square root of the sum of squares of $2N$ independent random variables distributed according to $\mathcal{N}(0, \frac{\beta_{Sk}(\theta_{\text{tilt}})}{2})$. Therefore, $\|\mathbf{h}_{Sk}\|$ has a chi-distribution with $2N$ degrees-of-freedom, and

$$\mathbb{E}\{\|\mathbf{h}_{Sk}\|\} = \sqrt{\beta_{Sk}(\theta_{\text{tilt}})} \frac{\Gamma(N + \frac{1}{2})}{\Gamma(N)}, \tag{B.7}$$

$$\text{Var}\{\|\mathbf{h}_{Sk}\|\} = \beta_{Sk}(\theta_{\text{tilt}}) \left(N - \left(\frac{\Gamma(N + \frac{1}{2})}{\Gamma(N)} \right)^2 \right). \tag{B.8}$$

Further, for any arbitrary $j \neq k$

$$\begin{aligned} \mathbb{E}\left\{ \left| \frac{\mathbf{h}_{Sk}^H \mathbf{h}_{Sj}}{\|\mathbf{h}_{Sk}\|} \right|^2 \right\} &= \mathbb{E}\left\{ \frac{\mathbf{h}_{Sk}^H \mathbf{h}_{S_j} \mathbf{h}_{Sj}^H \mathbf{h}_{Sk}}{\|\mathbf{h}_{Sk}\|^2} \right\} \\ &\overset{(a)}{=} \mathbb{E}\left\{ \frac{\mathbf{h}_{Sk}^H \mathbb{E}\{\mathbf{h}_{Sj} \mathbf{h}_{Sj}^H\} \mathbf{h}_{Sk}}{\|\mathbf{h}_{Sk}\|^2} \right\} \\ &\overset{(b)}{=} \beta_{Sj}(\theta_{\text{tilt}}), \end{aligned} \tag{B.9}$$

where (a) is because $\mathbf{h}_{Sj}$ and $\mathbf{h}_{Sk}$ are independent and (b) is because

$$\mathbb{E}\{\mathbf{h}_{Sj} \mathbf{h}_{Sj}^H\} = \beta_{Sj}(\theta_{\text{tilt}}) \mathbf{I}_N.$$

Hence, by using (B.7), (B.8), and (B.9) in $\bar{\gamma}_{k,\mathrm{MAC}}^{\mathrm{MF}}$, (5.21) is obtained.