

COOPERATIVE COMMUNICATIONS FOR 5G WIRELESS SYSTEMS

by

Jayamuni Mario Shashindra Srimal Silva

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

© Jayamuni Mario Shashindra Srimal Silva, 2020

Abstract

The wireless industry plays a vital role in the economic development and prosperity of a country. Future wireless systems aim to achieve over 20 Gbps of data rates and less than 1 ms latencies for a large number services and applications. Thus, Fifth Generation (5G) wireless systems have been proposed to satisfy these demands. Massive multiple-input multiple-output (mMIMO), non-orthogonal multiple access (NOMA), and cognitive radio (CR) are some of the novel technologies envisioned for 5G systems. On the other hand, cooperative communications or the use of wireless relays have proven to improve the reliability and the energy efficiency of a wireless system. Thus, the primary objective of this thesis is to investigate the use of two-way relay networks (TWRNs), multi-way relay networks (MWRNs), and relay selection for wireless systems. First, the performance of multi-pair mMIMO TWRNs is analysed with channel imperfections. Second, relay selection in underlay CR mMIMO TWRNs is investigated. Third, a NOMA based MWRN scheme is proposed to significantly improve the sum rate. Finally, a machine learning based multiple relay selection scheme is proposed. This thesis demonstrates that the use of cooperative technologies significantly improves the data rate and energy efficiency of wireless systems.

Preface

This thesis is an original work conducted by Jayamuni Mario Shashindra Srimal Silva.

Chapter 3 of this thesis has been published as S. Silva, G. A. A. Baduge, M. Ardakani and C. Tellambura, "Performance Analysis of Massive MIMO Two-Way Relay Networks With Pilot Contamination, Imperfect CSI, and Antenna Correlation," in *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 4831-4842, June 2018. A part of the contribution of this chapter has been also published as S. Silva, G. Amarasuriya, C. Tellambura and M. Ardakani, "Massive MIMO two-way relay networks with channel imperfections," 2016 *IEEE International Conference on Communications (ICC)*, Kuala Lumpur, 2016, pp. 1-7.

Chapter 4 of this thesis has been published as S. Silva, M. Ardakani and C. Tellambura, "Interference Suppression and Energy Efficiency Improvement with Massive MIMO and Relay Selection in Cognitive Two-way Relay Networks," in *IEEE Transactions on Green Communications and Networking*, January 2020. A part of the contribution has also been published as S. Silva, M. Ardakani and C. Tellambura, "Relay Selection for Cognitive Massive MIMO Two-Way Relay Networks," 2017 *IEEE Wireless Communications and Networking Conference (WCNC)*, San Francisco, CA, 2017, pp. 1-6.

Chapter 5 of this thesis has been accepted for publication as S. Silva, G. A. A. Baduge, M. Ardakani and C. Tellambura, "NOMA-Aided Multi-Way Massive MIMO Relaying," *IEEE Transactions on Communications*. A part of the contribution has also been published as S. Silva, G. A. A. Baduge, M. Ardakani and C. Tellambura, "NOMA-Aided Multi-Way massive MIMO Relay Networks," 2019 *IEEE International Conference on Communications (ICC)*, Shanghai, China, 2019, pp. 1-6.

Dedicated to my beloved wife and son...

Acknowledgements

First of all, I express my sincere gratitude and respect to my supervisors, Prof. Chintha Tellambura and Prof. Masoud Ardakani. They provided encouragement, guidance, and invaluable expertise throughout my PhD program while giving me freedom to define and conduct my research. Also, special thank goes to Prof. Gayan Amarasuriya Aruma Baduge for his guidance and support.

I extend my sincere gratitude to the PhD supervisory committee member Prof. Yindi Jing for evaluating my thesis and providing constructive feedback and comments. I would also like to convey my special thanks to Prof. Lutz Lampe and Prof. Bruce Cockburn for their valuable feedback during my final PhD defense. I am thankful to Prof. Behrad Gholipour for chairing my final PhD defense.

I gratefully acknowledge Alberta Innovates Technology Futures (AITF) for offering me Alberta Innovates Doctoral Graduate Student Scholarship to conduct my PhD research.

My most heartfelt gratitude goes to my loving wife for her immeasurable support and encouragement throughout my PhD journey. I would also like to thank my parents and my sister for their support. Furthermore, I would like to thank former and current members of the Wireless Communications lab W5-070 who helped me during my PhD studies.

Table of Contents

1	Introduction	1
1.1	Fifth Generation (5G) wireless systems	1
1.2	Overview	3
1.2.1	Massive multiple-input multiple-output (mMIMO) Systems	3
1.2.2	Cooperative Communications	3
1.2.3	Cooperative Relays	4
1.2.4	Cognitive Radio (CR) Systems	6
1.2.5	Non-orthogonal multiple access (NOMA)	7
1.3	Motivations and and Literature Review	7
1.3.1	Performance analysis of multi-pair mMIMO two-way relay networks (TWRNs)	8
1.3.2	Relay selection in mMIMO cognitive TWRNs	9
1.3.3	NOMA-aided mMIMO MWRNs	10
1.3.4	Machine learning for multiple relay selection	11
1.4	Significance of this thesis	11
1.5	Outline and Contributions	12
1.5.1	Novel contributions of this thesis	12
2	Background	14
2.1	Massive MIMO	14
2.1.1	Basic system model	14
2.1.2	Channel Estimation and Pilot Contamination	15
2.1.3	Downlink signal-to-interference-plus-noise ratio (SINR) values	16
2.1.4	SINR with perfect channel-state information (CSI)	17
2.1.5	Cell-free mMIMO	17
2.2	NOMA Systems	18

2.2.1	Uplink NOMA transmission	18
2.2.2	Downlink NOMA transmission	19
2.2.3	Comparison with orthogonal multiple access (OMA) methods	20
2.2.4	NOMA with multiple users	21
2.3	Conclusion	22
3	Performance Analysis of mMIMO TWRNs with Channel Imperfections	23
3.1	Introduction	23
3.2	System, channel, and signal model	25
3.2.1	System and channel model	25
3.2.2	Channel estimation	25
3.2.3	Signal model	26
3.2.4	Calculation of value $\hat{\beta}_l$	27
3.2.5	Overall sum rate of the system	28
3.2.6	Energy efficiency of the system	28
3.3	Asymptotic performance analysis	28
3.3.1	Transmit power scaling at the user nodes	29
3.3.2	Transmit power scaling at the relay	30
3.3.3	Transmit power scaling at the user nodes and relay	31
3.4	The optimal pilot sequence length	32
3.5	Analysis for a finite number of antennas	33
3.6	Asymptotic analysis for relay-antenna correlation	35
3.7	Simulation Results	37
3.8	Conclusion	41
4	Relay selection in mMIMO cognitive TWRNs	44
4.1	Introduction	44
4.2	System, channel, and signal model	47
4.2.1	System and channel model	47
4.2.2	Signal model	48
4.2.3	CSI requirements	49
4.2.4	Effect on the primary user	50
4.2.5	Outage of the secondary system	50
4.2.6	Exact conditional end-to-end SINR	51

4.2.7	Sum rate analysis when R_k is selected	52
4.3	Asymptotic analysis	53
4.3.1	Requirement of power scaling	53
4.3.2	Power scaling at the end nodes and the relay (PS-1)	53
4.3.3	Power scaling at the end nodes - (PS-2)	55
4.4	Optimal power allocation	56
4.4.1	Power scaling at the end nodes and the relay (PS-1)	57
4.4.2	Power scaling at the end nodes (PS-2)	57
4.5	Relay selection with optimum values	58
4.5.1	Power scaling at the end nodes and the relay (PS-1)	58
4.5.2	Power scaling at the end nodes (PS-2)	59
4.6	Energy efficiency analysis	61
4.7	Numerical results	62
4.8	Conclusions	69
5	NOMA-Aided Multi-Way mMIMO Relaying	73
5.1	Introduction	73
5.2	System, channel, and signal model	76
5.2.1	System and channel model	76
5.2.2	Channel estimation	77
5.2.3	Signal model	78
5.2.4	Imperfect successive interference cancellation (SIC) decoding at the user nodes	79
5.3	Achievable sum rate analysis	79
5.4	Asymptotic sum rate analysis	81
5.5	Energy efficiency of the system	83
5.6	Comparison with other MWRN operations	83
5.7	Design of power allocation matrix	84
5.8	Numerical results	86
5.9	Conclusion	93
6	Machine Learning for Multiple Relay Selection	94
6.1	Introduction	94
6.2	System, channel, and signal model	95
6.3	Deep neural network (DNN) solution	97

6.3.1	DNN structure	97
6.3.2	Dataset	98
6.3.3	Training and validation	98
6.3.4	Obtaining final results	98
6.4	Numerical results	99
6.4.1	Diversity order	101
6.4.2	Complexity analysis	102
6.5	Conclusion	102
7	Conclusion and Future Research Directions	104
7.1	Summary of contributions and conclusions	104
7.2	Future research directions	106
	Bibliography	107
	Appendix A Appendices for Chapter 3	123
A.1	Proof of limits	123
A.2	Proof of limits for power scaling at the user nodes	124
A.3	Proof of limits for power scaling at the relay nodes	125
	Appendix B Appendices for Chapter 4	126
B.1	Asymptotic SINR with power scaling at all nodes	126
	Appendix C Appendices for Chapter 5	128
C.1	Expected value results for Imperfect CSI	128
C.1.1	Derivation of $M_{k,m}$	128
C.1.2	Derivation of $N_{k,m}$	129
C.1.3	Derivation of Q_k	130
C.2	Derivation of Ψ for imperfect CSI	130
C.2.1	Computation of L_1	130
C.2.2	Computation of L_2	130
C.3	Important Results	131
C.4	Asymptotic SINR with power scaling	131

List of Symbols

Elementary & Special Functions

Notation	Definition
$\exp(\cdot)$	exponential function
$\log(\cdot)$	natural logarithm
$\log_2(\cdot)$	logarithm to base 2
$\Gamma(z)$	gamma function [1, Eqn. (8.310.1)]
$\gamma(\alpha, z)$	lower incomplete gamma function [1, Eqn. (8.350.1)]
$\Gamma(\alpha, z)$	upper incomplete gamma function [1, Eqn. (8.350.2)]
$E_1(z)$	exponential integral function [1, Eqn. (8.211)]

Probability & Statistics

Let X and A be a random variable and an arbitrary event, respectively.

Notation	Definition
$\mathbb{E}_X[\cdot]$	expected value with respect to X ; $\mathbb{E}[\cdot]$, if X is implied
$f_X(\cdot)$	probability density function (PDF) of X
$F_X(\cdot)$	cumulative distribution function (CDF) of X
$\bar{F}_X(\cdot)$	complementary cumulative distribution function (CCDF) of X
$\mathbb{V}_X[\cdot]$	variance of X ; $\mathbb{V}[\cdot]$, if X is implied.
$\Pr(A)$	probability of A
$X \sim \mathcal{CN}(\mu, \sigma^2)$	X is complex normal distributed with mean μ and variance σ^2

Vectors and Matrices

Let $\mathbf{a} \in \mathbb{C}^{1 \times N}$ and $\mathbf{A} \in \mathbb{C}^{M \times N}$ denote an $1 \times N$ complex vector and an $M \times N$ complex matrix, respectively.

Notation	Definition/ interpretation
$[\mathbf{a}]_i$ or \mathbf{a}_i	i th element of \mathbf{a}
$[\mathbf{A}]_{i,j}$ or $\mathbf{A}_{i,j}$	i th element on the j th column of \mathbf{A}
$[\mathbf{A}]_k$	k th diagonal element of \mathbf{A}
$\mathbf{A} = \text{diag}(a_k)$	\mathbf{A} is rectangular diagonal; a_1 through a_n are the non-zero diagonal elements
\mathbf{A}^{-1}	inverse of \mathbf{A} (for $m = n$)

\mathbf{A}^*	conjugate of \mathbf{A}
\mathbf{A}^H	Hermitian transpose of \mathbf{A}
\mathbf{A}^T	transpose of \mathbf{A}
$\det(\mathbf{A})$	determinant of \mathbf{A}
$\text{Tr}(\mathbf{A})$	trace of \mathbf{A} for $M = N$
\mathbf{I}_N	identity matrix of rank n
$\mathbf{O}_{M \times N}$	$(M \times N)$ zero matrix
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of \mathbf{A} and \mathbf{B}

List of Abbreviations

Abbreviation	Definition
4G	Fourth generation cellular networks
5G	Fifth Generation
AF	amplify-and-forward
AP	access point
AWGN	additive white Gaussian noise
BS	base station
CCDF	complementary cumulative distribution function
CCI	co-channel interference
CDF	cumulative distribution function
CPU	central processing unit
CR	cognitive radio
CSI	channel-state information
DF	decode-and-forward
DNN	deep neural network
eMBB	enhanced mobile broadband
i.i.d.	independent and identically distributed
IoT	Internet of Things
LTE-A	Long Term Evolution-Advanced
MIMO	multiple-input multiple-output
mMIMO	massive multiple-input multiple-output
mMTC	massive machine-type communication
MRC	maximal ratio combining
MRT	maximal ratio transmission
MWRN	multi-way relay network
NOMA	non-orthogonal multiple access
NR	new radio

Abbreviation	Definition
OMA	orthogonal multiple access
OWRN	one-way relay network
PDF	probability density function
PU	primary user
SIC	successive interference cancellation
SINR	signal-to-interference-plus-noise ratio
SNR	signal-to-noise ratio
SU	secondary user
TWRN	two-way relay network
UR-LLC	ultra-reliable, low-latency communication
ZF	zero forcing

Chapter 1

Introduction

Wireless communication plays a vital role in daily life. Similar to food, water, shelter, and clothing, access to the Internet has become a necessity for most people. Social media, file sharing, and online gaming have become the most popular applications of the wireless communication devices [2]. Global mobile data traffic grew by 71% during the year 2017 and is expected to increase sevenfold by the year 2022 [3]. Furthermore, the number of mobile-connected devices per capita will reach 1.5 by year 2022. The highest increase in the usage will be from the developing countries while the industry in developed countries is somewhat saturated [2]. At the same time, with the emergence of the Internet of Things (IoT), the total wireless connections globally is expected to increase by three-fold to 25 billion.

Mobile cellular industry has a huge impact on the global economy. In the year 2017, it generated \$3.6 trillion in economic value, which corresponds to 4.5% of the global GDP [2]. It supported 29 million jobs globally and generated over \$500 billion in revenue for the public sector through taxation and spectrum auctions [2]. Its impact is not limited to the economy, but has become a necessary infrastructure for the development of a country. Moreover, it plays a critical role in natural disaster relief such as recently in Nepal and USA [4]. Furthermore, it plays a vital role in agriculture, health-care, and finance sector to achieve sustainability goals [2] in developing countries.

1.1 Fifth Generation (5G) wireless systems

To cater for the aforementioned growth, 5G and beyond 5G wireless technologies are being developed [5]. 5G will be significantly faster than its predecessor 4G, enabling higher productivity across all 5G capable devices [6,7]. However, 5G must overcome

overcrowding and/or scarcity of the wireless spectrum. Unlike the previous wireless generations, 5G must support a wide variety of wireless applications that may not even be imagined yet [8,9].

The 5G physical layer was introduced as 5G-new radio (NR) [5]. Early standardization of 5G systems was done in 3GPP Release-15 [10] and will continue on 3GPP Release-16 [11]. According to International Mobile Telecommunications-2020 (IMT-2020) standards [5], 5G NR is expected to provide 20 Gbps peak data rates, with 1 ms of latency for 10^6 devices/ km^2 density with 100 times higher energy efficiencies and 3 times higher spectrum efficiencies than Fourth generation cellular networks (4G). Although 5G wireless must cater for a large variety of use-cases, three major categories have been identified as follows [5]:

1. Enhanced mobile broadband (eMBB)

This is the evolution of current 4G-LTE systems [12]. This service provides high data rates to enable various data consuming applications and must continuously provide these data rates even highly-mobile users. As an example, 5G NR must support uninterrupted connectivity for passengers in high speed trains. Thus, eMBB can be seen as the first phase of 5G. And Phase 2 will go beyond eMBB services to more transformative ultra-reliable, low-latency communication (UR-LLC) and massive machine-type communication (mMTC) applications.

2. mMTC

This service allows for different devices such as sensors and actuators to communicate without human intervention. These will include IoT applications, smart cities and homes and more. This service has the ability to provide connectivity to a large volume of devices which transmit a large number of small data packets [5,12].

3. UR-LLC

This use-case focuses on applications which requires higher reliability with a very low latency. Commercial health monitoring systems and self driving vehicles are some examples. The reliability required by these types of applications are in the order of 99.999% [13].

One major component of mMTC is the use of IoT. IoT devices can be classified

as resource-constrained and resource-rich devices [14]. However, most of IoT devices are considered to be low-powered resource-constrained devices with low data rate requirements and simple hardware capabilities [15, 16].

1.2 Overview

1.2.1 Massive multiple-input multiple-output (mMIMO) Systems

Massive MIMO systems are identified as a critical component of 5G wireless. They, also known as Large Scale Antenna Systems (LSAS) [17], are an extension of multi-user (MU) multiple-input multiple-output (MIMO). They are characterized by the use a large number of antennas at the base station (BS). This massive degree of spatial diversity enables the BS to use linear processing techniques to create almost orthogonal channels between users and it [18]. This factor results in high spectral and energy efficiencies [6, 7]. However, there are several performance limiting factors. The main one is pilot contamination, which is the residual interference caused by the reuse of non-orthogonal pilot sequences [17, 19]. Others include co-channel interference (CCI), imperfect channel-state information (CSI) and antenna correlation [6].

Massive MIMO, considered to be the next revolution of wireless systems, has lived up to these expectations. For instance, mMIMO base stations have been successfully implemented commercially. As an example, Sprint wireless (Texas) has obtained $3.4\times$ and $8.9\times$ gains in downlink and uplink sum rates in a LTE network by using a 64-antenna BS instead of a 8-antenna BS [20]. High data rates for thousands of fans were provided by using mMIMO bases stations in Russia during the FIFA 2018 World Cup [21]. Furthermore, 95% of BS solutions of Huawei are mMIMO-enabled [22]. Thus, mMIMO has become the cornerstone for 5G and other future wireless networks.

1.2.2 Cooperative Communications

Cooperative Communication is the utilization of neighbouring wireless nodes as helpers to support the communication between different nodes [23]. This means that wireless nodes work together to achieve different performance gains [24]. For instance, several single-antenna mobiles share their antennas and generate a virtual MIMO system. As a result, effective quality of service of the network can be improved. Some benefits of cooperation includes interference management, energy

savings, improved system throughput, seamless service provision, operational cost reductions, and efficient utilization of resources [24–26]. Thus, cooperative communications play a significant role in 5G systems.

Cooperative wireless communications can be implemented in various ways. 1. Using wireless nodes as repeaters (e.g., relays) to assist the communication between other wireless nodes, 2. Opportunistically exploiting resources allocated to other users without affecting their transmissions (e.g., cognitive radio), 3. Intentionally allowing interference to achieve a higher overall performance (e.g., non-orthogonal multiple access (NOMA)).

Despite the improved performance, cooperative communications faces several implementation challenges. Timing of cooperative transmissions, managing the interference, optimal selection of cooperative nodes, and incentives for cooperation are some of the challenges [24]. In this thesis, the following cooperative technologies are analyzed.

1.2.3 Cooperative Relays

Cooperative relays increase the coverage [27] by acting as intermediate repeaters between the transmitter and receiver. Since the use of relays results in an increase in the overall quality of service of the network, they have already been included in wireless standards such as IEEE 802.16 J [28] and Long Term Evolution-Advanced (LTE-A) [29]. In 5G systems, relay systems are expected to play a vital role [30], especially in IoT and mmWave systems.

Cooperative relay nodes can be categorized depending on the mode of their operation [26,31]. Amplify-and-forward (AF) relays amplify the received signal and forward it to the destination node, while decode-and-forward (DF) relays will decode the received data, encode it again and retransmit to the destination. The problem with AF relays is that they amplify the noise in the received signal and the amplified noise ends up arriving at the destination. That is not a problem in DF relays, but they can make decoding errors, which will be propagated to the destination node.

One-way relay networks (OWRNs)

These employ relays that are uni-directional at a given time. Thus, full mutual data exchange between two end nodes, say, S_1 and S_2 , requires four time slots. In the first two time slots, data from S_1 is sent to S_2 while the next two time slots are

used for the data transfer from S_2 to S_1 . However, note that only two time slots are required for the full data exchange if S_1 and S_2 have a direct communication link. Thus, this relaying approach results in an spectral efficiency loss of 100%.

Two-way relay networks (TWRNs)

TWRNs utilizes interference cancellation/physical layer network coding and require only two time slots to transfer data between two end nodes [32, 33]. In the first time slot, both S_1 and S_2 transmit their data to the relay. Next, the relay combines the received signal and broadcasts it in the next time slot. Thus, at the end of the second time slot, each end node has received the superimposed signal, which consists of the transmitted signal of its own and that of the other node. Each end node then performs self interference cancellation and recovers the data transmitted by the other node [32]. Thus, this relaying approach the same spectral efficiency as that of direct transmission. Thus, we see that the spectral efficiency of TWRNs is twice that of OWRNs. TWRNs can also be integrated with MIMO to further improve the achievable data rate and the reliability due to spatial multiplexing and diversity gains [34, 35].

Multi-way relay networks (MWRNs)

MWRNs are a generalization of both OWRNS and TWRNS and enable data exchange among three or more users (or nodes) [36–38]. They are suitable for IoT applications where a large number of devices mutually exchange their data [39]. Several time slots are utilized for the data transfer between the users. With the use of spatial multiplexing, for K users, K time slots are used for the data transfer [40].

Relay Selection

Several nodes may be available to relay signals between a sender and a destination nodes. Thus, one or more nodes from this pool of nodes can be selected to act as relays [41]. This can be single relay selection where only a single node is selected to act as a relay or multiple relay selection where several nodes are selected to support the transmission [42]. Single relay selection schemes include best relay selection, nearest neighbor selection and best worse channel selection [43, 44].

Selecting multiple relays will increase the complexity of the system as well as the complexity of the relay selection procedure. System complexity arises due to

the phase adjustments required at the sender to enable the coherent addition of multiple received signals at the destination. Relay selection complexity is due to the exponential number of possibilities when selecting multiple relays (if the number of potential relays are L , there are $2^L - 1$ possible ways to select multiple relays). In [43], relay ordering is presented for multiple relay selection.

1.2.4 Cognitive Radio (CR) Systems

Spectrum availability is a critical problem faced by wireless engineers [45]. Wireless spectrum is expensive and allocated only through regulatory bodies. For example, spectrum chunks at 700 MHz, 2.3 GHz, and 2.5 GHz for different Canadian provinces were auctioned for \$43 million in 2019 [46]. However, experimental results [47] demonstrate that some of the allocated spectrum is not being utilized due to outdated applications such as over-the-air TV in some locations. Such temporally unused allocated spectrum bands are called spectrum holes. They exist in different geographical locations and temporally when primary users are not transmitting. Cognitive radio (CR) systems attempt to utilize these spectrum holes cooperatively [45].

The licensed users of the spectrum are identified as primary users (PUs) and CR users are called secondary users (SUs). Based on the mode of operation, CRs are classified as follows [48].

1. **Underlay:** In this mode, the SUs can simultaneously access the PU spectrum provided that the interference on the PUs is below a certain threshold. The SU have the responsibility to meet this threshold by adjusting their transmit powers, and if they are unable to do so, they should stop their transmission.
2. **Overlay:** In this mode, PUs share knowledge of their signal codebooks and messages with the SUs. The SUs will assist in the data transfer between PUs. When the PUs are inactive, SUs can transmit their own data among each other.
3. **Interweave** In this mode, the SUs can only use the PU spectrum when no PUs are transmitting. Thus, before each transmission, SUs have to sense the spectrum to make sure that there are no active PUs in the vicinity.

With the use of a large number of devices in 5G, the efficient use of spectrum [9] is vital and thus, cognitive radio is one of the exciting technologies. An example

are IoT devices, which can reuse the licensed spectrum that has been allocated to different users when they are not actively transmitting. Furthermore, CRs are already included in different wireless standards such as IEEE 802.11.af [49], which highlights the use of Wi-Fi in VHF and UHF television bands.

1.2.5 NOMA

Historically, multiple access in wireless systems is implemented with orthogonal channels either in the time, frequency, code, or spatial domain and non-orthogonality caused by channel and system imperfections was considered as a drawback. In contrast, NOMA [50] is the use of non-orthogonal channels for multiple access, where the interference from other users are either decoded via cooperation or treated as noise. NOMA offers improved spectral efficiencies, reduced latency, and massive connectivity [51–55].

NOMA can be categorized into power-domain NOMA and code-domain NOMA [56–63]. In power-domain NOMA, different power levels are used to differentiate the signals transmitted by different users. The transmitter superimposes the signal and the receiver uses successive interference cancellation (SIC) to decode the data. NOMA has been identified as one of the main enabling components of 5G wireless systems [50,57,61,64] especially for mMTC applications due to its ability to provide connectivity for a higher number of users compared to orthogonal multiple access schemes. However, the use of NOMA for eMBB applications to provide higher sum rates has also been considered recently [58].

Some challenges faced by NOMA are imperfect CSI, imperfect SIC, power allocation, and user ordering methods [58]. Specifically, due to imperfect SIC, the errors during each stage of the iterative decoding process will propagate to the following stages. The end result can thus be a significant degradation of the overall error performance. Furthermore, the user pairing in NOMA requires considering an exponentially high number of possible orderings [58].

1.3 Motivations and and Literature Review

The aforementioned technologies have a significant potential to improve the performance of 5G systems. Specifically, TWRNs, MWRNs, relay selection, CR, and NOMA combined with mMIMO will reduce the inter-user interference and provide higher sum rates. However, the use of these technologies gives rise to a myriad of

challenges. Thus, the main objectives of this thesis are listed as follows.

1.3.1 Performance analysis of multi-pair mMIMO TWRNs

Massive MIMO TWRNs enable data exchange among multiple pairs of users [65, 66] with a two-fold increase in the achievable data rate compared to OWRNs [32]. Multiple TWRNs can accommodate more users spread throughout over different geographical locations. Massive MIMO TWRNs have a wide range of potential applications. For example, it can be used as a heterogeneous wireless entity in an existing cellular network for reducing the workload of the BS [67] by bypassing the BS. Service providers can thus use TWRNs to improve the data throughput without making drastic changes to their existing infrastructure. Another potential application scenario is IoT, which connects multiple wireless devices and sensors [14, 68]. For example, a cooling system might require data from temperature sensors while a security system might require data from the motion sensors. This scenario fits the model of a multi-pair relay network with a central relay node. Moreover, when multiple IoT networks coexist, CCI can be a significant impairment. Furthermore, these IoT devices often rely on battery power and thus, the energy efficiency is critical.

In [69], the asymptotic performance of multi-pair OWRNs with very large relay antenna arrays is investigated without CCI and for perfect CSI and asymptotic signal-to-interference-plus-noise ratio (SINR) and sum rate expressions are derived. In [70, 71], the achievable asymptotic sum rates for dual-hop OWRNs are derived when the relay or/and destination are mMIMO enabled. In [72], the sum rate of a MIMO TWRN has been analyzed under zero forcing (ZF) beamforming at the relay or user nodes and closed-form results and approximations for sum rate are obtained. Furthermore, in [73], the channel aging effects of multi-cell MWRNs with mMIMO are investigated. In [74], the asymptotic achievable sum rates are derived for mMIMO MWRNs with perfect CSI. In [65], multi-pair TWRNs with mMIMO are investigated by employing linear precoders and detectors, where again, the perfect CSI and CCI/pilot contamination free scenario is assumed. To this end, this considers three transmit power scaling laws. Several recent publications have analyzed multi-pair mMIMO TWRNs. In [75] and [76], multi-pair mMIMO TWRNs have been analyzed for maximal ratio combining (MRC)/maximal ratio transmission (MRT) beamforming. While [75] assumes perfect CSI, [76] analyzes the system under imperfect CSI scenario. Furthermore, [77] analyzes multi-pair

mMIMO TWRN with imperfect CSI under ZF beamforming. In [78], a full-duplex multi-pair mMIMO system is analyzed under imperfect CSI with ZF beamforming. In [79], a cognitive mMIMO TWRN has been analyzed with perfect CSI.

However, none of these work analyze the system under CCI, imperfect CSI, and pilot contamination in the context of two-way multi-pair mMIMO relaying. Furthermore, also lacking are closed-form solutions for sum rate and signal-to-noise ratio (SNR), when the relay has a finite number of antennas.

1.3.2 Relay selection in mMIMO cognitive TWRNs

In 5G wireless systems, the two design goals are improving the energy efficiency (the number of bits that can be sent per unit energy consumption [80, 81]) and the spectral efficiency (the number of bits that can be transmitted per second in a unit of bandwidth). The use of mMIMO TWRNs in an underlay CR environment can achieve both of these goals. The spectral and energy efficiency may further be improved by using relay selection [41, 42, 82, 83].

However, when secondary TWRNs operate over the same primary spectrum simultaneously with primary users, several interference problems will occur. The main one is that the interference on the PUs occur due to the two end nodes transmitting simultaneously during the first time slot. To reduce it, one option is to coordinate the transmissions of these two end nodes to adjust their transmit powers in unison. Such coordinated transmissions are unrealistic for user nodes, and thus TWRNs, despite their spectral efficiency advantage, have not been considered for cognitive radio networks.

Cognitive relays have been analyzed in [84–87]. Specifically, [84] analyzes a system where the primary and the secondary systems mutually coordinate to enable relay assisted communications and [85] derives detection and outage probabilities in the overlay mode. Also, [86] analyzes an underlay CR system with filter-and-forward relays. Overlay systems with multiple relays have been analyzed in [88]. Furthermore, relay selection and optimal power allocation is analyzed for single antenna nodes [89]. In [90], cognitive mMIMO systems are analyzed for power allocation under pilot contamination. Furthermore, in [91], the performance of a MIMO secondary network under a MIMO primary network has been analyzed, with asymptotic results when the number of antennas at PU increases to infinity. However, this work does not analyze the effect of multiple relays in the secondary

system. More recently, underlay mMIMO system have been analyzed in the presence of multiple primary MIMO nodes in [92]. This work analyses a secondary network with single-antenna users and a mMIMO BS. This work shows that a mMIMO PU can fully mitigate the co channel interference from the secondary network. However, none of these work focuses on relay selection for secondary mMIMO. Relay selection for cognitive full duplex one-way relay networks has been analyzed in [93] and relay selection for a full duplex energy harvesting system is analyzed in [94]. However, these result cannot be extended to TWRNs because of the interference constraints during the simultaneous data transmission of two nodes.

1.3.3 NOMA-aided mMIMO MWRNs

MWRNs enable data exchange between multiple users (or nodes) [36–38] and are suitable for IoT applications where a large number of devices mutually exchange their data [39]. Achievable information rates of the MWRN channel were first developed in [36]. In this network, multiple users first transmit their messages to the relay and then the relay will transmit signals over multiple time slots to enable message exchange among the users. In full data exchange, each user has a common message to all the others. There are no direct links among the users due to propagation impairments such as large-scale fading and/or shadowing effects. The applications of MWRNs include conference calls via a BS, wireless sensor networks, and satellite communications [36] and will increase with the emergence of IoT for the next-generation wireless systems [39]. Thus, beamforming methods [95, 96], relay selection schemes [97], and coding schemes [98] have been developed recently for MWRNs.

The integration of mMIMO with MWRN is an important topic [40, 99, 100]. For instance, the MWRN scheme in [38] is used on a multi-cell mMIMO system to quantify the achievable rates in the presence of imperfect CSI [40]. For the same system of [40], an efficient transmit power allocation scheme has been proposed in [99]. However, the number of time slots required for the full data exchange in a typical MWRN is equal to the number of users K [38, 40, 99] or more recently to $K/2$ [100]. As the data rate per user is the scaled version of total data rate by $\frac{1}{k}$, the data rate per user tends to saturate or reduces as the number of users is increased. A MWRN scheme where full data exchange is completed in 2 time slots, or at-least in a constant number of time slots independent of K , is not available in

the literature.

1.3.4 Machine learning for multiple relay selection

As mentioned in Section 1.2.3, multiple relay selection introduces several additional challenges compared to single relay selection. System complexity arises due to the phase adjustments required at the transmitter to enable the coherent addition of multiple received signals at the receiver. Also, if the number of potential relays is L , there are $2^L - 1$ possible ways to select multiple relays. Thus, the relay selection complexity is increased due to the exponential number of possibilities.

The work in [43] uses the relay ordering proposed in [101] to design a linear complexity relay selection scheme. Furthermore, it proposes a quadratic complexity multiple relay selection scheme. Although the performance of the quadratic complexity relay selection achieves almost the same performance as the optimum relay selection, the performance of the linear complexity relay selection method is lower. Thus, designing a multiple relay selection which achieves the performance of optimum method with a linear complexity with respect to the number of relays is a challenging problem.

1.4 Significance of this thesis

As mentioned earlier, 5G networks must provide higher sum rates with higher energy efficiencies and, thus, mMIMO, relay systems and NOMA will play a vital role in 5G networks. This thesis address the use of cooperative communications in 5G systems. Specifically, the work done in mMIMO TWRNs with channel imperfections will help the implementation of mMIMO systems by providing a closed form solutions for performance metrics. This will help the designers to decide on various system parameters such as number of antennas and the time used for pilot training. Furthermore, the use of cognitive radio will enable higher data rates by freeing up the spectrum that is underutilized in wireless systems. The use of the CR technologies with mMIMO and TWRNs combined with relay selection will enable future systems to achieve the required spectral and energy efficiencies. The work completed on this thesis on the use of NOMA for MWRN will allow wireless systems to exchange their signal within a fixed time period and thus will provide higher sum rates. Furthermore, it will be useful in mMTC communications as well. The last work done on this thesis on the use of machine learning for the relay selection will

be useful in systems where a lot of periodic data will be used in the system.

1.5 Outline and Contributions

The outline of this thesis is as follows. Chapter 1 provides the introduction and highlights the motivation, contributions, and the impact of the completed work. Chapter 2 reviews the basics about the topics covered in the thesis. The novel contributions of the thesis are provided in Chapters 3 to 6 and highlighted next. Chapter 7 provides the conclusion and the discusses possible future research based on this thesis.

1.5.1 Novel contributions of this thesis

The major contributions of this thesis are highlighted as follows.

- In Chapter 3, a multi-cell TWRN consisting of single-antenna user nodes and AF relay nodes with very large antenna arrays is considered. The total impact of CCI, imperfect CSI, pilot contamination, and the antenna correlation at the mMIMO node is investigated. Closed-form approximations for the sum rate are also derived. It is shown that the use of mMIMO mitigates the effect of CCI. However, the effects of imperfect CSI and pilot contamination degrade the performance even with a large antenna array. Yet, the use of mMIMO allows power scaling at the user nodes and relays and thus, even with channel imperfections, the benefits of employing a mMIMO enabled relay on transmit power savings are significant. Also, the analysis of antenna correlation shows that a large antenna array mitigates the resulting impact. Furthermore, the optimal pilot sequence length to maximize the sum rate of the system is obtained.
- In Chapter 4, a mMIMO TWRN with relay selection in CR setting is investigated. Specifically, optimal selection of an AF TWRN is considered to maximize the sum rate and to keep the interference on the PU below an interference threshold. Asymptotic SINR values for two scenarios are obtained: (1) the relays and the two end nodes use transmit power scaling and (2) only the end nodes use transmit power scaling. For these two cases, optimal power allocations subject to the PU interference constraints are derived. With these optimal power allocations, effect of relay selection on the outage, the sum rate,

and the energy efficiency of the network is analysed. For the first scenario, the outage can be reduced to zero with appropriate power allocation and the relay selection can be done offline. For the second scenario, outage will depend on the instantaneous channel state between the relays and the PU.

- In Chapter 5, a novel transmission scheme for MWRNs is proposed to enable the full data exchange among any number of users within two time slots by using NOMA. First, the users transmit their signals to the relay, which uses maximal ratio combining reception. Next, the relay transmits a superposition-coded signal for all users by using a maximal ratio transmission based precoder. Each user node then performs SIC decoding of data symbols of the other $K - 1$ user nodes. Worst-case Gaussian approximation is used to derive the overall sum rate. A relay power allocation matrix is designed to maximize the minimum sum rate achievable by each user and to maximize the fairness among the users. Furthermore, the effects of imperfect SIC and imperfect CSI on the overall sum rate are analyzed.
- In Chapter 6, machine learning techniques are used for multiple relay selection. Specifically, deep neural networks (DNNs) are utilized to perform multiple relay selection in a cooperative wireless system where L AF relays enable the data transmission between a source and a destination node. Relays either support the transmission with their full power or they do not transmit at all. This problem is modeled as a multi-class multi-label classification problem. Classification accuracy of up to 96% and sum rate accuracy of up to 99% is obtained compared to the optimum relay selection solution which has an exponential complexity with the number of relays. In contrast, the proposed method has a linear complexity with the number of relays and through simulations, it is shown that full diversity is achieved in 96% time by the proposed method.

Chapter 2

Background

This chapter presents necessary background for the thesis research. Specifically the basics of massive multiple-input multiple-output (mMIMO) and non-orthogonal multiple access (NOMA) systems are discussed.

2.1 Massive MIMO

2.1.1 Basic system model

A basic mMIMO system consists of L adjacent cells having K single antenna users in each cell. The users in the l th cell are denoted as $(U_{l,1}, \dots, U_{l,K})$, where the $U_{l,k}$ s are connected to the base station (BS) in each cell. The base stations are represented by B_l for $1 \leq l \leq L$ and each BS have M antennas where M is significantly higher than K . For instance K will be in the range of 10 to 100 while M will be in the range 100 to 1000. The performance of mMIMO will be analyzed as M goes to infinity.

The channel matrix from K users in the j th cell to the l th BS is represented as

$$\mathbf{G}_{jl} = \mathbf{F}_{jl} \mathbf{D}_{jl}^{\frac{1}{2}}, \quad (2.1)$$

where $\mathbf{F}_{jl} \sim \mathcal{CN}_{M \times K}(\mathbf{0}_{M \times K}, \mathbf{I}_M \otimes \mathbf{I}_K)$ accounts for small-scale fading, and $\mathbf{D}_{jl} = \text{diag}(\eta_{j,l,1}, \dots, \eta_{j,l,K})$ represents large-scale fading. The channel gains are independent and identically distributed (i.i.d.) and are assumed to remain fixed over the channel coherence time. Furthermore due to reciprocity, the BS to user channel matrix becomes \mathbf{G}_{jl}^T .

2.1.2 Channel Estimation and Pilot Contamination

The performance gains in mMIMO systems depends on the availability of channel-state information (CSI) [102]. Thus, obtaining the CSI is one of the main components of a mMIMO systems. The CSI acquisition is mainly done using uplink pilot transmissions. However, the time taken for pilot transmissions is limited by the coherence time of the system. Furthermore, if the time spent on pilot transmissions is high, it will result in lower data transmission time.

Due to the limited time allocated for pilot transmission, number of available pilot symbols will be limited. Thus, the same pilot has to reused in nearby cells. This will result in pilot contamination. Moreover, increasing the number of antennas at the BS does not reduce this problem. Some possible solutions include pilot scheduling algorithms and intelligent pilot sequence allocation algorithms [103]. Thus, whenever performance analysis of a mMIMO system is done it is important to consider pilot contamination and the effect of imperfect CSI.

The l th BS tries to estimate \mathbf{G}_l by using the pilot sequences transmitted by the users. During the training period, all users in the cell l simultaneously transmit mutually orthogonal pilot sequences of length τ . The pilot sequences used by K users can be represented by a $\tau \times K$ matrix $\sqrt{P_P}\mathbf{\Gamma}_l$, which satisfies $\mathbf{\Gamma}_l^H\mathbf{\Gamma}_l = \mathbf{I}_K$. Due to the unavailability of orthogonal pilot sequences for all users in L cells, the same sequence is reused by the $L - 1$ adjacent co-channel cells. Thus, $\mathbf{\Gamma}_j = \mathbf{\Gamma}_l = \mathbf{\Gamma}$ for $j \in \{1, 2, \dots, L\}$. Under these conditions, the received signal at B_l is given as

$$\mathbf{Y}_{P,l} = \sqrt{P_P} \sum_{j=1}^L \mathbf{G}_{jl}\mathbf{\Gamma}_j^T + \mathbf{N}_l, \quad (2.2)$$

where \mathbf{N}_l is an $K \times \tau$ noise matrix with i.i.d. $\mathcal{CN}(0, 1)$ elements and P_P is the transmit power of the pilot sequence. The minimum mean square error (MMSE) estimate of G_l corresponding to (2.2) can be derived as [104, 105]

$$\hat{\mathbf{G}}_l = \frac{1}{\sqrt{P_P}} \mathbf{Y}_{R_l} \mathbf{\Gamma}^* \tilde{\mathbf{D}}_l = \left(\sum_{j=1}^L \mathbf{G}_{jl} + \mathbf{V}_l / \sqrt{P_P} \right) \tilde{\mathbf{D}}_l, \quad (2.3)$$

where the elements of \mathbf{V}_l are distributed as independent $\mathcal{CN}(0, 1)$ random variables, $\tilde{\mathbf{D}}_l = \left(\frac{1}{P_P} \left(\sum_{j=1}^L \mathbf{D}_{jl} \right)^{-1} + \mathbf{I}_K \right)^{-1}$. The estimation error is $\mathbf{E}_l = \hat{\mathbf{G}}_l - \sum_{j=1}^L \mathbf{G}_{jl}$. The elements of the k th column of \mathbf{E}_l is Gaussian distributed with mean zero and variance $\sum_{j=1}^L \eta_{j,l,k} / (P_P \sum_{j=1}^L \eta_{j,l,k} + 1)$. Due to the MMSE properties, the matrices \mathbf{E}_l and $\hat{\mathbf{G}}_l$ are statistically independent. Here the pilot transmit power will be a

portion of the full transmit power of the user. Thus $P_p = \tau P_U$ where P_U is the transmit powers of the users and τ is the portion of the time allocated for the pilot transmission.

2.1.3 Downlink signal-to-interference-plus-noise ratio (SINR) values

Let $x_{l,k}$ be the data that has to be transmitted by B_l to $U_{l,k}$. B_l transmits $K \times 1$ signal vector \mathbf{x}_l , which is the concatenated data signals of K users in the l th cell. The signal vector \mathbf{x}_l satisfies $\mathcal{E}[\mathbf{x}_l \mathbf{x}_l^H] = \mathbf{I}_K$. The base stations use beamforming for transmitting data to the K users in the cell. Here, any linear beamforming method can be used and in this chapter maximal ratio transmission (MRT) is considered. The beamforming matrix at B_l is given as $\hat{\mathbf{G}}_{ll}^*$. Then the received signal at $U_{l,k}$ is written as

$$\mathbf{y}_{lk} = \sqrt{P_S} \sum_{j=1}^L \mathbf{g}_{jl,k}^T \hat{\mathbf{G}}_{jj}^* \mathbf{x}_j + n_{lk}, \quad (2.4)$$

where n_{lk} is the additive white Gaussian noise (AWGN) at the user with power σ^2 and P_S is the transmit power of the base stations. Here, $\mathbf{g}_{jl,k}$ is the k th column of the matrix \mathbf{G}_{jl} which represent the channel from $U_{l,k}$ to B_j . This can be expanded to the following form

$$\begin{aligned} \mathbf{y}_{lk} &= \sqrt{P_S} \mathbf{g}_{ll,k}^T \hat{\mathbf{G}}_{ll}^* \mathbf{x}_l + \sqrt{P_S} \sum_{j=1, j \neq l}^L \mathbf{g}_{jl,k}^T \hat{\mathbf{G}}_{jj}^* \mathbf{x}_j + n_{lk}, \\ &= \sqrt{P_S} \mathbf{g}_{ll,k}^T \hat{\mathbf{g}}_{ll,k}^* x_{l,k} + \sqrt{P_S} \mathbf{g}_{ll,k}^T \sum_{m=1, m \neq k}^K \hat{\mathbf{g}}_{ll,m}^* x_{l,m} \\ &\quad + \sqrt{P_S} \sum_{j=1, j \neq l}^L \sum_{m=1}^K \mathbf{g}_{jl,k}^T \hat{\mathbf{g}}_{jl,m}^* x_{j,m} + n_{lk} \end{aligned} \quad (2.5)$$

where the first term represents the intended signal to the user, the second term represent the inter-user interference within the cell, and the third term represent the inter-cell interference. Based on this equation the SINR of the user $U_{l,k}$ can be obtained as

$$\gamma_{lk} = \frac{P_S \left\| \mathbf{g}_{ll,k}^T \hat{\mathbf{g}}_{ll,k}^* \right\|^2}{P_S \sum_{m=1, m \neq k}^K \left\| \mathbf{g}_{ll,k}^T \hat{\mathbf{g}}_{ll,m}^* \right\|^2 + P_S \sum_{j=1, j \neq l}^L \sum_{m=1}^K \left\| \mathbf{g}_{jl,k}^T \hat{\mathbf{g}}_{jl,m}^* \right\|^2 + \sigma^2}. \quad (2.6)$$

The asymptotic value (when the number of antennas at the base stations goes to infinity) can be obtained as follows. For mMIMO, the base stations can scale their

power according to the number of antennas at the the BS. Thus, $P_S = E_S/\sqrt{M}$. Thus, (2.5) can be written as

$$\begin{aligned} \mathbf{y}_{lk} &= \frac{\sqrt{P_S}}{\sqrt[4]{M}} \mathbf{g}_{ll,k}^T \hat{\mathbf{g}}_{ll,k}^* x_{l,k} + \frac{\sqrt{P_S}}{\sqrt[4]{M}} \mathbf{g}_{ll,k}^T \sum_{m=1, m \neq k}^K \hat{\mathbf{g}}_{ll,m}^* x_{l,m} \\ &+ \frac{\sqrt{P_S}}{\sqrt[4]{M}} \sum_{j=1, j \neq l}^L \sum_{m=1}^K \mathbf{g}_{jl,k}^T \hat{\mathbf{g}}_{jj,m}^* \mathbf{x}_j + n_{lk}. \end{aligned} \quad (2.7)$$

Asymptotic results are obtained by utilizing the following two limit results.

$$\frac{\mathbf{g}_{ll,k}^T \hat{\mathbf{g}}_{jl,k}^*}{M} \xrightarrow[M \rightarrow \infty]{a.s.} \frac{E_P \eta_{j,l,k} \sum_{j=1}^L \eta_{j,l,k}}{1 + E_P \sum_{j=1}^L \eta_{j,l,k}} = \hat{\eta}_{j,l,k}, \quad P_P = E_P, \quad (2.8)$$

$$\frac{\mathbf{g}_{ll,k}^T \hat{\mathbf{g}}_{jl,m}^*}{M} \xrightarrow[M \rightarrow \infty]{a.s.} 0, \quad P_P = E_P, k \neq m, \quad (2.9)$$

By using (2.8) and (2.9), the SINR of the system can be written as

$$\gamma_{lk}^\infty = \frac{P_S \hat{\eta}_{l,l,k}^2}{P_S \sum_{j=1, j \neq l}^L \hat{\eta}_{j,l,k}^2}. \quad (2.10)$$

It is evident from (2.10), that mMIMO suppresses the effects of noise, co-channel interference (CCI), and the inter-cell interference.

2.1.4 SINR with perfect CSI

When there are no estimation errors on CSI, the precoding vector becomes $\sum_{j=1}^L \mathbf{G}_{jl}^*$. With this, $\hat{\eta}_{j,l,m} = \eta_{j,l,m}$ and the SINR is given as

$$\gamma_{lk}^\infty = \frac{P_S \eta_{l,l,k}^2}{P_S \sum_{j=1, j \neq l}^L \eta_{j,l,k}^2}. \quad (2.11)$$

2.1.5 Cell-free mMIMO

The mMIMO concept can be used in a distributed manner without cells. This is called cell-free mMIMO [106]. A cell-free mMIMO system consists of a central processing unit (CPU) and a set of distributed access points (APs) [107]. The APs are single or multiple antenna nodes that are connected to the CPU via a front-haul link. The large number of distributed APs will facilitate the data transfer for a smaller number of UEs in the system. All the APs will simultaneously take part in the communication process with all the UEs. The APs will send the received signal from UEs via the uplink to the CPU and CPU will perform joint decoding. During the downlink, the CPU will send the transmission coefficients and the downlink data

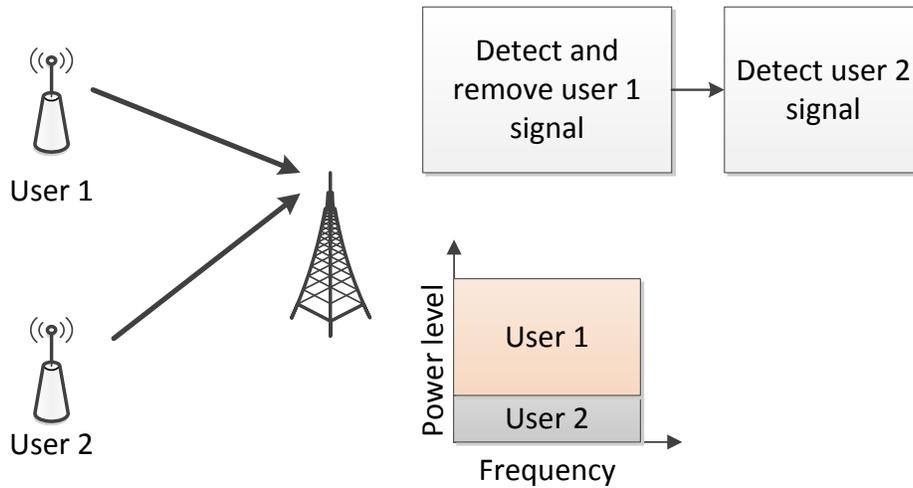


Figure 2.1: Uplink data transmission in NOMA systems

to all the APs. Both during the uplink and the downlink, the channel statistics (instead of the instantaneous CSI) are used for precoding and beamforming.

Cell-free mMIMO provides higher performance than the small cell systems [106] and as result has been identified as a potential technology for 5G wireless systems [107]. The main advantages of cell-free mMIMO are high energy efficiency, high spectral efficiency, flexibility and ease of deployment, and the uniform quality of service for UEs. Some potential research problems for cell-free mMIMO are channel estimation, power control, and AP selection methods.

2.2 NOMA Systems

This section provides the basic system model for a two-user NOMA system [50,61].

2.2.1 Uplink NOMA transmission

First the uplink transmission of NOMA is analysed. It is assumed that User 1 is near the BS and thus, the channel between the User 1 and the BS is stronger than the channel between User 2 and the BS. Both the users transmit at the same time using the same frequency and the BS receives a superimposed signal from both the users. However, the signal power from User 1 is higher due to its stronger channel. BS will first decode the signal from User 1 by assuming the signal from User 2 as

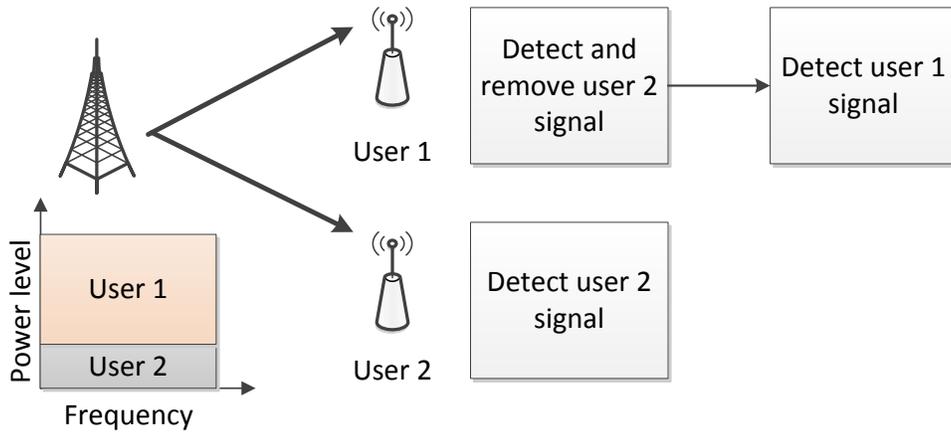


Figure 2.2: Downlink transmission of NOMA systems

noise. Next it will remove this decoded signal and then decode the signal from User 2. Similar steps can be done when the number of users is greater than 2. The mathematical expressions for this case can be given as follows.

The channel between the BS and the users 1 and 2 are given as h_1 and h_2 and $h_1 > h_2$ respectively. The transmit signals from User 1 and User 2 are x_1 and x_2 and the transmit powers from User 1 and User 2 are assigned according to α_1 and α_2 where $\alpha_1^2 + \alpha_2^2 = 1$. The received signal at the BS can be written as follows.

$$y_r = \alpha_1 P h_1 x_1 + \alpha_2 P h_2 x_2 + n_r, \quad (2.12)$$

where n_r is the noise at the BS and P is the maximum transmit power of users. The SINR of the User 1 $\gamma_{1,u}$ and User 2 $\gamma_{2,u}$ can be written as follows:

$$\gamma_{1,u} = \frac{\alpha_1^2 P^2 |h_1|^2}{\sigma_n^2}, \quad (2.13)$$

$$\gamma_{2,u} = \frac{\alpha_2^2 P^2 |h_2|^2}{\alpha_1^2 P^2 |h_2|^2 + \sigma_n^2}. \quad (2.14)$$

By changing the values of α_1 and α_2 , different data rate values can be achieved for User 1 and User 2.

2.2.2 Downlink NOMA transmission

For downlink NOMA, the BS will broadcast a a superimposed signal to both users. As the channel for User 2 is weaker, BS will assign more power to the User 2 signal.

After receiving the signal, User 2 will just decode its signal by considering the signal for User 1 as noise. However, User 1 will first decode the signal intended to User 2 and then remove this from the received signal. This decoding process is known as the successive interference cancellation. A similar setup can be used for more than two users as well. The transmit signal of the BS can thus be written as follows:

$$y_t = \beta_1 x_1 + \beta_2 x_2, \quad (2.15)$$

where x_1 and x_2 are the signal intended for User 1 and User 2 respectively. The received signal at User 1 and User 2 can be written as follows:

$$y_1 = h_1^T P_B (\beta_1 x_1 + \beta_2 x_2) + n_1, \quad (2.16)$$

and

$$y_2 = h_2^T P_B (\beta_1 x_1 + \beta_2 x_2) + n_2, \quad (2.17)$$

where n_1 and n_2 are the noise terms at User 1 and User 2. P_B is the transmit power of the BS. The SINR of User 1 $\gamma_{1,d}$ and User 2 $\gamma_{2,d}$ can be expressed as follows:

$$\gamma_{1,d} = \frac{\beta_1^2 P_B^2 |h_1|^2}{\sigma_{n_1}^2}, \quad (2.18)$$

$$\gamma_{2,d} = \frac{\beta_2^2 P_B^2 |h_2|^2}{\beta_1^2 P_B^2 |h_2|^2 + \sigma_{n_2}^2}. \quad (2.19)$$

2.2.3 Comparison with orthogonal multiple access (OMA) methods

The sum rate of a similar system with orthogonal multiple access with time division can be given as follows. The results for OMA is only given for uplink only. Similar rate regions can be obtained for downlink as well. The received SINR under OMA can be written as follows:

$$\gamma_{1,u}^O = \frac{P^2 |h_1|^2}{\sigma_n^2}, \quad (2.20)$$

and

$$\gamma_{2,u}^O = \frac{P^2 |h_2|^2}{\sigma_n^2}. \quad (2.21)$$

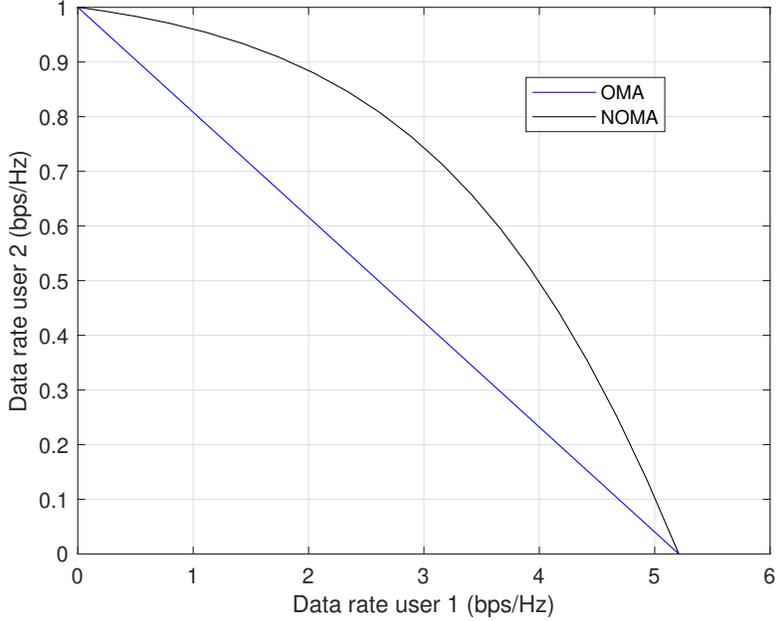


Figure 2.3: Uplink data rate for NOMA vs. and OMA

Suppose that τ_1 and τ_2 are the fractional times assigned for user 1 and user 2 respectively where $\tau_1 + \tau_2 = 1$. Based on this, the total sum rate of the system can be written as follows:

$$R^{OMA} = \tau_1 \log(1 + \gamma_{1,u}^O) + \tau_2 \log(1 + \gamma_{2,u}^O). \quad (2.22)$$

The sum rate of the NOMA system can be written as follows:

$$R^{NOMA} = \log(1 + \gamma_{1,u}) + \log(1 + \gamma_{2,u}). \quad (2.23)$$

The achievable rate regions under NOMA and OMA schemes are plotted in Fig. 2.3 by varying τ_1, τ_2, α_1 , and α_2 values. It can be seen that NOMA offers a higher sum-rate region compared to OMA.

2.2.4 NOMA with multiple users

The use of NOMA with more than two users for the downlink can be explained as follows. When there are K users in the systems, the users are ordered according to the channel quality between the users and the BS. Thus without loss of generality, $h_1 > h_2 > h_3 > \dots > h_{K-1} > h_K$. Let α_k represent the power allocation for the user U_k and $\sum_{k=1}^K \alpha_k^2 = 1$. User U_k will decode all the signals that are intended

to users U_{k+1} to U_K and perform successive interference cancellation (SIC). The signals that are transmitted to other users are considered as noise. Thus, the SINR of U_k is given as

$$\gamma_{k,d} = \frac{\beta_k^2 P_B^2 |h_k|^2}{P_B^2 |h_k|^2 \sum_{l=1}^{k-1} \beta_l^2 + \sigma_{n_2}^2}. \quad (2.24)$$

Similar steps are taken for the uplink as well.

2.3 Conclusion

This chapter provided basic principles of mMIMO and NOMA systems. In summary, mMIMO systems provide higher spectral efficiencies and energy efficiencies by utilizing the spatial degrees of freedom offered by the large antenna array. But CSI estimation errors and pilot contamination are as the main performance bottlenecks. On the other hand, NOMA utilizes non-orthogonal channels for multiple access and has shown significant spectral efficiency gains compared to orthogonal multiple access.

Chapter 3

Performance Analysis of mMIMO TWRNs with Channel Imperfections

This chapter considers multi-cell multi-pair two-way relay networks (TWRNs) consisting of single-antenna user nodes and massive multiple-input multiple-output (mMIMO) amplify-and-forward (AF) relay nodes. The combined effects of co-channel interference (CCI), imperfect channel-state information (CSI), pilot contamination, and the antenna correlation is analysed. It is shown that the effects of CCI and the antenna correlation can be completely mitigated by using mMIMO. However, the effects of imperfect CSI and pilot contamination degrade the performance even with a large antenna array. Yet, the use of mMIMO allows power scaling at the user nodes and relay and thus, even with channel imperfections, the benefits of employing a mMIMO enabled relay on transmit power savings are significant. Furthermore, the closed-form approximations for the sum rate are derived. This result helps to decide the required number of relay antennas to obtain a certain percentage of the asymptotic sum rate. Furthermore, the optimal pilot sequence length to maximize the sum rate of the system is derived.

3.1 Introduction

Multi-pair mMIMO TWRNs are used to enable pairwise data exchange among a set of users. With dense deployment of these wireless systems, CCI and pilot contamination are dominant performance limiting factors [19]. Moreover, because precoders/detectors need CSI, channel estimation errors (imperfect CSI) severely degrades the overall performance. However, many works [65, 75–78] assume per-

fect CSI. In contrast, this chapter analyzes the effects of CCI, imperfect CSI, pilot contamination, and antenna correlation for multi-pair mMIMO TWRNs.

More specifically, the contributions of this chapter can be listed as follows.

1. The asymptotic signal-to-interference-plus-noise ratio (SINR) and sum rate expressions are derived for three transmit power scaling laws: namely (i) power scaling at user nodes, (ii) power scaling at the relay, and (iii) power scaling at both the relay and user nodes. For these three cases, it is shown that the asymptotic SINR expressions become independent of the number of co-channel interferers (L). This suggests that the CCI degradation can be cancelled asymptotically if the relay antenna count grows unbounded. Nevertheless, the asymptotic performance is limited by the residual interference from pilot contamination, which cannot be completely mitigated even with infinitely many relay antennas. Notably, the asymptotic performance metrics are independent of the fast fading component of the wireless channel. The insight of this fact is that the cross-layer resource scheduling becomes simple. The analysis and Monte-Carlo simulations reveal that very large relay antenna arrays achieve substantial sum-rate gains.
2. The asymptotic results are valid only when the number of antennas at the relay is infinite. However, for practical purposes, a finite number of antennas must be considered. Thus closed-form sum rate results are derived in this chapter. To make analysis tractable, perfect channel conditions (no CCI, no pilot contamination and perfect CSI) are assumed. Nevertheless, the result is useful to decide the optimal number of relay antennas to obtain a certain percentage of the asymptotic performance and to determine how fast the performance of the system approaches the asymptotic performance. -
3. Asymptotic SINR and sum-rates for multi-pair mMIMO TWRNs with relay antenna correlation are derived. Although the effect of antenna correlation on the performance can be drastic, our analysis shows that the correlation impact can be mitigated by using a large antenna arrays.

3.2 System, channel, and signal model

3.2.1 System and channel model

The system model consists of L adjacent TWRNs having $2K$ number of users each. The users in the l th TWRN are denoted as $(U_{l,1}, \dots, U_{l,2K})$, where $U_{l,k}$ exchange data signals with its paired-user $U_{l,k'}$ via the half-duplex AF relay R_l for $k, k' \in \{1, \dots, 2K\}$ and $l \in \{1, \dots, L\}$. Here, R_l s are specialized relay nodes with higher power availability than users. Users are single-antenna terminals, and the relays are equipped with N antennas. The number of relay antennas are unbounded with respect to the total number of users ($N \gg 2K$). The channel matrix from $2K$ users in the j th TWRN to the l th relay is represented as $\mathbf{G}_{jl} = \mathbf{F}_{jl} \mathbf{D}_{jl}^{\frac{1}{2}}$, where $\mathbf{F}_{jl} \sim \mathcal{CN}_{N \times 2K}(\mathbf{0}_{N \times 2K}, \mathbf{I}_N \otimes \mathbf{I}_{2K})$ accounts for small-scale fading, and $\mathbf{D}_{jl} = \text{diag}(\eta_{j,l,1}, \dots, \eta_{j,l,2K})$ represents large-scale fading. As is customary, the channel gains are independent and identically distributed (i.i.d.) and are assumed to remain fixed over two consecutive time-slots and reciprocal, and hence, relay-to-user channel matrix becomes \mathbf{G}_{jl}^T . The CCI on the l th TWRN occurs due to the data transmissions of the other $L - 1$ TWRNs with relay R_j , where $j \in \{1, \dots, L\}$ and $j \neq l$.

3.2.2 Channel estimation

The l th relay tries to estimate \mathbf{G}_{ll} by using the pilot sequences transmitted by the users. During the training period, all users in the cell l simultaneously transmit mutually orthogonal pilot sequences of length τ . The pilot sequences used by $2K$ users can be represented by a $\tau \times 2K$ matrix $\sqrt{P_P} \mathbf{\Gamma}_l$, which satisfies $\mathbf{\Gamma}_l^H \mathbf{\Gamma}_l = \mathbf{I}_{2K}$. Yet, due to the unavailability of orthogonal pilot sequences for all users in L TWRNs, the same sequence is reused by the $L - 1$ adjacent co-channel TWRNs. Thus, $\mathbf{\Gamma}_j = \mathbf{\Gamma}_l = \mathbf{\Gamma}$ for $j \in \{1, 2, \dots, L\}$. Under these conditions, the received signal at the relay is given as

$$\mathbf{Y}_{R_l} = \sqrt{P_P} \sum_{j=1}^L \mathbf{G}_{jl} \mathbf{\Gamma}_j^T + \mathbf{N}_l, \quad (3.1)$$

where \mathbf{N}_l is a $2K \times \tau$ noise matrix with i.i.d. $\mathcal{CN}(0, 1)$ elements and P_P is the transmit power of the pilot sequence. The MMSE estimate of G_{ll} corresponding to

(3.1) can be derived as [104, 105]

$$\hat{\mathbf{G}}_{ll} = \frac{1}{\sqrt{P_p}} \mathbf{Y}_{R_l} \mathbf{\Gamma}^* \tilde{\mathbf{D}}_{ll} = \left(\sum_{j=1}^L \mathbf{G}_{jl} + \mathbf{V}_l / \sqrt{P_p} \right) \tilde{\mathbf{D}}_{ll}, \quad (3.2)$$

where the elements of \mathbf{V}_l are distributed as independent $\mathcal{CN}(0, 1)$ random variables, $\tilde{\mathbf{D}}_{ll} = \left(\frac{1}{P_p} \left(\sum_{j=1}^L \mathbf{D}_{jl} \right)^{-1} + \mathbf{I}_K \right)^{-1}$, and \mathbf{G}_{jl} is defined in the previous section. The estimation error is $\mathbf{E}_l = \hat{\mathbf{G}}_{ll} - \sum_{j=1}^L \mathbf{G}_{jl}$ ¹. The receive zero forcing (ZF) detector and transmit ZF precoder are constructed at the relay by using the CSI with estimation errors. The elements of the k th column of \mathbf{E}_l is Gaussian-distributed with mean zero and variance $\sum_{j=1}^L \eta_{j,l,k} / (P_p \sum_{j=1}^L \eta_{j,l,k} + 1)$. Due to the MMSE properties, the matrices \mathbf{E}_l and $\hat{\mathbf{G}}_{ll}$ are statistically independent.

3.2.3 Signal model

During two time-slots, $2K$ users in each TWRN exchange their information pairwise via their assigned relay. Specifically, the paired users $(U_{l,2i-1}, U_{l,2i})$ exchange their data signals $(x_{l,2i-1}, x_{l,2i})$, where $i \in \{1, \dots, K\}$ and $l \in \{1, \dots, L\}$. In the first time-slot, users transmit $2K \times 1$ signal vector \mathbf{x}_l , which is the concatenated data signals of $2K$ users in the l th TWRN, towards the relay R_l . The signal vector \mathbf{x}_l satisfies $\mathcal{E}[\mathbf{x}_l \mathbf{x}_l^H] = \mathbf{I}_{2K}$. The received signal at R_l is written as

$$\mathbf{y}_{R_l} = \sqrt{P_S} \sum_{j=1}^L \mathbf{G}_{jl} \mathbf{x}_j + \mathbf{n}_{R_l}, \quad (3.3)$$

where \mathbf{n}_{R_l} is the $N \times 1$ additive white Gaussian noise (AWGN) vector at the relay satisfying $\mathcal{E}[\mathbf{n}_{R_l} \mathbf{n}_{R_l}^H] = \mathbf{I}_N \sigma_{R_l}^2$ and P_S is the transmit power of the users. During the second-time slot, relay first amplifies and then forwards its received signal towards the users. The transmitted signal from the relay is $\mathbf{y}'_{R_l} = \hat{\beta}_l \hat{\mathbf{W}}_l \mathbf{y}_{R_l}$, where $\hat{\mathbf{W}}_l$ is the concatenated beamforming-and-amplification matrix at the relay and $\hat{\beta}_l$ is the amplification factor to satisfy the relay power constraint which is presented in the sequel. Here, $\hat{\mathbf{W}}_l$ is designed to cancel sub-channel interference within a given TWRN, and hence, it is constructed by using receive-ZF and transmit-ZF precoding and detection concepts as follows [108, 109]:²

$$\hat{\mathbf{W}}_l = \hat{\mathbf{G}}_{ll}^* (\hat{\mathbf{G}}_{ll}^T \hat{\mathbf{G}}_{ll}^*)^{-1} \mathbf{P} (\hat{\mathbf{G}}_{ll}^H \hat{\mathbf{G}}_{ll})^{-1} \hat{\mathbf{G}}_{ll}^H, \quad (3.4)$$

¹This definition for estimation error instead of $\mathbf{E}_l = \hat{\mathbf{G}}_{ll} - \mathbf{G}_{ll}$ is used, as even with a perfect MMSE estimation, the estimated channel will be the sum of all the channels that have used the same pilot sequence $(\sum_{j=1}^L \mathbf{G}_{jl})$.

²The results obtained in this chapter can also be obtained for other detection and precoding methods such as matched filter and MMSE.

where \mathbf{P} is the block diagonal permutation matrix for user pairing, constructed as $\mathbf{P} = \text{diag}(\mathbf{P}_1, \dots, \mathbf{P}_K)$ and $\mathbf{P}_i = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ for $i \in \{1, \dots, K\}$. Moreover, the relay power constraint is

$$P_R = \mathbb{E} \left[\text{Tr} \left(\mathbf{y}'_{R_l} \mathbf{y}'_{R_l}{}^H \right) \right]. \quad (3.5)$$

Next, the received signal at $U_{l,k'}$ is given as

$$\begin{aligned} y_{l,k'} &= \mathbf{g}_{l,k'}^T \mathbf{y}'_{R_l} + n_{l,k'} = \hat{\beta}_l \sqrt{P_S} \mathbf{g}_{l,k'}^T \hat{\mathbf{W}}_l \mathbf{G}_l \mathbf{x}_l \\ &+ \hat{\beta}_l \sqrt{P_S} \mathbf{g}_{l,k'}^T \hat{\mathbf{W}}_l \sum_{j=1, j \neq l}^L \mathbf{G}_{jl} \mathbf{x}_j + \hat{\beta}_l \mathbf{g}_{l,k'}^T \hat{\mathbf{W}}_l \mathbf{n}_{R_l} + n_{l,k'}, \end{aligned} \quad (3.6)$$

where $U_{l,k}$ and $U_{l,k'}$ are the paired-users exchanging their data signals with $(k, k') = (2i - 1, 2i)$ for $i \in \{1, \dots, 2K\}$. Moreover, $\mathbf{g}_{l,k'}$ is the k' th column vector of the matrix \mathbf{G}_l for $k \in \{1, \dots, K\}$ and $n_{l,k'}$ is the AWGN at the k' th user of the l th TWRN with variance $\sigma_{n_{l,k'}}^2$. (3.6) is further simplified as

$$\begin{aligned} y_{l,k'} &= \hat{\beta}_l \sqrt{P_S} \mathbf{g}_{l,k'}^T \hat{\mathbf{W}}_l \mathbf{g}_{lk} x_{l,k} + \hat{\beta}_l \sqrt{P_S} \mathbf{g}_{l,k'}^T \hat{\mathbf{W}}_l \sum_{m=1, m \neq k}^{2K} \mathbf{g}_{lm} x_{l,m} \\ &+ \hat{\beta}_l \sqrt{P_S} \mathbf{g}_{l,k'}^T \hat{\mathbf{W}}_l \sum_{j=1, j \neq l}^L \mathbf{G}_{jl} \mathbf{x}_j + \hat{\beta}_l \mathbf{g}_{l,k'}^T \hat{\mathbf{W}}_l \mathbf{n}_{R_l} + n_{l,k'}, \end{aligned} \quad (3.7)$$

where the first term is the desired signal at $U_{l,k'}$ and other terms are the interferences and noise. Thus the end-to-end SINR at $U_{l,k'}$, $\gamma_{l,k'}$ is given as

$$\gamma_{l,k'} = \frac{\hat{\beta}_l^2 P_S \left\| \mathbf{g}_{l,k'}^T \hat{\mathbf{W}}_l \mathbf{g}_{lk} \right\|^2}{\hat{\beta}_l^2 P_S \Theta + \hat{\beta}_l^2 \sigma_{R_l}^2 \sum_{j=1, j \neq l}^L \left\| \mathbf{g}_{l,k'}^T \hat{\mathbf{W}}_l \right\|^2 + \sigma_{n_{l,k'}}^2}, \quad (3.8)$$

where $\Theta = P_S \sum_{m=1, m \neq k}^{2K} \left\| \mathbf{g}_{l,k'}^T \hat{\mathbf{W}}_l \mathbf{g}_{lm} \right\|^2 + P_S \sum_{j=1, j \neq l}^L \left\| \mathbf{g}_{l,k'}^T \hat{\mathbf{W}}_l \mathbf{G}_{jl} \right\|^2$.

3.2.4 Calculation of value $\hat{\beta}_l$

By using (3.3) and $\mathbf{y}'_{R_l} = \hat{\beta}_l \hat{\mathbf{W}}_l \mathbf{y}_{R_l}$, the power constraint (3.5) is simplified as

$$\begin{aligned} P_R &= \hat{\beta}_l^2 \mathbb{E} \left[\text{Tr} \left(\hat{\mathbf{W}}_l \left(P_S \sum_{j=1}^L \mathbf{G}_{jl} \mathbf{G}_{jl}^H + \sigma_{R_l}^2 \mathbf{I}_N \right) \hat{\mathbf{W}}_l^H \right) \right] \\ &= \hat{\beta}_l^2 P_S \mathbb{E} \left[\text{Tr} \left(\hat{\mathbf{W}}_l \sum_{j=1}^L \mathbf{G}_{jl} \mathbf{G}_{jl}^H \hat{\mathbf{W}}_l^H \right) \right] + \hat{\beta}_l^2 \sigma_{R_l}^2 \mathbb{E} \left[\text{Tr} \left(\hat{\mathbf{W}}_l \hat{\mathbf{W}}_l^H \right) \right]. \end{aligned} \quad (3.9)$$

By substituting $\hat{\mathbf{W}}_l$ into (3.9), $\hat{\beta}_l$ is obtained as follows:

$$\hat{\beta}_l = \sqrt{\frac{P_R}{P_S \mathcal{T}_1 + \sigma_{R_l}^2 \mathcal{T}_2}}, \quad (3.10)$$

where $\mathcal{T}_1 = \sum_{j=1}^L \mathbb{E} \left[\text{Tr} \left(\mathbf{G}_{jl} \mathbf{G}_{jl}^H \hat{\mathbf{G}}_{ul} \left[\hat{\mathbf{G}}_{ul}^H \hat{\mathbf{G}}_{ul} \right]^{-1} \mathbf{P} \left[\hat{\mathbf{G}}_{ul}^T \hat{\mathbf{G}}_{ul}^* \right]^{-1} \mathbf{P} \left[\hat{\mathbf{G}}_{ul}^H \hat{\mathbf{G}}_{ul} \right]^{-1} \hat{\mathbf{G}}_{ul}^H \right) \right]$ and $\mathcal{T}_2 = \mathbb{E} \left[\text{Tr} \left(\left[\hat{\mathbf{G}}_{ul}^H \hat{\mathbf{G}}_{ul} \right]^{-1} \mathbf{P} \left[\hat{\mathbf{G}}_{ul}^T \hat{\mathbf{G}}_{ul}^* \right]^{-1} \mathbf{P} \right) \right]$.

3.2.5 Overall sum rate of the system

The average sum rate for the l th $2K$ -user TWRN with estimated CSI at the relay is defined as [17]

$$\mathcal{R}_l = \frac{(T_C - \tau)}{2T_C} \sum_{k=1}^{2K} \mathbb{E} [\log(1 + \gamma_{l,k})], \quad (3.11)$$

where T_C and τ are the coherence time of the wireless channel and length of the pilot sequence used for channel estimation. In particular, the pre-log factor $(T_C - \tau)/T_C$ accounts for the pilot overhead [17]. The two time-slots required for the data transmission between the paired users results in the pre-log factor of 1/2.

3.2.6 Energy efficiency of the system

As mentioned in the introduction, for many applications such as low-power sensor networks, the energy efficiency of the system will be very important. Thus, the energy efficiency of the system defined as [65]

$$\rho = \frac{\sum_{k=1}^{2K} \mathbb{E} [\log(1 + \gamma_{l,k})]}{2K P_S + P_R}, \quad (3.12)$$

where the denominator consists of the total power consumption of the system and the numerator consists of the overall sum rate.

3.3 Asymptotic performance analysis

This section derives asymptotic SINR and sum rate for the three power scaling laws³ [65] at the user nodes and relay whenever the relay antenna count grows unbounded ($N \rightarrow \infty$). Moreover, the detrimental impacts of CCI, imperfect CSI, and pilot contamination on the SINR of the system are analysed.

³Specifically, transmit power scaling laws are designed to achieve the highest power scaling coefficient at the transmitters that results in a non zero SINR. In this context, both the spectral and energy efficiency of the proposed mMIMO TWRNs can be substantially improved subject to a fundamental trade-off.

3.3.1 Transmit power scaling at the user nodes

Whenever the transmit power of user nodes is scaled inversely proportional to the relay antenna count, the power of the pilot sequence is also scaled accordingly. Thus, the overall transmit power can only be scaled inversely proportional to the square-root of the relay antenna count (\sqrt{N}) for estimated CSI. The normalized relay gain for unlimited number of relay antennas is obtained by letting $P_S = E_S/\sqrt{N}$, $P_P = \tau E_S/\sqrt{N}$ and $P_R = E_R$ where E_S and E_R are fixed, as

$$\lim_{N \rightarrow \infty} \frac{\hat{\beta}_l}{\sqrt{N}} = \left[\frac{E_R}{E_S \sum_{j=1}^L \sum_{i=1}^K \left(\frac{\eta_{j,l,2i-1}^2 + \eta_{j,l,2i}^2}{\tau E_S \hat{\eta}_{l,2i-1}^2 \hat{\eta}_{l,2i}^2} \right) + 2\sigma_{R_l}^2 \sum_{i=1}^K \left(\tau E_S \hat{\eta}_{l,2i-1} \hat{\eta}_{l,2i} \right)^{-2}} \right]^{\frac{1}{2}} \quad (3.13)$$

See Appendix A.2 for the proof. Here, $\hat{\eta}_{l,k}$ is defined as

$$\hat{\eta}_{l,k} = \sum_{j=1}^L \eta_{j,l,k}. \quad (3.14)$$

The parameter $\hat{\eta}_{l,k}$, is a measure of the pilot contamination experienced by $U_{l,k}$ and will extensively appear in SINR equations in the sequel. Furthermore, if pilot contamination is absent, then $\hat{\eta}_{l,k} = \eta_{l,l,k}$. Eqn. (3.7) is rewritten by dividing both sides by $\sqrt[4]{N}$ and substituting E_S values as

$$\begin{aligned} \frac{y_{l,k'}}{\sqrt[4]{N}} &= \frac{\hat{\beta}_l}{\sqrt{N}} \sqrt{E_S} \mathbf{g}_{lk'}^T \hat{\mathbf{W}}_l \mathbf{g}_{lk} x_{l,k} + \frac{\hat{\beta}_l}{\sqrt{N}} \sqrt{E_S} \mathbf{g}_{lk'}^T \hat{\mathbf{W}}_l \sum_{m=1, m \neq k}^{2K} \mathbf{g}_{lm} x_{l,m} \\ &\quad + \frac{\hat{\beta}_l}{\sqrt{N}} \sqrt{E_S} \mathbf{g}_{lk'}^T \hat{\mathbf{W}}_l \sum_{j=1, j \neq l}^L \mathbf{G}_{jl} \mathbf{x}_j \\ &\quad + \frac{\hat{\beta}_l}{\sqrt{N}} \sqrt[4]{N} \mathbf{g}_{lk'}^T \hat{\mathbf{W}}_l \mathbf{n}_{R_l} + \frac{1}{\sqrt[4]{N}} n_{l,k'}. \end{aligned} \quad (3.15)$$

Next, the asymptotic limit of the intended signal term in (3.15) is obtained by using the limits given in Appendix A.1, as

$$\begin{aligned} \sqrt{E_S} \mathbf{g}_{lk'}^T \hat{\mathbf{W}}_l \mathbf{g}_{lk} x_{l,k} &\xrightarrow[N \rightarrow \infty]{a.s.} \sqrt{E_S} \frac{\mathbf{g}_{lk'}^T \hat{\mathbf{G}}_{ll}^*}{\sqrt{N}} \left(\frac{\hat{\mathbf{G}}_{ll}^T \hat{\mathbf{G}}_{ll}^*}{\sqrt{N}} \right)^{-1} \mathbf{P} \left(\frac{\hat{\mathbf{G}}_{ll}^H \hat{\mathbf{G}}_{ll}}{\sqrt{N}} \right)^{-1} \frac{\hat{\mathbf{G}}_{ll}^H \mathbf{g}_{lk}}{\sqrt{N}} x_{l,k} \\ &= \sqrt{E_S} \frac{\eta_{l,l,k'} \eta_{l,l,k}}{\hat{\eta}_{l,k}^2} x_{l,k}. \end{aligned} \quad (3.16)$$

Similarly the asymptotic limit of the second term, which represents the inter-user interference in (3.15), is obtained as

$$\sqrt{E_S} \mathbf{g}_{lk'}^T \hat{\mathbf{W}}_l \sum_{m=1, m \neq k}^{2K} \mathbf{g}_{lm} x_{l,m} \xrightarrow[N \rightarrow \infty]{a.s.} 0. \quad (3.17)$$

Above, the inter-user interference is removed due to ZF receiving and transmission at the relay. Even with a different beamforming method, it can be shown that this value asymptotically goes to zero. Furthermore, the asymptotic limit of the third term in (3.15), which is the co-channel interference from the nearby j th TWRN is derived as

$$\begin{aligned}\sqrt{E_S} \mathbf{g}_{lk'}^T \hat{\mathbf{W}}_l \mathbf{G}_{jl} \mathbf{x}_j &\xrightarrow[N \rightarrow \infty]{a.s.} \sqrt{E_S} \frac{\mathbf{g}_{lk'}^T \hat{\mathbf{G}}_{ll}^* \left(\frac{\hat{\mathbf{G}}_{ll}^T \hat{\mathbf{G}}_{ll}^*}{\sqrt{N}} \right)^{-1} \mathbf{P} \left(\frac{\hat{\mathbf{G}}_{ll}^H \hat{\mathbf{G}}_{ll}}{\sqrt{N}} \right)^{-1} \hat{\mathbf{G}}_{ll}^H \mathbf{G}_{jl}}{\hat{\eta}_{l,k}^2} \mathbf{x}_j \\ &= \sqrt{E_S} \frac{\eta_{l,l,k'} \eta_{j,l,k}}{\hat{\eta}_{l,k}^2} x_{j,k}.\end{aligned}\quad (3.18)$$

Also, the fourth term in (3.15), which represents the added noise at the relay, is given as

$$\begin{aligned}\sqrt[4]{N} \mathbf{g}_{lk'}^T \hat{\mathbf{W}}_l \mathbf{n}_{R_l} &= \sqrt[4]{N} \mathbf{g}_{lk'}^T \hat{\mathbf{G}}_{ll}^* \left(\frac{\hat{\mathbf{G}}_{ll}^T \hat{\mathbf{G}}_{ll}^*}{\sqrt{N}} \right)^{-1} \mathbf{P} \left(\frac{\hat{\mathbf{G}}_{ll}^H \hat{\mathbf{G}}_{ll}}{\sqrt{N}} \right)^{-1} \hat{\mathbf{G}}_{ll}^H \mathbf{n}_{R_l} \\ &= \frac{\mathbf{g}_{lk'}^T \hat{\mathbf{G}}_{ll}^* \left(\frac{\hat{\mathbf{G}}_{ll}^T \hat{\mathbf{G}}_{ll}^*}{\sqrt{N}} \right)^{-1} \mathbf{P} \left(\frac{\hat{\mathbf{G}}_{ll}^H \hat{\mathbf{G}}_{ll}}{\sqrt{N}} \right)^{-1} \hat{\mathbf{G}}_{ll}^H \mathbf{n}_{R_l}}{\sqrt[4]{N}}.\end{aligned}\quad (3.19)$$

Based on (3.19), the asymptotic distribution of the fourth term is given as

$$\sqrt[4]{N} \mathbf{g}_{lk'}^T \hat{\mathbf{W}}_l \mathbf{n}_{R_l} \xrightarrow[N \rightarrow \infty]{d} \frac{\eta_{l,l,k'}}{\tau E_S \hat{\eta}_{l,k}^3} \tilde{n}_l, \quad (3.20)$$

where $\tilde{n}_l \sim \mathcal{CN}(0, \tau E_S \hat{\eta}_{l,k}^2 \sigma_{R_l}^2)$. Furthermore, the asymptotic limit of the last term in (3.15) is written as

$$\frac{1}{\sqrt[4]{N}} n_{l,k'} \xrightarrow[N \rightarrow \infty]{a.s.} 0. \quad (3.21)$$

By using the above asymptotic values and distributions of the terms in (3.15), the asymptotic SINR for transmit power scaling at the user nodes can be written as

$$\begin{aligned}\lim_{N \rightarrow \infty} \gamma_{l,k'} &= \frac{\frac{E_S \eta_{l,l,k'}^2 \eta_{l,l,k}^2}{\hat{\eta}_{l,k}^4}}{\frac{E_S \eta_{l,l,k'}^2 \sum_{j=1, j \neq l}^L \eta_{j,l,k}^2}{\hat{\eta}_{l,k}^4} + \frac{\eta_{l,l,k'}^2}{\tau E_S \hat{\eta}_{l,k}^4} \sigma_{R_l}^2} \\ &= \frac{\tau E_S^2 \eta_{l,l,k}^2}{\tau E_S^2 \sum_{j=1, j \neq l}^L \eta_{j,l,k}^2 + \sigma_{R_l}^2} = \gamma_\infty.\end{aligned}\quad (3.22)$$

3.3.2 Transmit power scaling at the relay

In this case, the asymptotic SINR is derived by scaling the transmit power at the relay inversely proportionally to the number of relay antennas. Thus the asymptotic

SINR is obtained by substituting $P_R = E_R/N$, $P_S = E_S$ and $P_p = \tau E_S$ where E_S and E_R are fixed. The asymptotic value of the normalized relay gain is obtained as

$$\lim_{N \rightarrow \infty} \hat{\beta}_l = \left[\frac{E_R}{E_S \sum_{j=1}^L \sum_{i=1}^K \frac{(\eta_{j,l,2i-1}^2(1+\tau E_S \hat{\eta}_{l,2i}) + \eta_{j,l,2i}^2(1+\tau E_S \hat{\eta}_{l,2i-1}))}{\tau E_S \hat{\eta}_{l,2i-1}^2 \hat{\eta}_{l,2i}^2}} \right]^{\frac{1}{2}}. \quad (3.23)$$

The proof of (3.23) is based on the results in Appendix A.3. Similar to the previous case, the received signal at $U_{l,k'}$ is given by (3.7). Limits on each term of (3.7) are used to obtain the SINR value at $U_{l,k'}$ as

$$\lim_{N \rightarrow \infty} \gamma_{l,k'} = \frac{\Psi_l^2 E_S \eta_{l,l,k'}^2 \eta_{l,l,k}^2}{\Psi_l^2 E_S \eta_{l,l,k'}^2 \sum_{j=1, j \neq l}^L \eta_{j,l,k}^2 + \hat{\eta}_{l,k}^4 \sigma_{n_{l,k'}}^2}, \quad (3.24)$$

where $\Psi_l = \lim_{N \rightarrow \infty} \hat{\beta}_l$ and given in (3.23).

3.3.3 Transmit power scaling at the user nodes and relay

This section obtains the asymptotic SINR for the transmit power scaling at the relay and user nodes (where $P_S = E_S/\sqrt{N}$, $P_P = \tau E_S/\sqrt{N}$, $P_R = E_R/\sqrt{N}$, and E_S and E_R are fixed). As in the previous section, the proofs are omitted due to their similarity to the results in Section 3.3.1. The asymptotic value of the normalized relay gain is obtained as

$$\lim_{N \rightarrow \infty} \frac{\hat{\beta}_l}{\sqrt[4]{N}} = \left[\frac{E_R}{E_S \sum_{j=1}^L \sum_{i=1}^K \left(\frac{\eta_{j,l,2i-1}^2 + \eta_{j,l,2i}^2}{\tau E_S \hat{\eta}_{l,2i-1}^2 \hat{\eta}_{l,2i}^2} \right) + 2\sigma_{R_l}^2 \sum_{i=1}^K (\tau E_S \hat{\eta}_{l,2i-1} \hat{\eta}_{l,2i})^{-2}} \right]^{\frac{1}{2}}. \quad (3.25)$$

By substituting the E_S values, (3.7) is rewritten for the received signal as

$$\begin{aligned} y_{l,k'} &= \frac{\hat{\beta}_l}{\sqrt[4]{N}} \sqrt{E_S} \mathbf{g}_{lk'}^T \hat{\mathbf{W}}_l \mathbf{g}_{lk} x_{l,k} \\ &+ \frac{\hat{\beta}_l}{\sqrt[4]{N}} \sqrt{E_S} \mathbf{g}_{lk'}^T \hat{\mathbf{W}}_l \sum_{m=1, m \neq k}^{2K} \mathbf{g}_{lm} x_{l,m} + \frac{\hat{\beta}_l}{\sqrt[4]{N}} \sqrt{E_S} \mathbf{g}_{lk'}^T \hat{\mathbf{W}}_l \sum_{j=1, j \neq l}^L \mathbf{G}_j \mathbf{x}_j \\ &+ \frac{\hat{\beta}_l}{\sqrt[4]{N}} \sqrt[4]{N} \mathbf{g}_{lk'}^T \hat{\mathbf{W}}_l \mathbf{n}_{R_l} + n_{l,k'}. \end{aligned} \quad (3.26)$$

By using the same procedure as in Section 3.3.1, the SINR is obtained as

$$\lim_{N \rightarrow \infty} \gamma_{l,k'} = \frac{\tau \Lambda_l^2 E_S^2 \eta_{l,l,k'}^2 \eta_{l,l,k}^2}{\tau \Lambda_l^2 E_S^2 \eta_{l,l,k'}^2 \sum_{j=1, j \neq l}^L \eta_{j,l,k}^2 + \Lambda_l^2 \eta_{l,l,k'}^2 \sigma_{R_l}^2 + \tau E_S \hat{\eta}_{l,k}^4 \sigma_{n_{l,k'}}^2}, \quad (3.27)$$

where $\Lambda_l = \lim_{N \rightarrow \infty} \frac{\hat{\beta}_l}{\sqrt{4N}}$ and given in (3.25).

Remark V.1: It can be seen that the asymptotic SINRs (3.22), (3.24), and (3.27) are affected by the pilot contamination. Interestingly, whenever $\eta_{j,l,k} = 0$ for $j \neq l$ or when $L = 1$, the asymptotic SINRs results in (3.22), (3.24), and (3.27) approaches the same SINRs provided in [65]. Furthermore, by substituting asymptotic SINRs (3.22), (3.24), and (3.27) into (3.11) and (3.12), the overall sum rate and the energy efficiency of the system is obtained.

3.4 The optimal pilot sequence length

This section analyses the optimal pilot sequence length to maximize the sum rate of the system for the case with $L = 1$ under power scaling at the user nodes scenario. For this case, (3.22) can be written as

$$\lim_{N \rightarrow \infty} \gamma_{l,k'} = \frac{\tau E_S^2 \eta_{l,l,k}^2}{\sigma_{R_l}^2} = \tau \xi_{l,k}, \quad (3.28)$$

where $\xi_{l,k} = \frac{E_S^2 \eta_{l,l,k}^2}{\sigma_{R_l}^2}$. By substituting this value into (3.11), the overall sum rate of the system is written as

$$\mathcal{R}_l = \frac{(T_C - \tau)}{2T_C} \sum_{k=1}^{2K} \log(1 + \tau \xi_{l,k}). \quad (3.29)$$

By taking the partial derivation of (3.29) with respect to τ and by equating it to zero, the following equation is obtained.

$$\sum_{k=1}^{2K} \frac{(T_C - \tau) \xi_{l,k}}{1 + \tau \xi_{l,k}} = \sum_{k=1}^{2K} \ln(1 + \tau \xi_{l,k}). \quad (3.30)$$

Since the total sum rate can be maximized by maximizing the sum rates of individual user nodes, above equation can be rewritten and rearranged as

$$\exp\left(\frac{(T_C - \tau) \xi_{l,k}}{1 + \tau \xi_{l,k}} + 1\right) = (1 + \tau \xi_{l,k}) \exp(1), \quad (3.31)$$

$$\frac{1 + T_C \xi_{l,k}}{1 + \tau \xi_{l,k}} \exp\left(\frac{1 + T_C \xi_{l,k}}{1 + \tau \xi_{l,k}}\right) = (1 + T_C \xi_{l,k}) \exp(1). \quad (3.32)$$

By using the Lambert-W function [110], (3.32) is solved as

$$\mathbb{W}([1 + T_C \xi_{l,k}] \exp(1)) = \frac{1 + T_C \xi_{l,k}}{1 + \tau \xi_{l,k}}. \quad (3.33)$$

The optimal value for τ to maximize the sum rate can be derived as

$$\tau_k^* = \left\lfloor \frac{1}{\xi_{l,k}} \left(\frac{1 + T_C \xi_{l,k}}{\mathbb{W}([1 + T_C \xi_{l,k}] \exp(1))} - 1 \right) \right\rfloor, \quad (3.34)$$

where $\lfloor \cdot \rfloor$ is the floor function. Analysis on the second partial derivative of R_l with respect to τ shows that the solution in (3.34) maximizes the sum rate of the system.

3.5 Analysis for a finite number of antennas

This section derives the approximate closed-form sum rate for the three power scaling laws at the user nodes and relay with a finite number of antennas. Moreover, for mathematical tractability, the analysis is limited for the CCI-free, perfect CSI and no pilot contamination case. Furthermore, the cumulative distribution function (CDF) and the probability distribution function (PDF) of the end-to-end SINR at $U_{l,k'}$ are obtained. For this case, the beamforming matrix $\hat{\mathbf{W}}_l$ can be rewritten as

$$\hat{\mathbf{W}}_l = \mathbf{W}_l = \mathbf{G}_l^* (\mathbf{G}_l^T \mathbf{G}_l^*)^{-1} \mathbf{P} (\mathbf{G}_l^H \mathbf{G}_l)^{-1} \mathbf{G}_l^H, \quad (3.35)$$

By using $\hat{\mathbf{W}}_l$, the amplification factor β_l is given as

$$\beta_l = \sqrt{\frac{P_R}{P_S \mathbb{E} \left[\text{Tr} \left([\mathbf{G}_l^T \mathbf{G}_l^*]^{-1} \right) \right] + \sigma_{R_l}^2 \mathbb{E} \left[\text{Tr} \left([\mathbf{G}_l^H \mathbf{G}_l]^{-1} \mathbf{P} [\mathbf{G}_l^T \mathbf{G}_l^*]^{-1} \mathbf{P} \right) \right]}}. \quad (3.36)$$

Using \mathbf{W}_l and the absence of co-channel interfering TWRNs, (3.6) is further simplified as

$$y_{l,k'} = \beta_l \sqrt{P_S} x_{l,k} + \beta_l \mathbf{1}_{k'} \mathbf{P} (\mathbf{G}_l^H \mathbf{G}_l)^{-1} \mathbf{G}_l^H \mathbf{n}_{R_l} + n_{l,k'}, \quad (3.37)$$

where $\mathbf{1}_{k'}$ represents a $1 \times 2K$ vector with value 1 at the k' location and zeros in all other places. Thus the end-to-end SINR at $U_{l,k'}$, $\gamma_{l,k'}$ is derived as

$$\gamma_{l,k'} = \frac{\beta_l^2 P_S}{\beta_l^2 \sigma_{R_l}^2 \left[(\mathbf{G}_l^H \mathbf{G}_l)^{-1} \right]_{k'} + \sigma_{n_{l,k'}}^2}, \quad (3.38)$$

where $\left[(\mathbf{G}_l^H \mathbf{G}_l)^{-1} \right]_{k'}$ is the k' th diagonal entry of the matrix $(\mathbf{G}_l^H \mathbf{G}_l)^{-1}$. To begin the derivation of the SINR value, the long-term power amplification factor $\tilde{\beta}_l$ is used. Here, the following matrix identities from [111] are used.

$$\text{Tr} \left(\mathbb{E} \left[(\mathbf{G}_l^H \mathbf{G}_l)^{-1} \right] \right) = \frac{\text{Tr}(\mathbf{D}_l^{-1})}{N - 2K} = \frac{\sum_{i=1}^{2K} \eta_{l,i}^{-1}}{N - 2K}. \quad (3.39)$$

$$\begin{aligned} \text{Tr} \left(\mathbb{E} \left[\left(\mathbf{G}_{ll}^H \mathbf{G}_{ll} \right)^{-1} \mathbf{P} \left(\mathbf{G}_{ll}^T \mathbf{G}_{ll}^* \right)^{-1} \mathbf{P} \right] \right) &= \frac{\text{Tr} \left(\mathbf{P} \mathbf{D}_{ll}^{-1} \mathbf{P} \mathbf{D}_{ll}^{-1} \right)}{(N-2K)^2 - 1} + \frac{\text{Tr} \left(\mathbf{P} \text{Tr} \left(\mathbf{P} \mathbf{D}_{ll}^{-1} \right) \mathbf{D}_{ll}^{-1} \right)}{\left((N-2K)^2 - 1 \right) (N-2K)} \\ &= \frac{2 \sum_{i=0}^K (\eta_{l,l,2i-1} \eta_{l,l,2i})^{-1}}{(N-2K)^2 - 1}. \end{aligned} \quad (3.40)$$

By using the above equations, the value of $\tilde{\beta}_l$ is rewritten as

$$\tilde{\beta}_l^2 = \frac{\left((N-2K)^2 - 1 \right) (N-2K) P_R}{\left((N-2K)^2 - 1 \right) P_S \sum_{i=1}^{2K} \eta_{l,l,i}^{-1} + 2(N-2K) \sigma_{R_l}^2 \sum_{i=0}^K (\eta_{l,l,2i-1} \eta_{l,l,2i})^{-1}}. \quad (3.41)$$

By substituting the value of $\tilde{\beta}_l$ to (3.38), the approximated SINR is obtained as

$$\tilde{\gamma}_{l,k'} = \frac{\alpha_{l,k'} X}{\eta_{l,k'} X + \zeta_{l,k'}}, \quad (3.42)$$

where $X = \left(\left[\left(\mathbf{G}_{ll}^H \mathbf{G}_{ll} \right)^{-1} \right]_{k',k'} \right)^{-1}$ and other symbols are given as follows:

$$\alpha_{l,k'} = \left((N-2K)^2 - 1 \right) (N-2K) P_R P_S. \quad (3.43)$$

$$\zeta_{l,k'} = \left((N-2K)^2 - 1 \right) (N-2K) P_R \sigma_{R_l}^2. \quad (3.44)$$

$$\begin{aligned} \eta_{l,k'} &= \left((N-2K)^2 - 1 \right) P_S \sigma_{n_{l,k'}}^2 \sum_{i=1}^{2K} \eta_{l,l,i}^{-1} \\ &+ 2(N-2K) \sigma_{n_{l,k'}}^2 \sigma_{R_l}^2 \sum_{i=0}^K (\eta_{l,l,2i-1} \eta_{l,l,2i})^{-1}. \end{aligned} \quad (3.45)$$

By using distribution of the k th diagonal element of the inverse Wishart matrix [112], the CDF of $\gamma_{l,k'}$ was obtained as [72]:

$$F_{\tilde{\gamma}_{l,k'}}(x) = \begin{cases} 1 - \frac{\Gamma \left(N-2K+1, \frac{\zeta_{l,k'} x}{\alpha_{l,k'} - \eta_{l,k'} x} \right)}{\Gamma(N-2K+1)}, & 0 < x < \frac{\alpha_{l,k'}}{\eta_{l,k'}} \\ 1, & x \geq \frac{\alpha_{l,k'}}{\eta_{l,k'}}. \end{cases} \quad (3.46)$$

By differentiating (3.46) by using the Leibniz integral rule the PDF is obtained as

$$\begin{aligned} f_{\tilde{\gamma}_{l,k'}}(x) &= \frac{d}{dx} \left[\frac{\zeta_{l,k'} x}{\alpha_{l,k'} - \eta_{l,k'} x} \right] \left(\frac{\zeta_{l,k'} x}{\alpha_{l,k'} - \eta_{l,k'} x} \right)^{N-2K} \frac{e^{-\frac{\zeta_{l,k'} x}{\alpha_{l,k'} - \eta_{l,k'} x}}}{\Gamma(N-2K+1)} \\ &= \frac{\alpha_{l,k'} (\zeta_{l,k'})^{N-2K+1} x^{N-2K} e^{-\frac{\zeta_{l,k'} x}{\alpha_{l,k'} - \eta_{l,k'} x}}}{\Gamma(N-2K+1) (\alpha_{l,k'} - \eta_{l,k'} x)^{N-2K+2}}, \end{aligned} \quad (3.47)$$

where $0 \leq x < \frac{\alpha_{l,k'}}{\eta_{l,k'}}$. The average sum rate can be approximated by solving $\bar{R}_l = \frac{1}{2 \ln(2)} \int_0^\infty \ln(1+x) f_{\gamma_{l,k'}}(x) dx$ as

$$\bar{R}_{L^*} \approx \frac{1}{2 \ln(2)} \frac{1}{(N-2K)!} \mathbb{I}_1, \quad (3.48)$$

where \mathbb{I}_1 is defined as follows:

$$\begin{aligned} \mathbb{I}_1 = & \alpha_{l,k'} \zeta_{l,k'}^{N-2K+1} \int_0^{\frac{\alpha_{l,k'}}{\eta_{l,k'}}} \frac{x^{N-2K}}{(\alpha_{l,k'} - \eta_{l,k'} x)^{N-2K+2}} \\ & \times \exp\left(-\frac{\zeta_{l,k'} x}{\alpha_{l,k'} - \eta_{l,k'} x}\right) \ln(1+x) dx, \end{aligned} \quad (3.49)$$

By substituting the dummy variable $t = \zeta_{l,k'} x / (\alpha_{l,k'} - \eta_{l,k'} x)$ into (3.49) the integral \mathbb{I}_1 can be simplified as

$$\mathbb{I}_1 = \int_0^\infty t^{N-2K} e^{-t} \ln\left(\frac{\zeta_{l,k'} + (\alpha_{l,k'} + \eta_{l,k'})t}{\zeta_{l,k'} + \eta_{l,k'} t}\right) dt, \quad (3.50)$$

Next, \mathbb{I}_1 in (3.50) can be solved in closed-form as follows:

$$\mathbb{I}_1 = \mathbb{J}(N-2K, \zeta_{l,k'}, \alpha_{l,k'} + \eta_{l,k'}) - \mathbb{J}(N-2K, \zeta_{l,k'}, \eta_{l,k'}), \quad (3.51)$$

where the function $\mathbb{J}(x, y, z)$ is defined as

$$\begin{aligned} \mathbb{J}(x, y, z) &= \int_0^\infty \lambda^x \exp(-\lambda) \ln(y+z\lambda) d\lambda \\ &= \Gamma(x+1) \times \\ & \left(\ln(y) + \sum_{p=0}^x \frac{1}{\Gamma(x-p+1)} \left(\left(\frac{-y}{z}\right)^{x-p} \exp\left(\frac{y}{z}\right) E_1\left(\frac{y}{z}\right) + \sum_{q=1}^{x-p} \Gamma(q) \left(\frac{-y}{z}\right)^{x-p-q} \right) \right) \end{aligned} \quad (3.52)$$

By substituting (3.51) into (3.48), an approximation of the average sum rate can be derived in closed-form as:

$$\bar{R}_{L^*} = \frac{\mathbb{J}(N-2K, \zeta_{l,k'}, \alpha_{l,k'} + \eta_{l,k'}) - \mathbb{J}(N-2K, \zeta_{l,k'}, \eta_{l,k'})}{2 \ln(2) (N-2K)!}. \quad (3.53)$$

The sum rates under different power scaling scenarios can be obtained by substituting the P_S and P_R values to $\alpha_{l,k'}$, $\zeta_{l,k'}$ and $\eta_{l,k'}$. The results obtained in (3.53) will be useful to identify the number of antennas required to obtain a certain percentage of the asymptotic sum rate.

3.6 Asymptotic analysis for relay-antenna correlation

This section derives asymptotic results for the SINR and sum-rate under the antenna correlation at the relay nodes under the three power scaling scenarios identified in

Section 3.3. Moreover, for mathematical tractability the analysis is limited for the CCI-free, perfect CSI and no pilot contamination case.

When there is antenna correlation at the relay the channel vector to the relay from the user k is written as $\bar{\mathbf{g}}_{ll,k}^T = (\Psi_{l,k})^{\frac{1}{2}} \mathbf{f}_{ll,k}^T d_{lk}^{\frac{1}{2}}$, where $\Psi_{l,k}$ is the $N \times N$ correlation matrix at the relay. Furthermore, $\mathbf{f}_{ll,k}^T$ is the k th column vector in the matrix \mathbf{F}_{ll} and d_{lk} is the k th diagonal entry of the matrix \mathbf{D}_{ll} that is given in 3.2.1. Accordingly the channel matrix between all relay and all the users is given as $\bar{\mathbf{G}}_{ll} = [\bar{\mathbf{g}}_{ll,1}^T \quad \bar{\mathbf{g}}_{ll,2}^T \quad \cdots \quad \bar{\mathbf{g}}_{ll,2K}^T]$. This corresponds to the max-semi-correlated Rayleigh fading scenario presented in [34,38]. For this case, the beamforming matrix $\hat{\mathbf{W}}_l$ can be rewritten as

$$\bar{\mathbf{W}}_l = \bar{\mathbf{G}}_{ll}^* (\bar{\mathbf{G}}_{ll}^T \bar{\mathbf{G}}_{ll}^*)^{-1} \mathbf{P} (\bar{\mathbf{G}}_{ll}^H \bar{\mathbf{G}}_{ll})^{-1} \bar{\mathbf{G}}_{ll}^H, \quad (3.54)$$

By using \mathbf{W}_l , the amplification factor $\bar{\beta}_l$ is given as

$$\bar{\beta}_l = \sqrt{\frac{P_R}{P_S \text{Tr} \left(\mathbb{E} \left[\left[\bar{\mathbf{G}}_{ll}^T \bar{\mathbf{G}}_{ll}^* \right]^{-1} \right] \right) + \sigma_{R_l}^2 \text{Tr} \left(\mathbb{E} \left[\left[\bar{\mathbf{G}}_{ll}^H \bar{\mathbf{G}}_{ll} \right]^{-1} \mathbf{P} \left[\bar{\mathbf{G}}_{ll}^T \bar{\mathbf{G}}_{ll}^* \right]^{-1} \mathbf{P} \right] \right)}}. \quad (3.55)$$

By using the similar steps as in Section 3.5, the end-to-end SINR at $U_{l,k'}$, $\gamma_{l,k'}$ is obtained as

$$\gamma_{l,k'} = \frac{\bar{\beta}_l^2 P_S}{\bar{\beta}_l^2 \sigma_{R_l}^2 \left[\left(\bar{\mathbf{G}}_{ll}^H \bar{\mathbf{G}}_{ll} \right)^{-1} \right]_{k'} + \sigma_{n_{l,k'}}^2}, \quad (3.56)$$

where $\left[\left(\bar{\mathbf{G}}_{ll}^H \bar{\mathbf{G}}_{ll} \right)^{-1} \right]_{k'}$ is the k' th diagonal entry of the matrix $\left(\bar{\mathbf{G}}_{ll}^H \bar{\mathbf{G}}_{ll} \right)^{-1}$.

To analyse the asymptotic performance of antenna correlation at the relay, the limit results relevant to the channel matrix $\bar{\mathbf{G}}_{ll}$ are considered. Here,

$$\left[\frac{\bar{\mathbf{G}}_{ll}^H \bar{\mathbf{G}}_{ll}}{N} \right]_{i,j} = d_{li}^{1/2} \frac{\mathbf{f}_{ll,i}^H \left(\Psi_{l,i}^H \Psi_{l,j} \right)^{1/2} \mathbf{f}_{ll,j}}{N} d_{lj}^{1/2}. \quad (3.57)$$

By using the limit results, it can be shown that if $i \neq j$, then the value of (3.57) goes to zero. If $i = j$ then the above value equals $\text{Tr}(\Psi_{l,j})$, which is equal to N for correlation matrices⁴. Based on this, the limit result can be given as

$$\left[\frac{\bar{\mathbf{G}}_{ll}^H \bar{\mathbf{G}}_{ll}}{N} \right] \xrightarrow[N \rightarrow \infty]{a.s.} \mathbf{D}_{ll}, \quad (3.58)$$

⁴It is assumed that the correlation matrices have full rank. However, in mMIMO channels, the correlation matrices can be rank-deficient [113]. In this context, the similar results will hold if the spectrum of the correlation matrices follow the conditions in [114, Theorem 3.4].

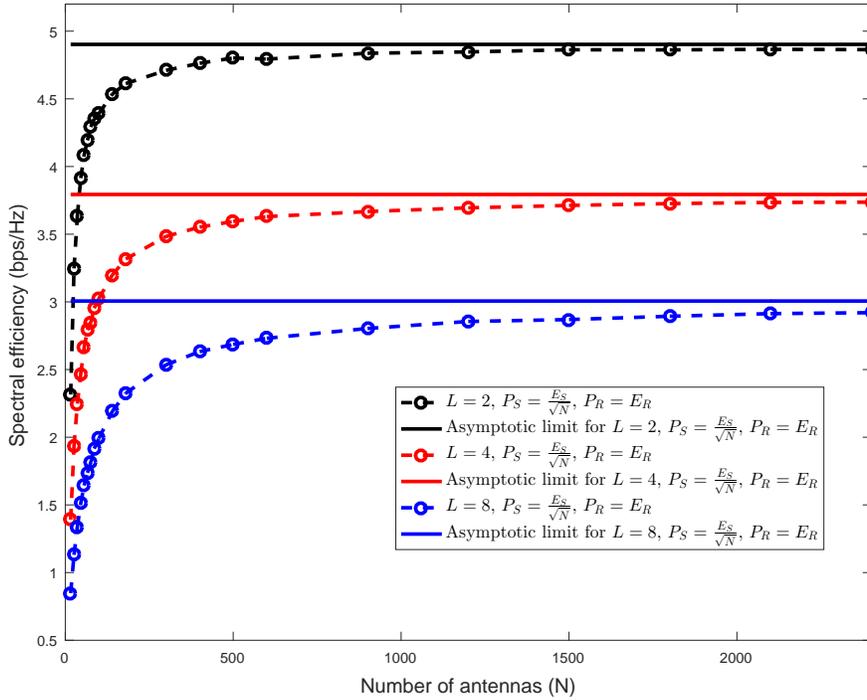


Figure 3.1: Spectral efficiency versus the number of relay antennas of an 12-user TWRN with different L values. The channels in \mathbf{G}_{jl} are i.i.d. Rayleigh RVs with $\mathbf{D}_u = \mathbf{I}_{2K}$ and $\mathbf{D}_{jl} = \frac{1}{2}\mathbf{I}_{2K}$, where $j, l \in \{1, \dots, L\}$ and $j \neq l$.

and coincidentally the asymptotic results for the case with antenna correlation is equal to the results obtained for the case without any antenna correlation. This shows that by using mMIMO, the degenerative effect of antenna correlation can be removed in the system model.

3.7 Simulation Results

This section presents the simulation results and comparisons with the derived asymptotic results. The power at the user nodes and the relay nodes is taken as $E_S = E_R = 10$, and the noise powers at user and relay nodes is taken as $\sigma_{n_{l,k}}^2 = \sigma_{n_R}^2 = 1$. The pathloss exponent η is assumed to be two. The normalized pilot sequence power factor τ/T_C is 0.8. Spectral and energy efficiencies under different power scaling scenarios, different K values, and different L values are presented in the sequel. What is the effect of having multiple TWRNs on the spectral efficiency and the energy efficiency? In Fig. 3.1 and Fig. 3.2, the spectral efficiency and the energy efficiency

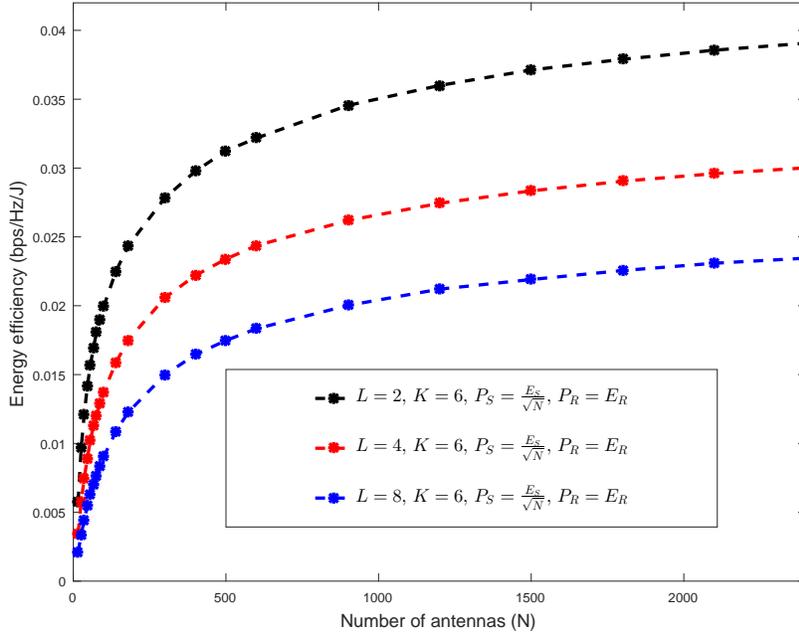


Figure 3.2: Energy efficiency versus the number of relay antennas of 12-user, L TWRNs. The channels in \mathbf{G}_{jl} are i.i.d. Rayleigh RVs with $\mathbf{D}_{ll} = \mathbf{I}_{2K}$ and $\mathbf{D}_{jl} = \frac{1}{2}\mathbf{I}_{2K}$, where $j, l \in \{1, \dots, L\}$ and $j \neq l$.

are presented for power scaling at user nodes (case 1), for $L = 2$, $L = 4$ and $L = 8$ values, when $K = 6$, respectively. The obtained asymptotic values (3.22) are also plotted for comparison. It can be seen in Fig. 3.1 that for all L values, the spectral efficiency asymptotically reaches the analytical results, validating the derived results. Furthermore, as the number of TWRNs (L) in the system is increased, the achievable spectral efficiency and the energy efficiency of a single TWRN decreases due to the interference and pilot contamination introduced by other TWRNs. As an example, a $L = 2$ system can obtain 4.9 bps/Hz efficiency while an $L = 8$ system can only achieve a spectral efficiency of 3 bps/Hz. However, if the spectral efficiency of the whole system is considered (by multiplying the spectral efficiency of a single TWRN by L), it can be concluded that the bandwidth can be utilized further by increasing the number of TWRNs. According to the values obtained in Fig. 3.1, the total spectral efficiency is 9.8 bps/Hz when $L = 2$ and approximately 15 bps/Hz when $L = 8$. Thus it can be concluded that by using multi-pair mMIMO TWRNs for pairwise communications between nodes, that the limited bandwidth can be utilized effectively.

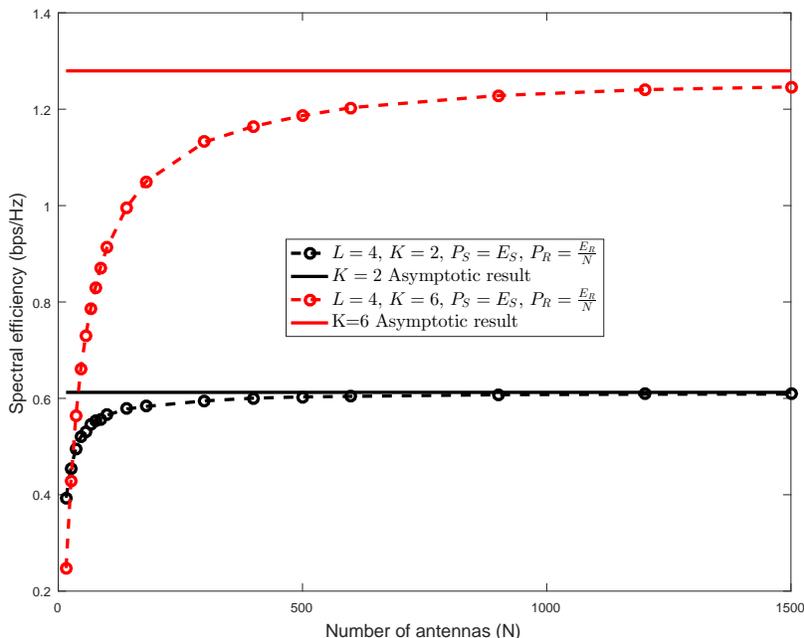


Figure 3.3: Spectral efficiency versus the number of relay antennas of a $L = 4$ TWRN systems with different number of user pairs. The channel gains of \mathbf{G}_{jl} are i.i.d. Rayleigh RVs. with $\mathbf{D}_l = \mathbf{I}_{2K}$ and $\mathbf{D}_{j,l} = \frac{1}{2}\mathbf{I}_{2K}$, where $j, l \in \{1, \dots, L\}$ and $j \neq l$.

The effect of number of users in a single system ($2K$) is analysed in Fig. 3.3. Specifically, the sum rate is plotted for a system with eight relay networks under power scaling at the relay nodes for $K = 2$ and $K = 6$ (case 2). A four-user TWRN achieves 0.62 bps/Hz while a 12-user TWRN obtains 1.28 bps/Hz. Fig. (3.24) shows that the analytical results match the simulated values. Note that the spectral efficiency increases as the number of users increases, as the same bandwidth is used by the additional users. Thus, a mMIMO pairwise TWRN improves the bandwidth utilization by serving more users. However, there will be countering factors that will limit the number of user pairs in a network, such as the number of available orthogonal pilot sequences which is limited by the coherence time of the system.

The spectral efficiency gains and energy efficiency gains of different power scaling scenarios are compared in Fig. 3.4 and Fig. 3.5 for 12-user TWRNs ($L = 8$). The analytical results (3.22), (3.24), and (3.27) are also plotted for comparison purposes. Power scaling at the user nodes has the highest asymptotic of 7.2 bps/Hz out of

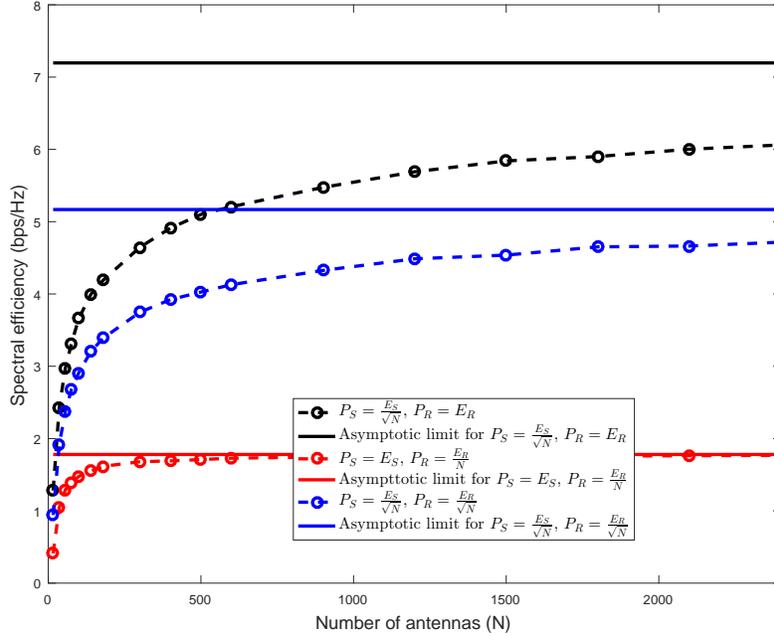


Figure 3.4: Spectral efficiency versus the number of relay antennas of 12-user, $L = 8$ TWRNs. The channels in \mathbf{G}_{jl} are i.i.d. Rayleigh RVs with $\mathbf{D}_{ll} = \mathbf{I}_{2K}$ and $\mathbf{D}_{jl} = \frac{1}{2}\mathbf{I}_{2K}$, where $j, l \in \{1, \dots, L\}$ and $j \neq l$.

all the three cases while case-2 and case-3 achieved asymptotically 5.2 bps/Hz and 1.8 bps/Hz, respectively. Moreover, in Fig. 3.5 power scaling at both user and relay nodes obtains the highest energy efficiency as expected. Furthermore, the energy efficiency of case-2 (power scaling at the relay only) is very low compared to other two cases. This result is expected as the power of user nodes are kept unchanged in this scenario. Thus the numerator in (3.12) is relatively constant for different N values, and thus the power efficiency will be low. How accurate are the analytical results when the number of antennas is finite? In Fig. 3.6, the sum rate results are plotted to answer this question. The simulated sum rate values match with closed-form results in (3.53), justifying the accuracy of the approximation. For example, as few as 14 relay antennas yields about 85% of the asymptotic performance ($N = \infty$). Moreover, increasing to $N = 22$ relay antennas yields a 92%. This is good news because more or less the performance of mMIMO is possible with a finite number of antennas. How much degradation of the sum rate occurs due to antenna correlation? In Fig. 3.7, the antenna correlation model in [115, Eqn. (4)] is used. Here, the (p, q) th element of correlation matrix is given as $e^{-j2\pi(p-q)l \cos(\theta)} e^{-\frac{1}{2}(2\pi(p-q)l \sin(\theta)\sigma)^2}$ [38], where l is

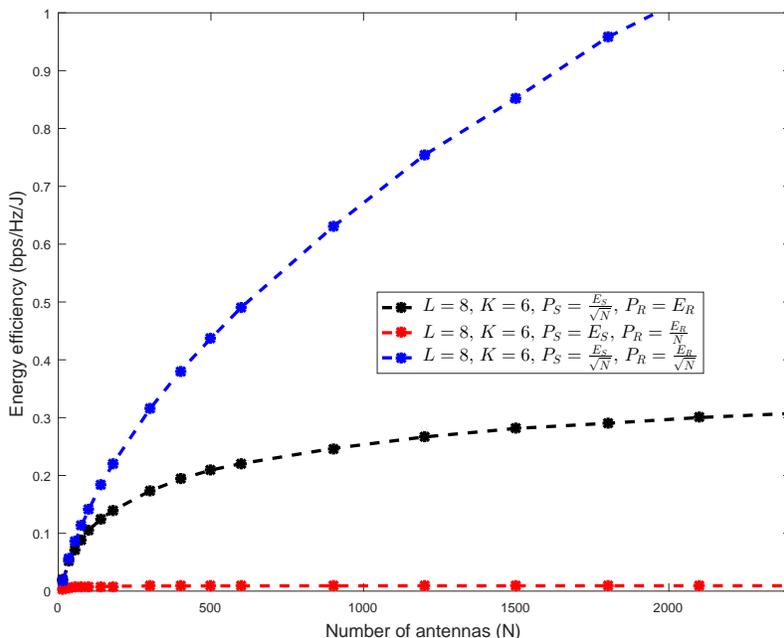


Figure 3.5: Energy efficiency versus the number of relay antennas of 12-user, 8 TWRNs. The channels in \mathbf{G}_{jl} are i.i.d. Rayleigh RVs with $\mathbf{D}_{ll} = \mathbf{I}_{2K}$ and $\mathbf{D}_{jl} = \frac{1}{2}\mathbf{I}_{2K}$, where $j, l \in \{1, \dots, L\}$ and $j \neq l$.

the relative antenna spacing, θ is the average angle of arrival/departure, and σ is the standard deviation of the angle of arrival/departure. The sum rate of the system under low correlation and high correlation at the relay as well as with no correlation is plotted. Fig. 3.7 shows that although antenna correlation degrades the sum rate, as the number of relay antennas increases, it approaches to that of the uncorrelated antenna case. This observation corroborates the analysis that antenna correlation can be mitigated by a massive antenna array at the relay.

3.8 Conclusion

This chapter investigated the impact of several key impairments, namely co-channel interference, imperfect CSI, antenna correlation and pilot contamination for multi-pair mMIMO TWRNs. The asymptotic SINR, sum rate, and energy efficiency are derived for the three transmit power scaling laws. Importantly, it is shown that the user transmit power can be scaled down inversely proportionally to the square-root of the number of relay antennas. Even better, if power scaling is limited to the relay node, it can be inversely proportional to the number of relay antennas without any

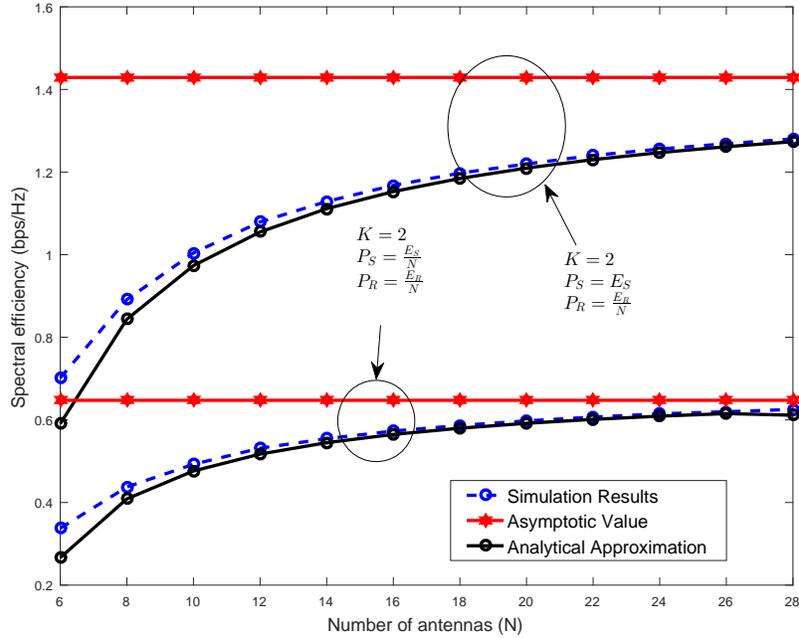


Figure 3.6: Sum rate versus the number of relay antennas of a 4-user TWRN ($L = 1$) under different power scaling scenarios. The channel gains of \mathbf{G}_{ll} are i.i.d. Rayleigh RVs.

performance penalty. The analytical and simulation results reveal that substantial sum-rate gains and energy-efficiency gains can be achieved by adopting mMIMO at the relay. Also, the closed-form results for a finite number of relay antennas may help wireless researchers to estimate the number of antennas required to obtain a certain percentage of the asymptotic sum rate. Furthermore, the results in this chapter show that mMIMO is an antidote to the degenerative effects of antenna correlation.

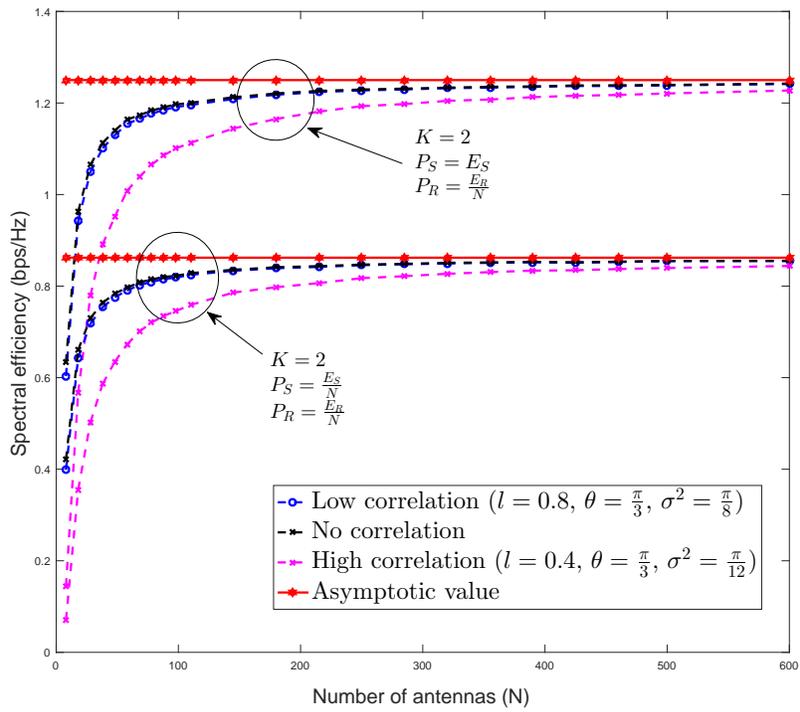


Figure 3.7: Sum-rate versus the number of relay antennas of a 4-user TWRN ($L = 1$) under relay antenna correlation.

Chapter 4

Relay selection in massive multiple-input multiple-output (mMIMO) cognitive two-way relay networks (TWRNs)

This chapter analyses a relay-assisted wireless communication link between two underlay mMIMO terminals. Specifically, an amplify-and-forward (AF) relay is optimally selected to maximize the sum rate and to keep the interference on the primary user (PU) below an interference threshold. Signal-to-interference-plus-noise ratios (SINRs) are obtained for two scenarios: (1) the relays and the two end nodes use transmit power scaling and (2) only the end nodes use transmit power scaling. For these two cases, optimal power allocations to satisfy PU interference are derived. With these optimal power allocations, the impact of relay selection on the outage, the sum rate, and the energy efficiency of the network is analysed. For the first scenario, the outage can be reduced to zero with appropriate power allocation and the relay selection can be done offline. For the second scenario, outage will depend on the instantaneous channel-state information (CSI) between the relays and the PU. Furthermore, the energy efficiency of the system is obtained by using a realistic power consumption model and it is shown that relay selection increases the energy efficiency.

4.1 Introduction

Several significant problems are faced by cognitive radio (CR) TWRNs. These problems are due to the fact their transmit powers must not cause too much interference

at the primary users. During the operation of the TWRN, interference on the PU is generated over two time slots. In the first one, both the end nodes generate it and in the second slot, the relay will generate it. How to control this interference on the PU so that the outage of the underlay network is avoided? What is the limiting behaviour of interference for a large number of antennas? Does relay selection help to reduce the interference? How to scale down the transmit powers of the end nodes as a function of the number of antennas? What power allocation maximizes the sum rate while limiting the interference?

In order to answer some of these questions, the cognitive two-way relay selection problem of a particular network (Fig. 4.1) is analyzed in this chapter. The source and destination nodes S_1 and S_2 (also called the end nodes) are mMIMO-enabled and the relay nodes (R_1 to R_k) are multiple-input multiple-output (MIMO) enabled. S_1 and S_2 perform zero forcing (ZF) based transmission and receiving while enabling multiple data sub-streams and the relay performs amplify and forwarding. Furthermore, S_1 and S_2 perform self-interference cancellation. S_1 , S_2 , and the relay nodes all act as secondary users in the presence of a PU. Relay selection is performed to maximize the sum rate between S_1 and S_2 while limiting the interference on the PU. This system setup is ideal for Internet of Things (IoT) applications where networks are often deployed in an ad-hoc fashion. Such IoT networks operate in the underlay mode without affecting the licensed systems while improving energy efficiency [116].

In the considered network, the main interference issue arises during the first time slot when the two end nodes transmit simultaneously to the relay. This interference must be below the threshold at the PU. If not, the secondary system will go to outage. The key question is how to reduce or eliminate this outage. There are two immediate solutions. The first solution is to coordinate the transmissions of the two end nodes to reduce the joint interference. The problem is, however, that implementing coordination among two different end nodes is difficult. For instance, to do so, each end node will require the CSI between the other node and the PU. For example, S_1 needs to know the channel between S_2 and the PU node. This will require additional pilot overheads. The second solution is to use two separate time slots for S_1 and S_2 to transmit data to the relay. This solution avoids the cumulative interference, and it is proposed in [117, 118]. But the penalty is the 33% reduction of spectral efficiency as three time-slots are required for bidirectional data transfer.

Herein, an alternative solution without constraining the end users to transmit in

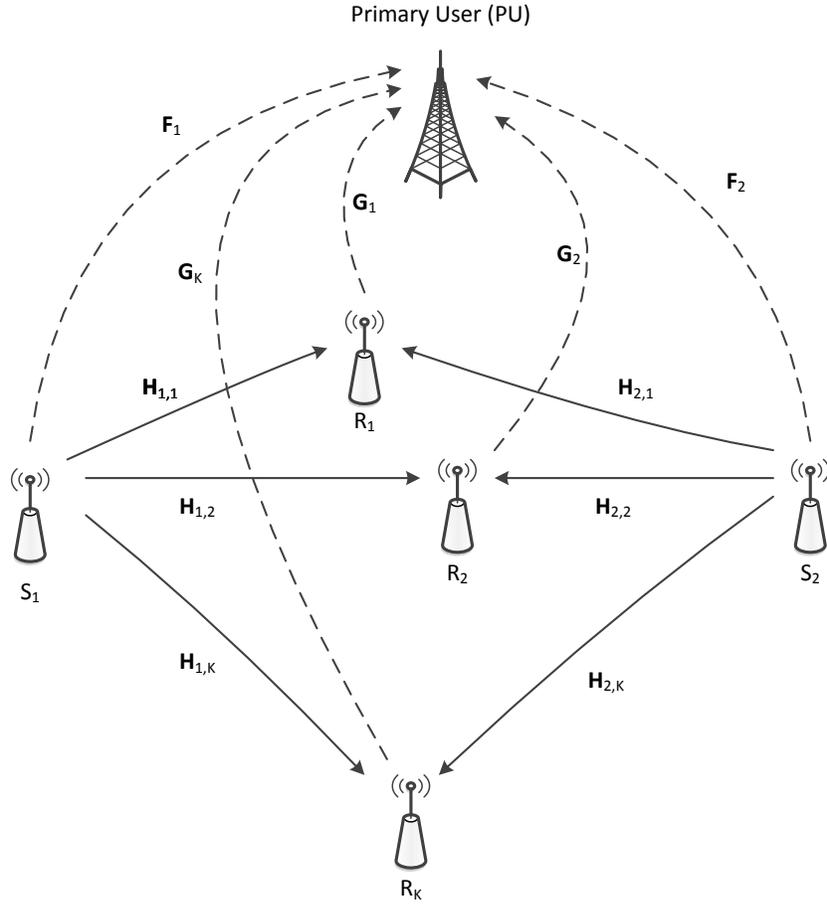


Figure 4.1: The network with K relay nodes and two end nodes, S_1 and S_2 .

different time slots is proposed. It is based on the use of a large number of antennas, power scaling, and power allocation. Furthermore, relay selection is employed to obtain significant sum rate and energy efficiency gains. Specifically, the objective is to maximize the sum rate of an underlay TWRN through relay selection subject to the interference constraints on the PU.

Contributions:

- The SINR and the sum rate for a certain data sub-stream are obtained, when a certain relay is selected.
- Two possible power scaling scenarios are proposed to limit the interference at the PU. In power scaling one (PS-1), the transmit powers at the end nodes and the relays are scaled down according to the number of antennas at the

end nodes. In power scaling two (PS-2), only the power at the end nodes is scaled down according to the number of antennas at the end nodes.

- Asymptotic sum rates under the proposed power scaling scenarios and the aggregated sum rate for the selected relay is obtained.
- Optimum power allocation to maximize the sum rate and to satisfy the maximum interference constraint at PU for each relay under the two power scaling methods is proposed.
- Three relay selection criteria are proposed to maximize the sum rate. Specifically, one relay selection method for PS-1 and two relay selection methods for PS-2 are proposed. These relay selection methods are analysed for their performance in terms of the achieved sum rate, the outage probability of the secondary system, and the requirement of instantaneous channel information.
- Energy efficiency is analysed under PS-1 and PS-2 as the number of antennas at the end nodes are increased. For this analysis, a practical model is used which takes the power consumption in transceiver chains and the computational power consumption at the end nodes into account.

4.2 System, channel, and signal model

4.2.1 System and channel model

The system model consists of one PU and the secondary network (Fig. 1). Secondary TWRN consists of two user nodes (S_1 and S_2) and K relay nodes (R_k for $k \in \{1, \dots, K\}$). The PU is equipped with N antennas and user node S_i is equipped with N_i antennas for $i \in \{1, 2\}$, and the k th relay node has N_{R_k} antennas. All secondary nodes are assumed to be half-duplex terminals, and all channel amplitudes are assumed to be independently distributed, frequency flat-Rayleigh fading. The wireless channel from S_i to PU is defined as $\mathbf{F}_i = \hat{\mathbf{D}}_i^{1/2} \tilde{\mathbf{F}}_i$, where $\tilde{\mathbf{F}}_i \sim \mathcal{CN}_{N \times N_i}(\mathbf{0}_{N \times N_i}, \mathbf{I}_N \otimes \mathbf{I}_{N_i})$ captures the fast fading and $\hat{\mathbf{D}}_i = \hat{\eta}_i \mathbf{I}_N$ accounts for the pathloss. The channel between R_k and PU is defined as $\mathbf{G}_k = \mathbf{D}_k^{1/2} \tilde{\mathbf{G}}_k$, with $\tilde{\mathbf{G}}_k \sim \mathcal{CN}_{N \times N_{R_k}}(\mathbf{0}_{N \times N_{R_k}}, \mathbf{I}_N \otimes \mathbf{I}_{N_{R_k}})$ and $\mathbf{D}_k = \eta_k \mathbf{I}_N$. Similarly, the channel matrix from S_i to R_k is defined as $\mathbf{H}_{i,k} = \mathbf{D}_{i,k}^{1/2} \tilde{\mathbf{H}}_{i,k}$, with $\tilde{\mathbf{H}}_{i,k} \sim \mathcal{CN}_{N_{R_k} \times N_i}(\mathbf{0}_{N_{R_k} \times N_i}, \mathbf{I}_{N_{R_k}} \otimes \mathbf{I}_{N_i})$ and $\mathbf{D}_{i,k} = \eta_{i,k} \mathbf{I}_{N_{R_k}}$. The detailed system model

is shown in Fig. 4.1. Here, the sets $\{\eta_k\}$, $\{\hat{\eta}_i\}$ and $\{\eta_{i,k}\}$ represent the pathloss components in the R_k -to-PU, S_i -to-PU and S_i -to- R_k channels.

The channel coefficients are assumed to be fixed during two consecutive time-slots (a time-slot is the time used for a single transmission between two wireless nodes) and, hence, the reverse channels are assumed to be the transpose of forward channel by using the reciprocity property of wireless channels. The additive noise at all the receivers is modelled as complex zero mean additive white Gaussian noise (AWGN). The direct channel between S_1 and S_2 is assumed to be unavailable due to large pathloss and heavy shadowing effects [32, 35].

4.2.2 Signal model

S_1 and S_2 exchange their signal vectors \mathbf{x}_1 and \mathbf{x}_2 by selecting one of the available relays using two time-slots. Here the selected relay is denoted as R_k for the analysis. First, S_1 and S_2 transmit \mathbf{x}_1 and \mathbf{x}_2 , respectively, towards R_k by employing transmit-ZF precoding over the multiple access channel¹. The received signal at R_k can then be written as

$$\mathbf{y}_{R_k} = m_{1,k} \mathbf{H}_{1,k} \mathbf{V}_{T_{1,k}} \mathbf{x}_1 + m_{2,k} \mathbf{H}_{2,k} \mathbf{V}_{T_{2,k}} \mathbf{x}_2 + \mathbf{n}_{R_k} + \mathbf{i}_{R_k}, \quad (4.1)$$

where the $N_{R_k} \times 1$ signal vector \mathbf{x}_i satisfies $\mathcal{E}[\mathbf{x}_i \mathbf{x}_i^H] = \mathbf{I}_{N_{R_k}}$ for $i \in \{1, 2\}$ and $k \in \{1, \dots, K\}$. Thus, the $N_i \times 1$ precoded-transmit signal vector at S_i is given by $\mathbf{V}_{T_{i,k}} \mathbf{x}_i$. Here, \mathbf{i}_{R_k} is the $N_{R_k} \times 1$ interference vector on the secondary relay by the PU, which is modelled as AWGN with average power $\sigma_{I_{R_k}}^2 = P_U \eta_k$, where P_U is the transmit power of PU.² In (4.1), $m_{i,k}$ is the power normalizing factor at S_i and is designed to constrain its transmit power as follows [35]:

$$m_{i,k} = \sqrt{P_{i,k} / \text{Tr}(\mathbf{V}_{T_{i,k}} \mathbf{V}_{T_{i,k}}^H)}, \quad (4.2)$$

where $P_{i,k}$ is the transmit power at S_i to satisfy the interference constraints at PU. Further, in (4.1), \mathbf{n}_{R_k} is the $N_{R_k} \times 1$ zero mean AWGN vector at R_k satisfying $\mathcal{E}(\mathbf{n}_{R_k} \mathbf{n}_{R_k}^H) = \mathbf{I}_{N_{R_k}} \sigma_{N_{R_k}}^2$, and $\mathbf{V}_{T_{i,k}}$ is the transmit-ZF precoding matrix at S_i given by [119]

$$\mathbf{V}_{T_{i,k}} = \mathbf{H}_{i,k}^H (\mathbf{H}_{i,k} \mathbf{H}_{i,k}^H)^{-1}. \quad (4.3)$$

¹In order to use transmit-ZF at S_1 and S_2 , the constraint $\min(N_1, N_2) \geq \max_{k \in \{1, \dots, K\}} N_{R_k}$ needs to be satisfied.

²Here, as no coordination between the primary and secondary networks is assumed and the average interference power at the relay when the PU is transmitting. Similar assumptions are made in [89].

In the second time-slot, R_k amplifies \mathbf{y}_{R_k} and broadcasts this amplified-signal towards both user nodes. Each node then performs ZF receiving and obtains the following signal vector:

$$\mathbf{y}_{S_{i,k}} = \mathbf{V}_{R_{i,k}} (M_k \mathbf{H}_{k,i} \mathbf{y}_{R_k} + \mathbf{n}_i + \mathbf{i}_i), \quad (4.4)$$

where M_k is the amplification factor at R_k and is defined as

$$M_k = \sqrt{P_{R_k} / (m_{1,k}^2 + m_{2,k}^2 + \sigma_{N_{R_k}}^2 + \sigma_{I_{R_k}}^2)}, \quad (4.5)$$

where $m_{i,k}$ is defined in (4.2) and P_{R_k} is the transmit power at R_k to satisfy the interference constraints at PU. Moreover, in (4.4), $\mathbf{H}_{k,i} = \mathbf{H}_{i,k}^T$, and \mathbf{n}_i is the $N_i \times 1$ zero mean AWGN at S_i satisfying $\mathcal{E}(\mathbf{n}_i \mathbf{n}_i^H) = \mathbf{I}_{N_i} \sigma_{N_i}^2$. Also, \mathbf{i}_i is the interference on S_i by PU with average power $\sigma_{I_i}^2 = P_U \hat{\eta}_i$. Besides, $\mathbf{V}_{R_{i,k}}$ is the receive-ZF matrix at S_i and can be written as [119]

$$\mathbf{V}_{R_{i,k}} = (\mathbf{H}_{k,i}^H \mathbf{H}_{k,i})^{-1} \mathbf{H}_{k,i}^H. \quad (4.6)$$

4.2.3 CSI requirements

The CSI in the system can be categorized as (1) CSI within the secondary system and (2) CSI between the PU and the secondary system. Since no coordination between the PU and the secondary system is assumed, acquiring type two CSI is harder than acquiring type one CSI (for related issues see [120–122] and references therein). The full CSI requirements can be listed as follows.

1. **CSI within the secondary system:** End node S_1 (S_2) requires the CSI between it and the selected relay R_k , $\mathbf{H}_{1,k}$ ($\mathbf{H}_{2,k}$) to perform transmit beamforming and receive beamforming. The selected relay R_k only performs the AF operation. Thus, it will only require the information about the amplification coefficients at the end nodes $m_{1,k}$ and $m_{2,k}$. Although $m_{1,k}$ ($m_{2,k}$) depends on the CSI between the S_1 and R_k (S_2 and R_k), full CSI is not required at the R_k . For self interference cancellation, M_k should be known by both the end nodes. This knowledge can be acquired by the end nodes through a broadcast from the relay. Alternatively, all the nodes can decide to use the long term average values for $m_{i,k}$ s and M_k . This will relax the CSI requirements on the system.

2. **CSI between the PU and secondary system:** Apart from the CSIs required for the data transmission between S_1 and S_2 , the underlay operation requires all the nodes to limit the interference on the PUs. Thus when R_k transmits its signal, it should have the CSI between itself and the PU. Furthermore, when the end nodes transmit simultaneously in the first time slot, the end nodes must have the CSI between S_1 and PU and also between S_2 and PU (i.e. both \mathbf{F}_1 and \mathbf{F}_2 should be known by S_1 and S_2). This work will show that this CSI is not required when mMIMO and appropriate power scaling methods are used.

4.2.4 Effect on the primary user

In underlay cognitive radio, the secondary users should transmit their data without exceeding the interference temperature at the PU. This section obtains the equations for the received interference at the PU during the two time-slots. During the first time-slot, the received interference signal at PU can then be written as

$$\mathbf{i}_{1,k} = m_{1,k} \mathbf{F}_1 \mathbf{V}_{T_{1,k}} \mathbf{x}_1 + m_{2,k} \mathbf{F}_2 \mathbf{V}_{T_{2,k}} \mathbf{x}_2. \quad (4.7)$$

Similarly, in the second time-slot, the received interference signal at PU can then be written as

$$\begin{aligned} \mathbf{i}_{2,k} &= M_k \mathbf{G}_k y_{R_k} \\ &= m_{1,k} M_k \mathbf{G}_k \mathbf{H}_{1,k} \mathbf{V}_{T_{1,k}} \mathbf{x}_1 + m_{2,k} M_k \mathbf{G}_k \mathbf{H}_{2,k} \mathbf{V}_{T_{2,k}} \mathbf{x}_2 \\ &\quad + M_k \mathbf{G}_k \mathbf{n}_{R_k} + M_k \mathbf{G}_k \mathbf{i}_{R_k}. \end{aligned} \quad (4.8)$$

The interference constraints at the PU are given as

$$I_{1,k} = P_{1,k} \text{Tr}(\mathbf{F}_1^H \mathbf{F}_1) + P_{2,k} \text{Tr}(\mathbf{F}_2^H \mathbf{F}_2) \leq I_t, \quad (4.9)$$

$$I_{2,k} = P_{R_k} \text{Tr}(\mathbf{G}_k^H \mathbf{G}_k) \leq I_t, \quad (4.10)$$

where I_t is the interference threshold.

4.2.5 Outage of the secondary system

The outage in the secondary TWRN with relay R_k will occur if the interference constraints (4.9) and (4.10) are not satisfied. Let the random variable $O_{S_k} = 1$

denote outage and $O_{S_k} = 0$ denote non-outage of the secondary TWRN with relay R_k . The probability of outage can be defined as

$$P_{\text{out},k} = \Pr[O_{S_k} = 1] = \Pr[I_{1,k} > I_T \text{ or } I_{2,k} > I_T]. \quad (4.11)$$

Since $I_{1,k}$ and $I_{2,k}$ are independent of each other $P_{\text{out},k}$ can also be written as

$$P_{\text{out},k} = 1 - \Pr[I_{1,k} \leq I_T] \Pr[I_{2,k} \leq I_T]. \quad (4.12)$$

The outage of the whole system happens if all the relays are in a outage (i.e. $O_{S_k} = 1$ for all $k \in \{1, \dots, K\}$). Let I_1 be the interference on the PU when the two end nodes transmit simultaneously in the first time slot. As I_1 is common for all the relays, this outage can be written as

$$P_{\text{out}} = \Pr[I_1 > I_T \text{ or } (I_{2,1} > I_T \text{ and } \dots I_{2,K} > I_T)]. \quad (4.13)$$

Since $I_1, I_{2,1}, \dots, I_{2,K}$ are independent of each other P_{out} can also be written as

$$P_{\text{out}} = 1 - \Pr[I_1 \leq I_T] (1 - \Pr[I_{2,1} > I_T] \dots \Pr[I_{2,K} > I_T]). \quad (4.14)$$

4.2.6 Exact conditional end-to-end SINR

Theorem 1 *The post processing end-to-end SINR of the l th data sub-stream at S_i under the transmit powers of $P_{1,k}, P_{2,k}$ and P_{R_k} when $O_{S_k} = 0$ and relay R_k is used, is given as*

$$\gamma_{S_{i,k}|O_{S_k}=0}^{(l)} = \frac{M_k^2 m_{i',k}^2}{M_k^2 \sigma_{R_k}^2 + \frac{\sigma_i^2}{\eta_{i,k}} \left[\left(\tilde{\mathbf{H}}_{k,i}^H \tilde{\mathbf{H}}_{k,i} \right)^{-1} \right]_{l,l}}, \quad (4.15)$$

where $\{i, i'\} \in \{\{1, 2\}, \{2, 1\}\}$, $l \in \{1, \dots, N_{R_k}\}$, and $k \in \{1, \dots, K\}$. Also the summed noise terms are defined as $\sigma_{R_k}^2 = \sigma_{I_{R_k}}^2 + \sigma_{N_{R_k}}^2$ and $\sigma_i^2 = \sigma_{I_i}^2 + \sigma_{N_i}^2$.

Proof

By substituting (4.1), (4.3), and (4.6) into (4.4), the received signal vector at S_i can be written in an alternative form as follows:

$$\mathbf{y}_{S_{i,k}} = M_k (m_{i,k} \mathbf{x}_i + m_{i',k} \mathbf{x}_{i'} + \mathbf{n}_{R_k} + \mathbf{i}_{R_k}) + \tilde{\mathbf{n}}_i, \quad (4.16)$$

where $\{i, i'\} \in \{\{1, 2\}, \{2, 1\}\}$. Further, $\tilde{\mathbf{n}}_i$ is the filtered, colored noise and is given by $\tilde{\mathbf{n}}_i = \mathbf{V}_{R_{i,k}}(\mathbf{n}_i + \mathbf{i}_i)$. Next, by using self-interference cancellation to (4.16), the signal vector of $S_{i'}$ received at S_i can be extracted as follows:

$$\tilde{\mathbf{y}}_{S_{i,k}} = M_k (m_{i',k} \mathbf{x}_{i'} + \mathbf{n}_{R_k} + \mathbf{i}_{R_k}) + \tilde{\mathbf{n}}_i. \quad (4.17)$$

By using (4.17), the post-processing end-to-end SINR of the l th data sub-stream at S_i can be derived as (4.15). \blacksquare

Corollary 1.1 *By substituting M_k (4.5), $m_{i',k}$ (4.2) and the result $\text{Tr}(\mathbf{V}_{T_{i,k}} \mathbf{V}_{T_{i,k}}^H) = \text{Tr}([\mathbf{H}_{i,k} \mathbf{H}_{i,k}^H]^{-1})$ into (4.15), the end-to-end conditional SINR in (4.15) can be written in a more insightful form as*

$$\gamma_{S_{i,k}^{(l)} | O_{S_k}=0} = \frac{P_{R_k} P_{i',k} / \text{Tr}([\mathbf{H}_{i',k} \mathbf{H}_{i',k}^H]^{-1})}{P_{R_k} \sigma_{R_k}^2 + \left(\frac{P_{1,k}}{\text{Tr}([\mathbf{H}_{1,k} \mathbf{H}_{1,k}^H]^{-1})} + \frac{P_{2,k}}{\text{Tr}([\mathbf{H}_{2,k} \mathbf{H}_{2,k}^H]^{-1})} + \sigma_{R_k}^2 \right) \frac{\sigma_i^2}{\eta_{i,k}} \left[(\tilde{\mathbf{H}}_{k,l}^H \tilde{\mathbf{H}}_{k,l})^{-1} \right]_{l,l}}. \quad (4.18)$$

4.2.7 Sum rate analysis when R_k is selected

This subsection obtains sum rate expressions when the relay R_k is selected assuming that the interference constraints at PU are satisfied. To correctly decode the data sub-streams by the receivers, each node needs to transmit the data with a common rate in MIMO TWRNs. Thus, the sum rate obtained by selecting the k th relay can be defined as follows:

$$\mathcal{R}_k = \begin{cases} 2 \min(\mathcal{R}_{S_{1,k}}, \mathcal{R}_{S_{2,k}}), & \text{if } O_{S_k} = 0 \\ 0, & \text{if } O_{S_k} = 1 \end{cases} \quad (4.19)$$

where \mathcal{R}_{S_i} is the sum of data sub-streams rates at $S_{k,i}$ for $i \in \{1, 2\}$, and can be written as

$$\mathcal{R}_{S_{i,k}} = \begin{cases} \frac{1}{2} \sum_{l=1}^{N_R} \log \left(1 + \gamma_{S_{i,k}^{(l)} | O_{S_k}=0} \right), & \text{if } O_{S_k} = 0 \\ 0, & \text{if } O_{S_k} = 1 \end{cases} \quad (4.20)$$

The factor of two appears in (4.19) due to the presence of two user nodes in the TWRN of interest. Further, the pre-log factor of one-half in (4.20) is due to the use of two time-slots.

4.3 Asymptotic analysis

The asymptotic analysis when the number of antennas at the end nodes goes to infinity (i.e. $N_1, N_2 \rightarrow \infty$) is provided in this section. Specifically, while N_1 and N_2 grow unbounded, the number of antennas at relay R_k and PU remains constant. For simplicity, it is assumed that as $N_1, N_2 \rightarrow \infty$, the ratio between them remains constant:

$$\alpha = \frac{N_2}{N_1}. \quad (4.21)$$

Next the asymptotic sum rate and the interference constraints for the two different power scaling scenarios are obtained.

4.3.1 Requirement of power scaling

Scaling the transmit power inversely with the number of antennas has been widely used in the literature to limit the total transmit power of a node. In [123] three power scaling scenarios has been analyzed for mMIMO two-way relay networks. These includes power scaling at the end nodes, power scaling at the relay, and power scaling at both the end and relay nodes. However, not all these power scaling methods can be used in an underlay setup. To clarify this situation, the interference $I_{1,K}$ during the first time slot is rewritten as

$$I_{1,k} = P_{1,k} \text{Tr}(\mathbf{F}_1^H \mathbf{F}_1) + P_{2,k} \text{Tr}(\mathbf{F}_2^H \mathbf{F}_2). \quad (4.22)$$

But if $P_{1,k}$ and $P_{2,k}$ are constants, as N_1 and N_2 become extremely large the value $\text{Tr}(\mathbf{F}_1^H \mathbf{F}_1)$ and $\text{Tr}(\mathbf{F}_2^H \mathbf{F}_2)$ reaches infinity (the asymptotic limit results in Appendix B.1). Thus without scaling down $P_{1,k}$ and $P_{2,k}$, the interference constraint at the PU can not be satisfied for a large number of transmit antennas. Thus, there are two ways to scale down $P_{1,k}$ and $P_{2,k}$, and they are analyzed in the next two subsections.

4.3.2 Power scaling at the end nodes and the relay (PS-1)

Theorem 2 *When, the transmit powers of user nodes S_1 and S_2 and relay nodes R_k for $k \in \{1, \dots, K\}$ are scaled inversely with the number of antennas at the end nodes(i.e. PS-1) and condition (4.22) is satisfied, the asymptotic SINR is given as*

$$\gamma_{S_{1,k}}^{\infty} = \frac{E_{R_k} E_{2,k} \eta_{1,k} \eta_{2,k}}{\sigma_1^2 (E_{1,k} \eta_{1,k} + E_{2,k} \eta_{2,k}) + \sigma_{R_k}^2 N_{R_k} (E_{R_k} \eta_{1,k} + \sigma_1^2)}, \quad (4.23)$$

and

$$\gamma_{S_{2,k}}^{\infty} = \frac{\alpha E_{R_k} E_{1,k} \eta_{1,k} \eta_{2,k}}{\sigma_2^2 (E_{1,k} \eta_{1,k} + E_{2,k} \eta_{2,k}) + \sigma_{R_k}^2 N_{R_k} (\alpha E_{R_k} \eta_{2,k} + \sigma_2^2)}. \quad (4.24)$$

Proof

The two end nodes and the relays can scale down their transmit power as follows:

$$P_{i,k} = \frac{E_{i,k}}{N_i} \quad \text{for } i \in \{1, 2\} \quad \text{and} \quad P_{R_k} = \frac{E_{R_k}}{N_1}, \quad (4.25)$$

where $E_{1,k}$, $E_{2,k}$ and E_{R_k} are fixed constants. To prove the asymptotic SINRs in (23) and (24), the exact expression of SINR (4.18) are utilized, which contains several matrix trace terms and the diagonal term of an inverse matrix. All these terms have asymptotic limits (see proof in Appendix B.1). ■

Interestingly, the asymptotic SINRs (4.23) and (4.24) are independent of the small-scale fading and only depend on the pathloss. Note that they are also independent of the data sub-stream index, l , and hence, $\gamma_{S_{i,k}}^{\infty} = \gamma_{S_{i,k}}^{\infty}$ for $l \in \{1, \dots, N_{R_k}\}$.

Next it is shown that the interference constraints are satisfied during the two slots given the power scaling scenario (25). First, consider the second time slot where only the selected relay is transmitting. By substituting $P_{i,k}$ and P_{R_k} , (4.10) is rewritten as

$$I_{2,k} = E_{R_k} \frac{\text{Tr}(\mathbf{G}_k^H \mathbf{G}_k)}{N_1} \leq I_t, \quad (4.26)$$

and as $\frac{\text{Tr}(\mathbf{G}_k^H \mathbf{G}_k)}{N_1} \rightarrow 0$ as $N_1 \rightarrow \infty$, this condition is asymptotically satisfied for any value of E_{R_k} . Second, consider the first time slot where the two end nodes are transmitting. By using the limits on (4.22), the interference constraint is obtained as (see Appendix B.1)

$$\lim_{N_i \rightarrow \infty} I_{1,k} = E_{1,k} \hat{\eta}_1 N + E_{2,k} \hat{\eta}_2 N \leq I_t. \quad (4.27)$$

The constraint (4.27) does not contain any random quantities. Thus, it can be strictly satisfied by carefully selecting $E_{1,k}$ and $E_{2,k}$ at the end nodes of the secondary network. Therefore, the secondary system can operate without an outage (i.e. $P_{out} = 0$).

By using asymptotic SINR (4.23) and (4.24), the sum rate between U_i and R_k is obtained as

$$\mathcal{R}_{S_{i,k}}^\infty = \frac{1}{2} N_{R_k} \log \left(1 + \gamma_{S_{i,k}}^\infty \right). \quad (4.28)$$

Further, the total sum rate when R_k is selected is given as

$$\mathcal{R}_k^\infty = N_{R_k} \log \left(1 + \min \left(\gamma_{S_{1,k}}^\infty, \gamma_{S_{2,k}}^\infty \right) \right). \quad (4.29)$$

Remark I: The results obtained in (4.23) and (4.24), only depend on the pathloss component and the average noise values. Thus, the total sum rate will be a fixed value for a given relay. Thus, relay selection can be done off-line and no instantaneous channel state is necessary for the relay selection process. Furthermore, outage $P_{out,k}$ can be set to zero by scaling the transmit powers of S_1 and S_2 .

4.3.3 Power scaling at the end nodes - (PS-2)

Theorem 3 *The asymptotic SINR when the transmit power at the user nodes S_1 and S_2 are scaled inversely proportional to the number of antennas at the user nodes (i.e PS-2) is given as*

$$\gamma_{S_{1,k}}^\infty = \frac{E_{2,k} \eta_{2,k} N_{R_k}}{\sigma_{R_k}^2}, \quad (4.30)$$

and

$$\gamma_{S_{2,k}}^\infty = \frac{E_{1,k} \eta_{1,k} N_{R_k}}{\sigma_{R_k}^2}. \quad (4.31)$$

Proof

The transmit power scaling can be given as

$$P_{i,k} = \frac{E_{i,k}}{N_i} \quad \text{for } i \in \{1, 2\} \quad \text{and} \quad P_{R_k} = E_{R_k}, \quad (4.32)$$

where $E_{1,k}$, $E_{2,k}$ and E_{R_k} are fixed. The equations for SINRs are obtained by substituting $P_{i,k}$ and P_{R_k} values and by using the asymptotic limit results on (4.18). Proofs are omitted due to the similarity to the proof of SINRs in Appendix B.1. ■

Remark II: Similar to (4.23) and (4.24), the SINR values in (4.30) and (4.31) only depend on the pathloss component and the average noise values. Furthermore, they are independent of the interference and noise levels at the end nodes (σ_i^2).

Proposition 3.1 *The average sum rate of the secondary system under PS-2 is given as follows:*

$$\begin{aligned}\bar{\mathcal{R}}_k^\infty &= (1 - P_{\text{out},k})N_{R_k} \log \left(1 + \min \left(\gamma_{S_{1,k}}^\infty, \gamma_{S_{2,k}}^\infty \right) \right) \\ &= \left(1 - \frac{\Gamma \left(NN_{R_k}, \frac{I_t}{\eta_k E_{R_k}} \right)}{\Gamma(NN_{R_k})} \right) N_{R_k} \log \left(1 + \min \left(\frac{E_{2,k} \eta_{2,k} N_{R_k}}{\sigma_{R_k}^2}, \frac{E_{1,k} \eta_{1,k} N_{R_k}}{\sigma_{R_k}^2} \right) \right)\end{aligned}\quad (4.33)$$

Proof

As mentioned before, (4.27) provides the interference constraint during the first time-slot. This constraint can be satisfied by selecting $E_{1,k}$ and $E_{2,k}$ appropriately. However, the PU interference from the selected relay during the second time slot must also satisfy the interference constraint:

$$I_{2,k} = E_{R_k} \text{Tr} \left(\mathbf{G}_k^H \mathbf{G}_k \right) = \eta_k E_{R_k} \text{Tr} \left(\tilde{\mathbf{G}}_k^H \tilde{\mathbf{G}}_k \right) \leq I_t. \quad (4.34)$$

Unlike the PS-1 case, this is a random event, whose likelihood depends on channel quality between the selected relay and the PU. Thus the outage of the secondary system (i.e. $P_{\text{out},k}$) is unpredictable. The outage probability may be expressed as

$$P_{\text{out},k} = \Pr[I_{2,k} \geq I_t] = \Pr \left[\text{Tr} \left(\tilde{\mathbf{G}}_k^H \tilde{\mathbf{G}}_k \right) \geq \frac{I_t}{\eta_k E_{R_k}} \right]. \quad (4.35)$$

It is well known that $\text{Tr} \left(\tilde{\mathbf{G}}_k^H \tilde{\mathbf{G}}_k \right)$ is Gamma-distributed with shape parameter of $N \times N_{R_k}$ [124]. Thus $P_{\text{out},k}$ can be expressed as

$$P_{\text{out},k} = \frac{1}{\Gamma(NN_{R_k})} \Gamma \left(NN_{R_k}, \frac{I_t}{\eta_k E_{R_k}} \right). \quad (4.36)$$

By using the obtained SINR values, the sum rate can be written as

$$\mathcal{R}_k^\infty = \begin{cases} N_{R_k} \log \left(1 + \min \left(\gamma_{S_{1,k}}^\infty, \gamma_{S_{2,k}}^\infty \right) \right), & \text{if } O_{S_k} = 0 \\ 0, & \text{if } O_{S_k} = 1, \end{cases} \quad (4.37)$$

and (4.33) can be obtained. ■

4.4 Optimal power allocation

Optimal power allocation schemes are designed to maximize the sum rates given in (4.29) and (4.37) under the interference constraint in (4.27) for PS-1 and PS-2. The optimal power allocations at S_1 and S_2 are presented as $E_{1,k}^*$ and $E_{2,k}^*$. Furthermore, it is assumed that the transmit power at the selected relay E_{R_k} is a fixed value for the following two cases.

4.4.1 Power scaling at the end nodes and the relay (PS-1)

Maximizing R_k^∞ in (4.29) corresponds to maximizing the minimum of $\gamma_{S_{1,k}}^\infty$ and $\gamma_{S_{2,k}}^\infty$ by selecting $E_{1,k}$ and $E_{2,k}$ which satisfies (4.27). It can be seen that both $\gamma_{S_{1,k}}^\infty$ and $\gamma_{S_{2,k}}^\infty$ are increasing with respect to $E_{1,k}$ and $E_{2,k}$. Thus, the maximum of R_k^∞ occurs when (4.27) becomes an equality. Thus

$$E_{1,k}^* = \frac{1}{\hat{\eta}_1} \left(\frac{1}{(1+\delta)N} I_t - E_{2,k}^* \hat{\eta}_2 \right), \quad (4.38)$$

where $0 \leq \delta \ll 1$ is used to make sure that the interference on PU is significantly less than the threshold interference level. Furthermore, it can be seen that the maximum value for the minimum between two asymptotic SINR values occur when $\gamma_{S_{1,k}}^\infty = \gamma_{S_{2,k}}^\infty$. Thus, the following equation is obtained:

$$\begin{aligned} \sigma_2^2 \quad & \eta_{2,k} N_{R_k} E_{2,k}^{*2} - \alpha \sigma_1^2 \eta_{1,k} N_{R_k} E_{1,k}^{*2} + (\sigma_2^2 \eta_{1,k} - \alpha \sigma_1^2 \eta_{2,k}) N_{R_k} E_{1,k}^* E_{2,k}^* \\ & + (\alpha E_{R_k} \eta_{2,k} + \sigma_2^2) \sigma_{R_k}^2 E_{2,k}^* - \alpha (E_{R_k} \eta_{1,k} + \sigma_1^2) \sigma_{R_k}^2 E_{1,k}^* = 0. \end{aligned} \quad (4.39)$$

Then solving (4.38) and (4.39) provides the optimal values $E_{1,k}^*$ and $E_{2,k}^*$. By using these values, $\gamma_{S_{1,k}}^{\infty,*}$, $\gamma_{S_{2,k}}^{\infty,*}$, and $\mathcal{R}_k^{\infty,*}$ is obtained which are the asymptotic optimal values of SINRs and sum rate, respectively.

As an special case, if $\alpha = 1$ and $\eta_{1,k} = \eta_{2,k}$, then it can be concluded that $E_{1,k}^* = E_{2,k}^*$. And the optimum power allocation values are given by using (4.38), as

$$E_{1,k}^* = E_{2,k}^* = \frac{1}{(1+\delta)(\hat{\eta}_1 + \hat{\eta}_2)N} I_t. \quad (4.40)$$

4.4.2 Power scaling at the end nodes (PS-2)

This section provides the optimal power allocation scenario for the power scaling at the end nodes case. Unlike in the previous scenario, the sum rate also depends on the outage probability $P_{out,k}$ which depends on the transmit power at the relay node P_{R_k} . However, the asymptotic SINR values in (4.30) and (4.31) are independent of E_{R_k} . Thus, similar to the previous case this corresponds to the case where the equality condition in (4.27) is satisfied. Thus,

$$\begin{aligned} \gamma_{S_{1,k}}^\infty &= \gamma_{S_{2,k}}^\infty \\ \frac{E_{2,k} \eta_{2,k} N_{R_k}}{\sigma_{R_k}^2} &= \frac{E_{1,k} \eta_{1,k} N_{R_k}}{\sigma_{R_k}^2}. \end{aligned} \quad (4.41)$$

By using value for $E_{1,k}^*$ from (4.38) in (4.41), the following is obtained:

$$E_{1,k}^* = \frac{\eta_{2,k}}{(1 + \delta) (\hat{\eta}_1 \eta_{2,k} + \hat{\eta}_2 \eta_{1,k}) N} I_t, \quad (4.42)$$

$$E_{2,k}^* = \frac{\eta_{1,k}}{(1 + \delta) (\hat{\eta}_1 \eta_{2,k} + \hat{\eta}_2 \eta_{1,k}) N} I_t. \quad (4.43)$$

Thus, the optimal average sum rate is written as

$$\bar{\mathcal{R}}_k^{\infty,*} = \left(1 - \frac{\Gamma \left(NN_{R_k}, \frac{I_t}{\eta_k E_{R_k}} \right)}{\Gamma(NN_{R_k})} \right) N_{R_k} \log \left(1 + \frac{\eta_{1,k} \eta_{2,k} N_{R_k}}{(1 + \delta) (\hat{\eta}_1 \eta_{2,k} + \hat{\eta}_2 \eta_{1,k}) N \sigma_{R_k}^2} I_t \right). \quad (4.44)$$

4.5 Relay selection with optimum values

This section presents the relay selection criteria for the two power scaling strategies, PS-1 and PS-2. For both these cases, the objective of relay selection is to maximize the achievable asymptotic sum rate of the secondary system. The relay selection criterion may be stated as:

$$K^* = \operatorname{argmax}_{k \in \{1, \dots, K\}} [\mathcal{R}_k^{\infty,*}], \quad (4.45)$$

where K^* is the index of the selected relay. Next two subsections analyse the relay selection criteria under PS-1 and PS-2.

4.5.1 Power scaling at the end nodes and the relay (PS-1)

The relay selection can further be simplified to the following:

$$K^* = \operatorname{argmax}_{k \in \{1, \dots, K\}} [\mathcal{R}_k^{\infty,*}] = \operatorname{argmax}_{k \in \{1, \dots, K\}} \left[N_{R_k} \log \left(1 + \gamma_{S_{1,k}}^{\infty} \right) \right], \quad (4.46)$$

Thus, relay selection is deterministic as all the $R_k^{\infty,*}$ are fixed values. Further, (4.46) shows that the choice of optimal relay highly depends on the maximum value of N_{R_k} . Since the log of $(1 + \gamma_{S_{1,k}}^{\infty})$ is taken, the impact of N_{R_k} is higher than that of the logarithm. Thus, for most cases the relay selection simply boils down to the selection of the relay with the maximum number of antennas.

4.5.2 Power scaling at the end nodes (PS-2)

Unlike in the previous case, here the sum rate of R_k^∞ and outage are inter-related. Since outage is random, the relay selection will no longer be a deterministic process. Yet, motivated by the simplicity of relay selection for deterministic sum rates (46), two relay selection methods are proposed where, one of them does not require the instantaneous channel knowledge. The proposed relay selection methods are to

1. Maximize the instantaneous asymptotic sum rate (RSS-1),
2. Maximize the average sum rate (RSS-2).

Next the advantages and disadvantages of these two relay selection schemes are discussed, and their performances are compared in Section 4.7.

Maximize the instantaneous sum rate

The relay selection criterion may be stated as

$$K^* = \operatorname{argmax}_{k \in \{1, \dots, K\}} [\mathcal{R}_k^{\infty, *}] . \quad (4.47)$$

However, when some of the relays are in outage, the optimal selected relay will be among the relays that are not in outage. Thus, in this case relay selection clearly provides higher sum rates and lower outage probabilities. However, the knowledge of the channel states is necessary for the relay selection process, whose complexity will thus increase. By satisfying $\Pr[I_1 \leq I_T] = 1$ with power allocation, and by using the outages related to $I_{2,k}$ s from (4.36), (4.14) is rewritten as

$$P_{\text{out}} = \prod_{k=1}^K \frac{\Gamma\left(NN_{R_k}, \frac{I_t}{\eta_k E_{R_k}}\right)}{\Gamma(NN_{R_k})} . \quad (4.48)$$

As a special case, if all the relay nodes are identical (i.e. $N_{R_k} = N_R$ and $E_{R_k} = E_R$ for all $k \in (1, \dots, K)$), the average sum rate with relay selection is given as ³

$$\bar{\mathcal{R}}^{\infty, *} = \left(1 - \frac{\prod_{k=1}^K \Gamma\left(NN_R, \frac{I_t}{\eta_k E_R}\right)}{\Gamma(NN_R)^K}\right) N_R \log \left(1 + \frac{\eta_{1,k} \eta_{2,k} N_R}{(1 + \delta) (\hat{\eta}_1 \eta_{2,k} + \hat{\eta}_2 \eta_{1,k}) N \sigma_R^2} I_t\right) . \quad (4.49)$$

³Even in this case, the parameter values $\eta_{1,k}$ and $\eta_{2,k}$ will be different. However, for simplification it is assumed that $\eta_{1,k}$ and $\eta_{2,k}$ are approximately equal for all k .

Eq. (4.48) shows that relay selection significantly decreases the outage probability and hence increases the average sum rate compared to a single relay system. However, instantaneous channel knowledge for the relay selection will cost bandwidth, energy and complexity. The reason is that channel estimation requires the transmission of pilot sequences and channel state feed back to the transmit node and to the relay selector.

Maximize the average sum rate

For this case, the relay selection equation can be given as

$$\begin{aligned} K^* &= \operatorname{argmax}_{k \in \{1, \dots, K\}} [\bar{\mathcal{R}}_k^{\infty,*}] \\ &= \operatorname{argmax}_k \left[\left(1 - \frac{\Gamma\left(NN_{R_k}, \frac{I_t}{\eta_k E_{R_k}}\right)}{\Gamma(NN_{R_k})} \right) N_{R_k} \log \left(1 + \frac{\eta_{1,k} \eta_{2,k} N_{R_k}}{(1+\delta)(\hat{\eta}_1 \eta_{2,k} + \hat{\eta}_2 \eta_{1,k}) N \sigma_{R_k}^2} I_t \right) \right] \end{aligned} \quad (4.50)$$

Since the average sum rate, which is a deterministic value, is used, the relay selection becomes deterministic. Thus the relay selection complexity for this scenario is significantly less than the previous case as the instantaneous channel conditions of all the relays are not required for the relay selection. However, the channel conditions of the required relay is still required for the beamforming and interference control. Similar to the power scaling at the end nodes and the relay case, the number of antennas at the relay N_{R_k} has the highest impact on the relay selection.

As a special case, if the case with identical relays is considered, the average sum rate is given by

$$\bar{\mathcal{R}}_k^{\infty,*} = \left(1 - \frac{\Gamma\left(NN_R, \frac{I_t}{\eta_k E_R}\right)}{\Gamma(NN_R)} \right) N_R \log \left(1 + \frac{\eta_{1,k} \eta_{2,k} N_R}{(1+\delta)(\hat{\eta}_1 \eta_{2,k} + \hat{\eta}_2 \eta_{1,K}) N \sigma_R^2} I_t \right). \quad (4.51)$$

It is evident from (4.51) and (4.49) that the increased sum rate in the previous case is due to the increased relay selection complexity.

Remark III: As evident from the above two cases, the number of antennas at a given relay N_{R_k} has the highest impact on the achievable sum rate of a system. Thus, if all other parameters are approximately equal, then selecting the relay with the highest number of antennas will provide the highest sum rate for the secondary system.

Remark IV: It can be seen that the sum rate $R_k^{\infty,*}$ is a deterministic variable. Thus, the relay selection also becomes a deterministic task. Thus, relay selection can be done during the system design stage. That is, the end and relay nodes need not implement real-time relay selections.

4.6 Energy efficiency analysis

This section analyses the energy efficiency under the power allocation conditions in Section 4.4.

Some analyses are based on the assumption that the power consumption at the nodes is only due to the wireless data transfers between nodes. In this case, energy efficiency of PS-1 approaches infinity and the energy efficiency of PS-2 approaches an asymptotic limit. However, the processing done at the nodes consumes more than 50% of the total in mMIMO [80]. Thus, both power consumption at circuit components and that for ZF computations done at the end nodes must be considered. Therefore, a more accurate mMIMO energy efficiency analysis is proposed in [125]. According to this, the power consumption during time-slot 1, $P_{tot,ts1}$ can be given as

$$P_{tot,ts1} = P_{1,k} + P_{2,k} + P_{ts1,TC} + P_{ts1,C} + P_{1,k,LP} + P_{2,k,LP}, \quad (4.52)$$

where $P_{i,k}$ is the transmit power at the i th end node, $P_{ts1,TC}$ is the power consumed in transceiver chains in time-slot 1, $P_{ts1,C}$ is the power consumed for coding in time-slot 1, and $P_{i,k,LP}$ is the power consumed in node i for linear processing (in this case ZF). The values for these power consumption components are given as follows:

$$P_{ts1,TC} = N_1 P_{EN} + N_2 P_{EN} + K P_{Rel} + P_{SYN}, \quad (4.53)$$

$$P_{ts1,C} = 2\mathcal{R}_k P_{COD}, \quad (4.54)$$

$$P_{i,k,LP} = \frac{2BN_i N_{R_k}}{L_{BS}} + \frac{B}{U} \left(\frac{N_{R_k}^3}{3L_{BS}} + \frac{3N_i N_{R_k}^2 + N_i N_{R_k}}{L_{BS}} \right), \quad (4.55)$$

where P_{EN} and P_{Rel} are the powers required to run the circuit components at the end nodes and relays, P_{SYN} is the power of the local oscillator, P_{COD} is the coding power consumption at the end nodes, B is the bandwidth, L_{BS} is the computational efficiency (given in flops/W) of the end nodes, and U is the coherence block. Similarly, the power consumption during time-slot 2 is given as

$$P_{tot,ts2} = P_{R_k} + P_{ts2,TC} + P_{ts2,D} + P_{k,1,LP} + P_{k,2,LP}, \quad (4.56)$$

where $P_{ts2,TC} = P_{ts1,TC}$, $P_{k,i,LP} = P_{i,k,LP}$, and

$$P_{ts2,D} = 2\mathcal{R}_k P_{DEC}, \quad (4.57)$$

where P_{DEC} is the decoding power consumption at the end nodes. Based on these values the average energy efficiency of the secondary system is defined as

$$\rho_k = \frac{\mathcal{R}_k}{P_{tot,ts1} + P_{tot,ts2}}. \quad (4.58)$$

Based on (4.58), the energy efficiency under PS-1 can be written as

$$\rho_{k,PS1} = \frac{\mathcal{R}_k}{\frac{E_{1,k}}{N_1} + \frac{E_{2,k}}{N_1} + \frac{E_{R_k}}{N_1} + \mathcal{P}_{res}}. \quad (4.59)$$

Here, $\mathcal{P}_{res} = 2((N_1 + N_2)P_{EN} + KP_{Rel} + P_{SYN}) + 2\mathcal{R}_k(P_{COD} + P_{DEC}) + 2P_{1,k,LP} + 2P_{2,k,LP}$. Furthermore, $P_{1,k,LP}$ and $P_{2,k,LP}$ is given by (4.55). Note that the consumed power does not go to zero as the number of antennas is increased. Instead, the consumed power increase with N_1 and N_2 due to the second term in the above equation. Similarly, the energy efficiency under PS-2 is written as

$$\rho_{k,PS2} = \frac{\mathcal{R}_k}{\frac{E_{1,k}}{N_1} + \frac{E_{2,k}}{N_1} + E_{R_k} + \mathcal{P}_{res}}. \quad (4.60)$$

4.7 Numerical results

This section presents the numerical and simulation results of the proposed selection strategies. Specifically, the average and asymptotic sum rates are computed and simulated. The interference threshold at PU is $I_T = 10$ dB, the pathloss components between PU and S_1 and S_2 as $\hat{\eta}_1 = \hat{\eta}_2 = 1/16$, and the noise powers $\sigma_{N_i}^2 = \sigma_{R_k}^2 = 1$. Also $\alpha = 1$ and transmit power at the relay R_k is taken as $E_k = 25$ dBm. The primary user has 6 antennas ($N = 6$). Figs. 4.2 to 4.7 represent the results for PS-1 while the Figs. 4.8 to 4.12 show the simulation results for PS-2.

In Fig. 4.2, four cases with different numbers of relays for end-node power scaling are analysed. Case 1 consists of a single relay node ($K = 1$) with a single antenna. In Case 2, $K = 4$ and the numbers of relay antennas (N_{R_k}) are 1, 1, 4, 4, respectively. In Case 3, $K = 8$ and the numbers of relay antennas are 1, 1, 4, 4, 8, 8, 8, 8, respectively. In Case 4, the sum rate is plotted, when the relay with the highest number of antennas is selected (relay with 8 antennas). The optimal power allocation scheme with $\delta = 0.1$ is used. It can be seen that multiple relays bring about significant sum rate gains. For example a single relay with a single antenna can only achieve a sum rate of 2.5bps/Hz while the four-relay system, which has 1, 1, 4, and 4 antennas each can obtain a sum rate of 5.5bps/Hz. Also, with eight relays, the achievable

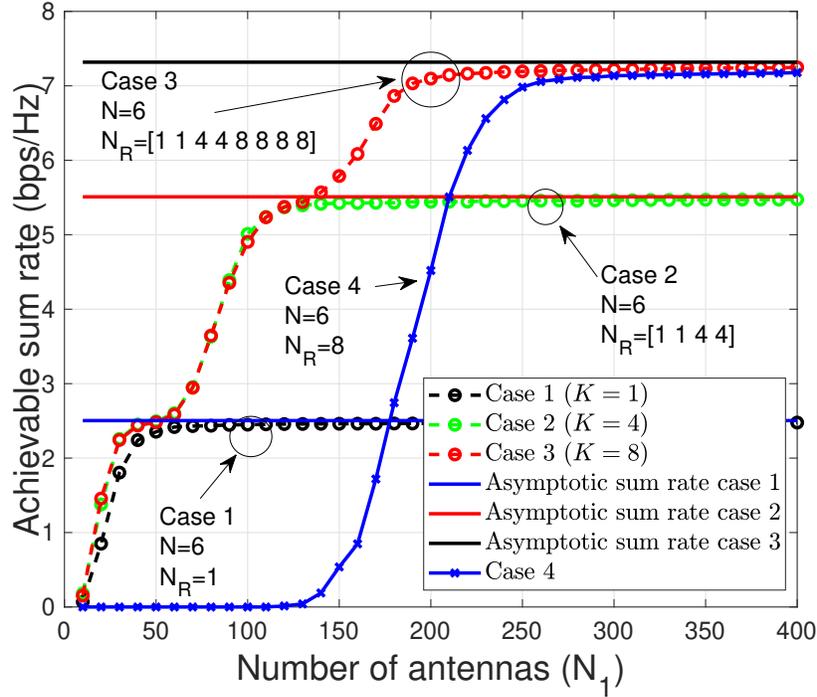


Figure 4.2: Sum rate as a function of the number of S_1 antennas.

sum rate is 7.2bps/Hz, which is a significant increase compared to previous two cases. The asymptotic analysis perfectly matches the simulated sum rates when the number of antennas are increased. Selecting the relay with the highest number of antennas (Case 4) provides the same asymptotic sum rate as Case 3, which uses relay selection. However, without relay selection, more antennas are required to achieve asymptotic performance.

Fig. 4.3 plots the energy efficiency given in (4.59) for system configuration cases that were analyzed in the previous figure. The parameter values for P_{EN} , P_{Rel} , P_{SYN} , P_{COD} , P_{DEC} , B , U , and L_{BS} are adopted from [125]. For Case 1, the energy efficiency increases as the number of antennas are increased up-to around 40. However, after that, energy efficiency reduces as the number of antennas are increased and asymptotically reach zero. For Case 2 and Case 3, there are multiple local maxima for the energy efficiency. The reason behind this is that different relays have different number of antennas (i.e. for Case 3, two relays have single antennas and two relays have 4 antennas and the other relays have 8 antennas). Regardless of the number of antennas available at the relays, more relays always improve energy efficiency.

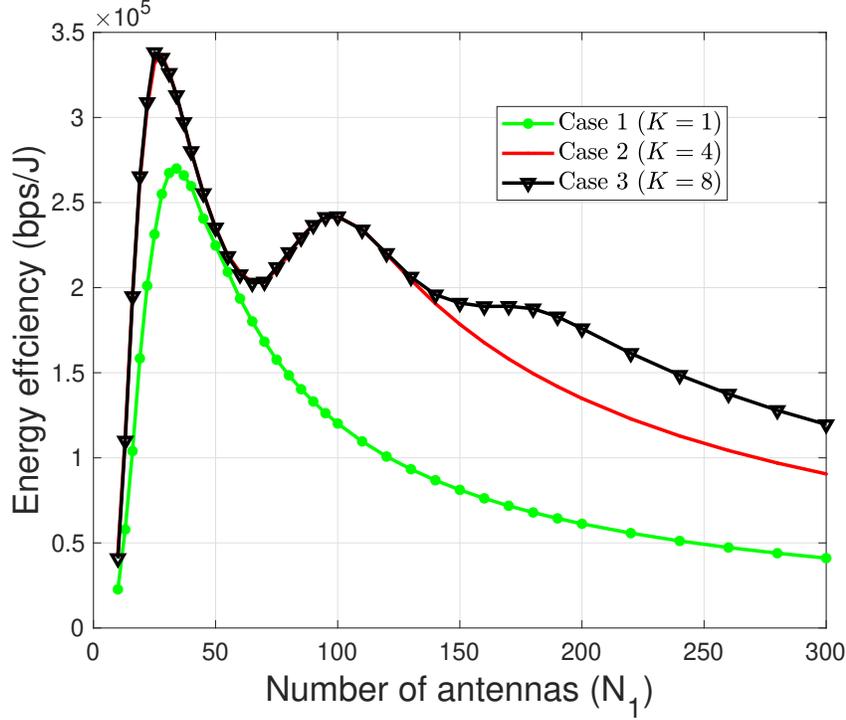


Figure 4.3: Energy efficiency comparisons as function of the number of S_1 antennas for PS-2.

To investigate the effect of identical antenna configurations, the energy efficiency is plotted when all the relays have 8 antennas for PS-1 in Fig. 4.4. It clearly shows that more relays improve energy efficiency at any number of antennas at S_1 and also, the peak energy efficiency value can be obtained for fewer antennas at S_1 . For instance, the peak energy efficiency of 1.9×10^5 bps/J can be obtained with around 170 antennas at the end nodes for Case 3 while the peak energy efficiency of 1.7×10^5 bps/J can be obtained with around 190 antennas for Case 2. The energy efficiency increase with the number of relays (especially in the low antenna regime) is due to the fact that more relays increase the sum rate of the system.

Fig. 4.5 analyzes the sum rate with different number of relays, but when those relays have the same number of antennas ($N_{R_k} = 8$ for all the relays). Cases one, two, and three corresponds to 1, 4, and 8 relays respectively. Optimal power allocation with $\delta = 0.1$ is used. Unlike in Fig. 4.2, it can be seen that the achievable asymptotic sum rates for different cases are the same. The reason is that the achievable sum rate depends on the relay with the highest number of antennas. However, with multiple relays, the asymptotic performance can be obtained by using a smaller number of

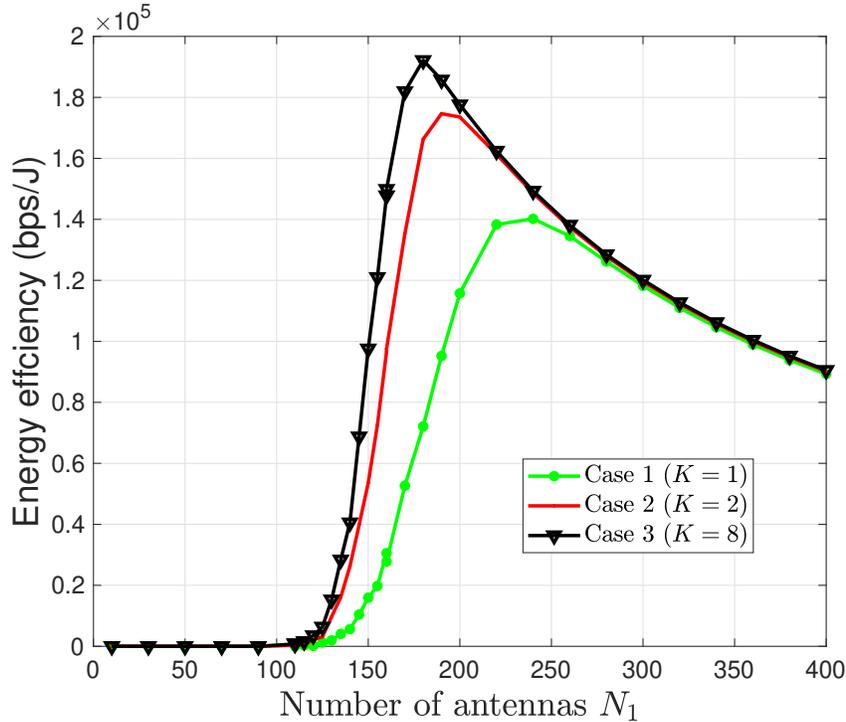


Figure 4.4: Energy efficiency as a function of the number of S_1 antennas. All the relays are equipped with 8 antennas each.

antennas at the end nodes. Also it can be seen that until the number of antennas at the end nodes surpasses a threshold (in this case around 120), the achievable sum rate is zero. This is due to the interference constraint at the PU. Although (4.22) is satisfied asymptotically for any E_{R_k} , when N_1 is low this constraint may not be satisfied and the secondary system will remain in outage state. Thus, until the number of antennas at the end nodes increases to a certain limit, the secondary network cannot start the transmission.

Fig. 4.6 plots the outage when the PU interference exceeds the threshold. (1) interference during the first time-slot (when both the end nodes are transmitting) causes the outage, (2) interference during the second time-slot causes the outage when having 8 antennas, and (3) interference during the second time-slot causes the outage when the relay with the highest number of antennas is selected is plotted in this figure. As seen from Fig. 4.6, the first-time slot outage rapidly reduces with the number of antennas. For instance, with 200 antennas the outage is less than 10^{-12} . This shows that, with a large antenna array at the end nodes and with power scaling, the interference on the PU approaches zero and thus, the end nodes do not

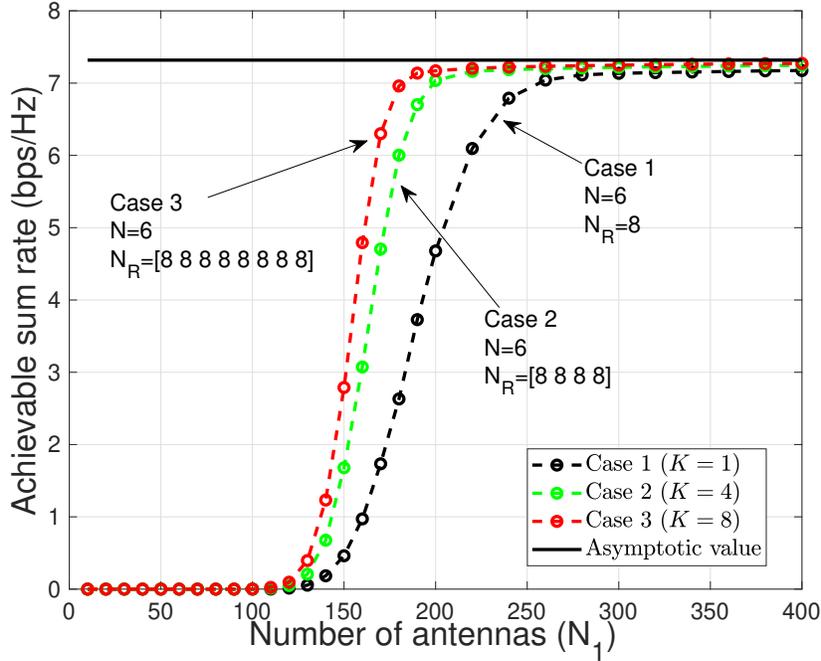


Figure 4.5: Sum rate comparisons as a function of the number of S_1 antennas. All relays are equipped with 8 antennas.

need to know each other's channel matrices to decide whether to transmit or not. Also, the interference during the second time-slot approaches zero but with a lower convergence rate. As an example with 250 antennas, relay selection provides less than 10^{-12} outage while the outage with a single relay is around 10^{-5} .

Fig. 4.7 plots the achievable sum rates for three cases for different power allocations at the end nodes. In Case 1, the derived optimal power allocation scheme in Section 4.4.1 is used. In Case 2, $\frac{E_1}{E_2} = 3/2$ and in Case 3, $\frac{E_1}{E_2} = 7/3$. It can be seen from the plot that the sum rate approaches the asymptotic sum rates obtained through equation (4.29) when the number of antennas at the end nodes increases. Furthermore, the optimal power allocation scheme in Section 4.4.1 obtains the highest achievable sum rate while the sum rates of other power allocations are significantly less than that for the optimal power allocation.

Figs. 4.8, 4.9, and 4.10 plot the outage, sum rate, and the energy efficiency for end-node power scaling (PS-2) where different number of relays are available for each case. Specifically, Case 1 has a single relay ($L = 1$), whereas Case 2 and Case 3 involve 2 and 8 relays, respectively. In all the three cases, each relay has 8 antennas. Relay selection is used to maximize the instantaneous sum rate, and the simulation

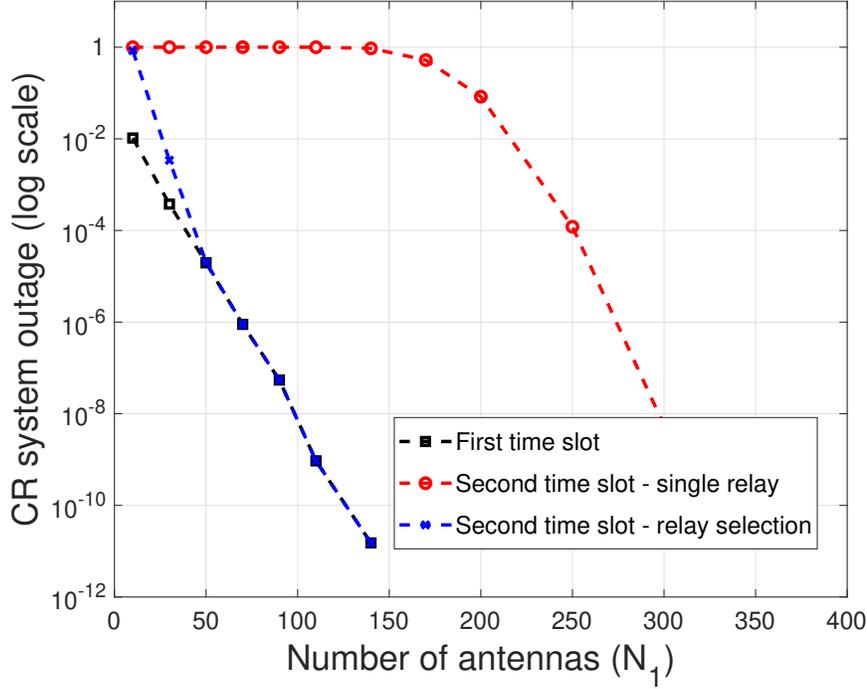


Figure 4.6: CR TWRN outage during different time slots with and without relay selection.

results are compared with the analytical asymptotic results obtained in (4.48) and (4.49). These two figures show that the analytical results match with the simulation results. Also, more relays will reduce the outage and increase the sum rate. For instance, just one relay ($L = 1$) will experience an outage of 0.96 while the outage of a system with 8 relays is 0.74. With further calculations, it can be shown that the outage will drop to 0.47, if 20 relays are deployed. Furthermore, system with 8 relays can provide sum rate of 2.1bps/Hz while a system with a single relay can only provide a sum rate of 0.3 bps/Hz. This is significantly different than that for the previous case with all-node power scaling. In that case, the achievable asymptotic sum rate is not increased when the number of relays are increased while all of them had the same number of antennas. However, when power scaling is only used at the end nodes, more relays will significantly improve the sum rate.

The energy efficiency (4.60) is plotted in Fig. 4.10. Similar to the case in PS-1, the energy efficiency increases while the number of antennas is increased and then reduces after a certain threshold. For all the cases, this threshold is around 25 antennas. This is in contrast to the energy efficiency based only on the transmit

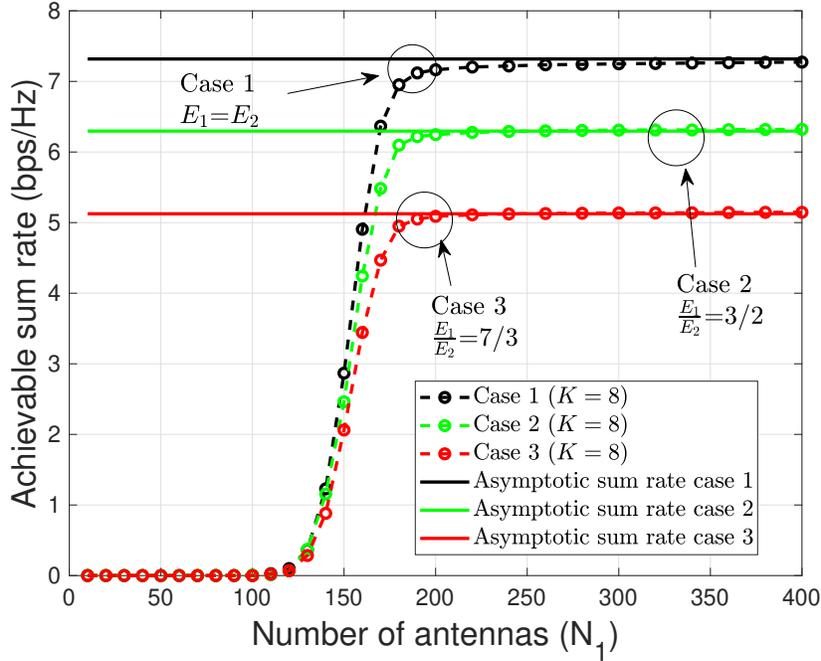


Figure 4.7: The achievable sum rate against the number of antennas at the end nodes for different power allocation cases.

power where the energy efficiency approaches infinity for extremely large number of antennas. Furthermore, more relays increases the energy efficiency and thus, relay selection helps to make green communications a reality.

Fig. 4.11 plots the sum rate against the number of antennas at the end nodes S_1 when power scaling is only used at the end nodes. In Case 1, a relay with 8 antennas is used and in Case 2 two relays with 8 antennas each are used. However in Case 3, 8 relays where 2 relays contain 8 antennas and 6 relays with 4 antennas each are used. It can be seen that multiple relays increase the sum rate. When the number of antennas at different relays are not equal, no closed-form solution exists for the achievable asymptotic sum rate.

Fig. 4.12 plots the outage against the number of antennas at the end nodes S_1 when power scaling is only used at S_1 and S_2 . The outage is plotted when the relay is selected based on the average sum rate, labeled as relay selection scheme 2 (RSS-2). Although the outage is significantly reduces when the number of relays is increased for relay selection to maximize the instantaneous sum rate scenario, RSS-2 can only obtain the performance equal to the case with a single relay antenna. As an example, with 8 relay nodes, the relay selection scheme can reduce the outage

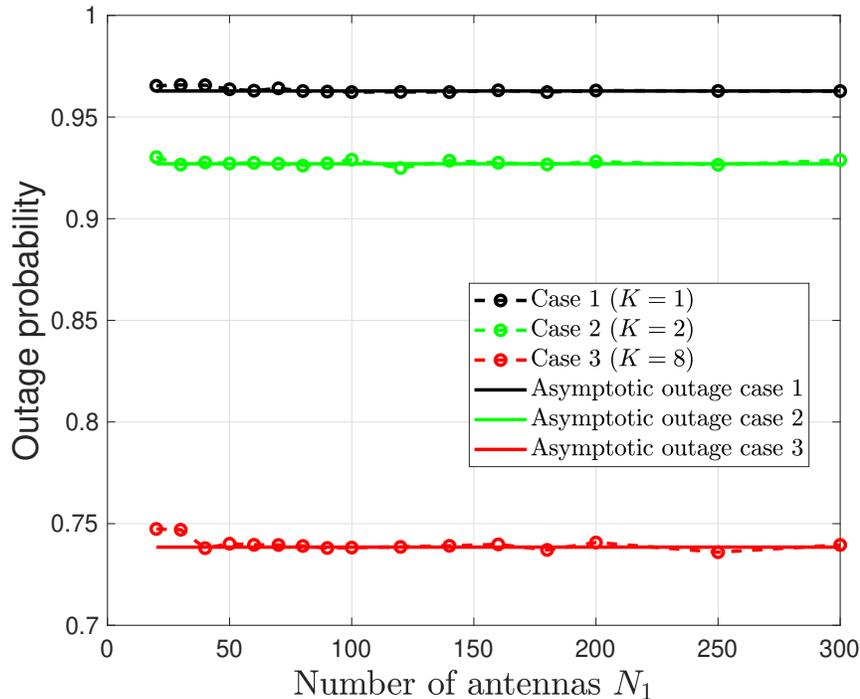


Figure 4.8: Outage as a function the number of antennas at S_1 for end-node power scaling.

asymptotically to zero while the outage is 0.35 with just a single relay. Furthermore, two relays will reduce the outage to 0.12. If RSS-2 is used, the outage value does not change even if the number of relays is increased to 8. However, the overhead used for the relay selection for RSS-2 is significantly low as the same relay is selected every time regardless of the instantaneous channel conditions. The trade-off for this simplicity will be the increased outage and eventually the reduced achievable sum rate. Moreover, when all the relays have the same number of antennas, the advantage of having more relays will be absent if the relay selection is based on the average sum rate.

4.8 Conclusions

This chapter analyzed a cognitive mMIMO TWRN with relay selection which maximizes the sum rate. Two power scaling scenarios are identified: (1) (PS-1) the transmit powers of S_1 , S_2 , and the relay are scaled and (2) (PS-2) only those of S_1 and S_2 are scaled. Furthermore, optimal power allocation was derived for S_1 and S_2 subject to interference constraints at PU. Next, the power consumption was

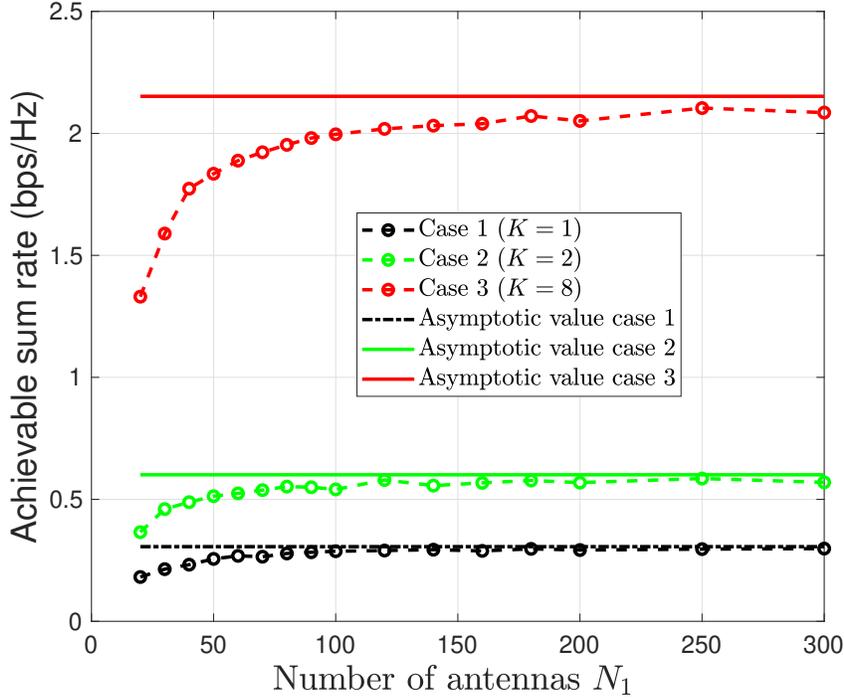


Figure 4.9: Achievable sum rate vs. the number of antennas at S_1 for end-node power scaling.

analyzed realistically for the proposed system. The analysis reveals the following.

1. It is well known that mMIMO flattens the effect of small-scale fading in mMIMO, and that holds for cognitive mMIMO as well. In particular, the asymptotic SINR and sum rate do not depend on small-scale fading for PS-1. Thus, the PS-1 relay selection can be done offline. Further, the interference on the PU can easily be reduced or eliminated when the number of antennas at the end nodes becomes extremely large. Thus, the secondary system can function with zero outage.
2. In PS-2, the relay will have a finite number of antennas. Then, the outage of the TWRN will depend on the channel conditions between the relays and the PU. Thus, although the end-to-end SINR is independent of the instantaneous channel state, the sum rate will still depend on the instantaneous values.
3. Relay selection improves the sum rate of the secondary system and reduces the interference on the PU. The first relay selection method is based on the instantaneous sum rate maximization and thus, will require instantaneous CSI

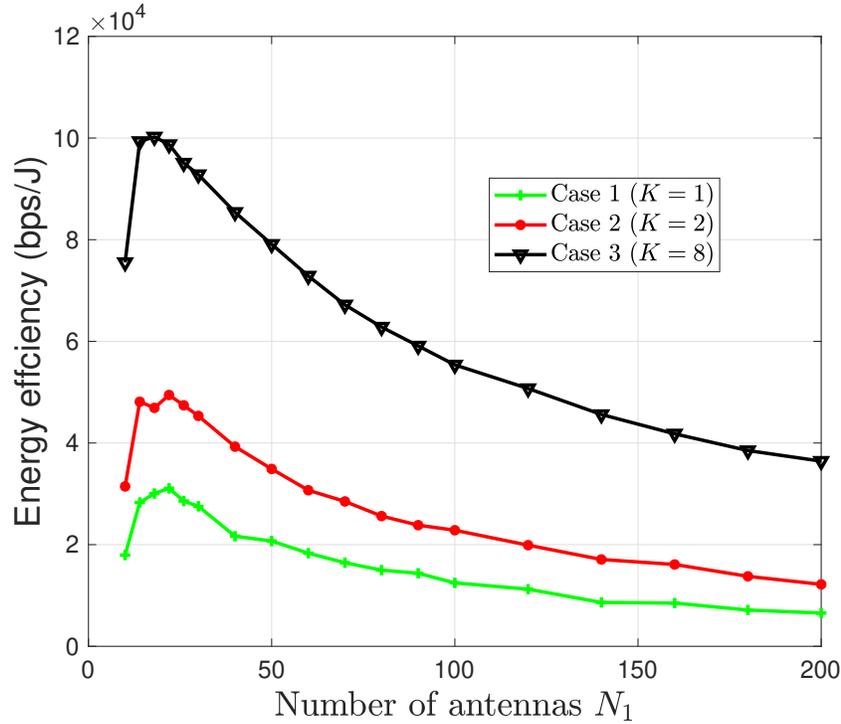


Figure 4.10: Energy efficiency vs. the number of antennas at S_1 for end-node power scaling.

between relays and PU. The second relay selection aims for the highest average sum rate and does not require instantaneous CSI. Thus, for this case, relay selection can be performed offline.

- Also, as the number of available relays are increased, the energy efficiency is improved significantly for both PS-1 and PS-2. However, as the number of antennas at the end nodes are increased, energy efficiency peaks and then declines. The reduced transmit power due to power scaling improves energy efficiency while the increased power consumption in the transceiver chains reduces energy efficiency as the number of antennas at the end nodes are increased.

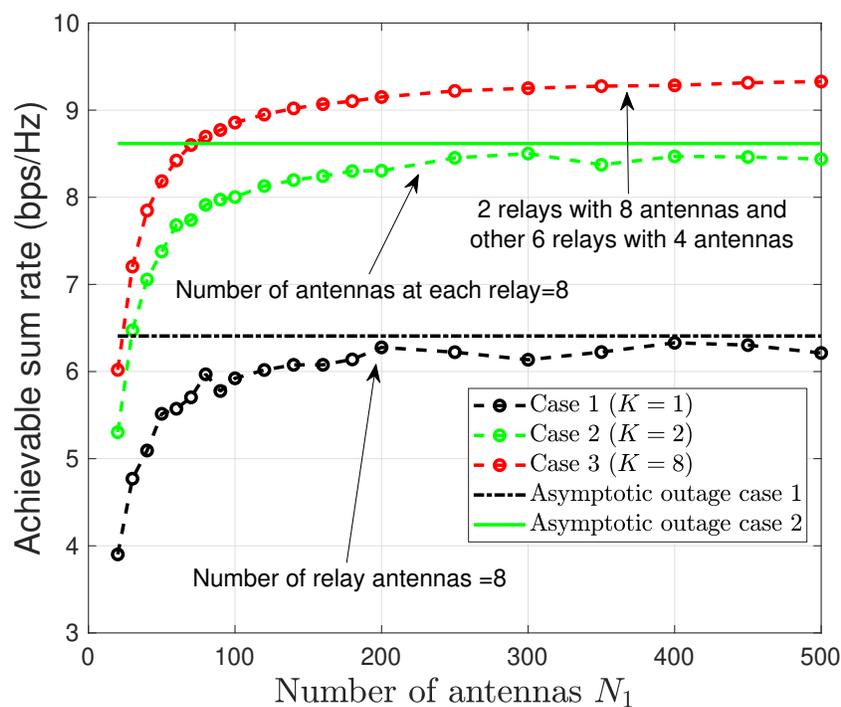


Figure 4.11: Achievable sum rate vs. the number of S_1 antennas for end-node power scaling.

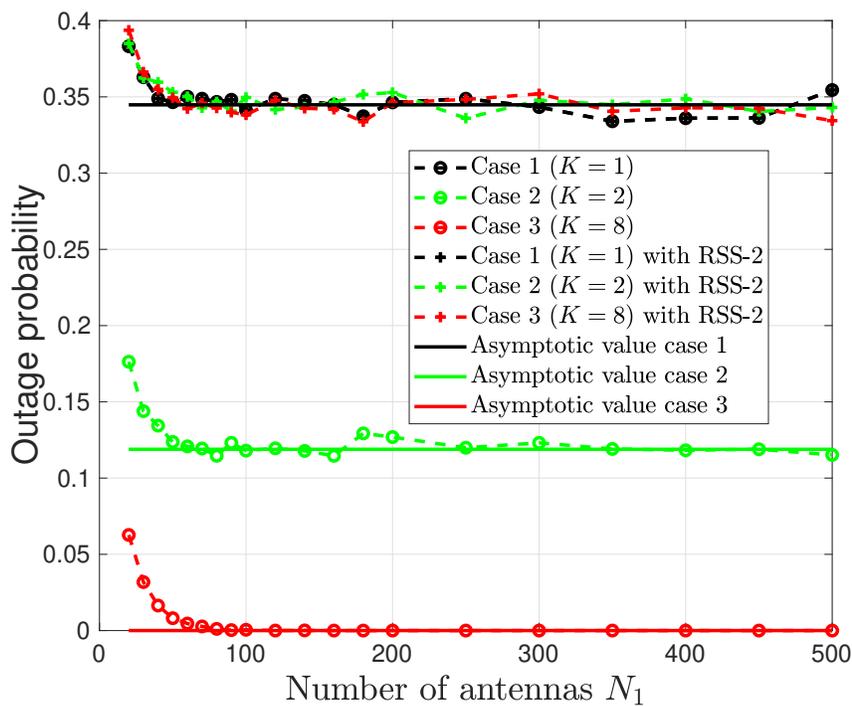


Figure 4.12: Outage vs. the number of S_1 antennas for end-node power scaling.

Chapter 5

Non-orthogonal multiple access (NOMA)-Aided Multi-Way mMIMO Relaying

This chapter proposes a multi-way relay network (MWRN) transmission scheme that can complete the full data exchange among any number of users within two time slots. This is accomplished by exploiting the performance gains of NOMA and massive multiple-input multiple-output (mMIMO). First, the users transmit their signals to the relay, which uses maximal ratio combining (MRC). Next, the relay transmits a superposition-coded signal for all users by using maximal ratio transmission (MRT). Each user then performs successive interference cancellation (SIC) decoding of data symbols of the other user nodes. The derived closed form overall sum rates demonstrate significant spectral-efficiency gains and energy-efficiency gains over the existing MWRN counterparts. Furthermore, the relay power allocation matrix is designed to maximize the minimum sum rate among the users, thus maximizing the user fairness. Furthermore, the effects of imperfect SIC and imperfect channel-state information (CSI) on the sum rate are analyzed.

5.1 Introduction

This chapter considers the fundamental question of improving the spectral efficiency of a K -user MWRN with full data exchange. In the basic configuration, with an intermediate multiple-input multiple-output (MIMO) relay, this is possible in exactly K orthogonal time slots [38]. Similarly, the MWRN protocols in [40] and [99] require K time slots for K user nodes. Thus, the overall spectral efficiency is low due to the $1/K$ pre-log factor, appearing in achievable rate expressions, and hence

it diminishes as the number of users grows. To attack this fundamental issue, [100] develops a new MWRN transmit protocol to do this in $\lceil (K - 1)/2 \rceil + 1$ time slots, where $\lceil x \rceil$ is the ceiling function. This number is roughly $K/2$. Thus, this state-of-the-art protocol [100] halves the number of time slots, equivalently doubling the spectral efficiency. This improvement stems from the adoption of linear processing, self-interference cancellation, and SIC decoding. However, the achievable sum rate of [100] still suffers the $O(1/K)$ decline of spectral efficiency with the number of users.

Herein, a novel MWRN protocol is proposed, which reduces the number of time slots to just two, regardless of the number of users. Thus, it can provide significant spectral efficiency gains compared to those of [40, 99, 100]. To realize these remarkable gains, the concept of power-domain NOMA is integrated with mMIMO MWRNs. It is designed as follows. In the first time-slot, all user nodes transmit simultaneously to the relay, which in turn uses MRC for reception. The relay then generates a superposition-coded signal, applies an amplification factor, and transmits back to the user nodes by using a MRT precoder. Then each user node performs SIC to decode the symbols belonging to the remaining $K - 1$ users' messages.

Critically important for NOMA, the power allocation matrix allows different user signals be assigned with different power levels in order to achieve the desired performance targets. Thus, it can be designed to achieve max-min fairness, sum rate maximization, and energy efficiency maximization [62, 126]. This chapter designs the power allocation to maximize the minimum achievable sum rate of the MWRN. In addition, the performance of NOMA degrades heavily due to imperfect SIC [58]. Indeed, modelling this effect has been the focus of several recent papers. Imperfect SIC has been modelled as a portion of the decoded signal in [127] and as the estimation error of the decoded signal in [128, 129]. This chapter uses the model of [128, 129] to analyze the effect of SIC on the data rate of the proposed NOMA MWRN protocol.

More specifically, the contributions of this chapter are summarized as follows.

1. A NOMA-based mMIMO MWRN transmit protocol is proposed to enable full-mutual data exchange among $K > 2$ users. The key feature is that it uses just two time slots. Thus, it potentially achieves a sum rate gain of $K/2$ (approximately) over the current state-of-the-art counterpart in [100], and the

gain scales up with the number of user nodes (K). Of course, in actuality, the gain could be less depending on the interference issues, imperfect CSI, and imperfect SIC.

2. Closed-form results are obtained for the sum rate of the proposed scheme by using the so-called additive white Gaussian noise (AWGN) approximation [130]. The simulations and numerical results confirm the tightness of this approximation and compare the sum rate performances of the proposed MWRN protocol against the existing counterparts.
3. To reflect more accurately on practical systems, it is assumed that nodes do not have perfect CSI. In practical systems, CSI acquisition is done through uplink pilot transmissions and inevitable estimation errors degrade the performance of the system. Thus, the effects of both imperfect SIC and imperfect/erroneous CSI are analysed. They both degrade the performance, especially when the number of users increases.
4. The energy efficiency of the proposed scheme is analysed and it is shown that the proposed scheme provides significant energy efficiency gains compared to other methods.
5. The asymptotic sum rate value when the number of antennas at the relay grows unbounded is also obtained.
6. A power allocation matrix at the relay is proposed based upon the asymptotic sum rate values. A closed-form solution for the matrix based on the asymptotic sum rates of the system is obtained.

This protocol achieves a full data exchange among K spatially-distributed user nodes in exactly two time-slots. This amounts to a time-slot reduction of $(1 - 4/K) \times 100\%$ over the current state-of-the-art MWRN protocol [100]. For example, with $K = 8$ users, the time-slot saving is 50%. This reduction directly translates into a significant spectral efficiency gain over all existing MWRN counterparts [40, 99, 100]. This gain is the result of power-domain NOMA and mMIMO via superposition coding, SIC, and linear detection/precoding.

As the number of time slots required in the proposed scheme is constant regardless of the number of users, it may be a helpful step in the context of massive

connectivity envisaged in Fifth Generation (5G) and beyond, where a massive number of new Internet of Things (IoT) devices will connect to a next-generation wireless network. Industry estimates that the total number of IoT connected devices will be 14.7B by 2023 [131]. They will generate both data and connection traffic. Clearly, some of these devices could be supported in the MWRN configuration, and the spectral efficiency gains of the proposed protocol may help to mitigate the overall traffic growth.

5.2 System, channel, and signal model

5.2.1 System and channel model

This chapter considers a system with K users and denotes the k -th user by S_k , $k \in \{1, \dots, K\}$. They all are single-antenna terminals. The data exchange requirement can be stated as follows. For all $k \in \{1, \dots, K\}$, S_k must transmit its data to the remaining $K - 1$ users and must receive data from all of them too. The relay, R , is equipped with M antennas and is of mMIMO type ($M \gg K$). Here, R is a specialized relay with a fixed higher power availability than users. The direct channels between the user pairs are not available due to unfavorable channel propagation conditions [32, 35] or not utilized in order to minimize mutual interference among the users. Thus, the purpose of R is to accommodate the data transfer among the users.

The wireless channel between S_k and R is represented as $\mathbf{h}_k = \sqrt{\beta_k} \tilde{\mathbf{h}}_k \in \mathcal{C}^M$ where β_k is the large-scale fading coefficient (may include range-dependent path loss and shadow fading), and $\tilde{\mathbf{h}}_k$ is the small-scale fading vector, which is distributed as

$$\tilde{\mathbf{h}}_k \sim \mathcal{CN}(0, \mathbf{I}_M). \quad (5.1)$$

The matrix $\tilde{\mathbf{H}} = [\tilde{\mathbf{h}}_1, \dots, \tilde{\mathbf{h}}_K] \in \mathcal{C}^{M \times K}$ represents the small-scale channel fading channel coefficients from all the users towards the relay. Thus, this channel matrix incorporates both large-scale and small-scale propagation effects, and it may be expressed as

$$\mathbf{H} = \tilde{\mathbf{H}}\mathbf{D}^{1/2}, \quad (5.2)$$

where $\mathbf{D} = \text{diag}(\beta_1, \dots, \beta_K)$.

5.2.2 Channel estimation

To design the MRC/MRT detector/precoder (\mathbf{W}_R and \mathbf{W}_T), the relay requires CSI between itself and the users. This CSI could be estimated via the conventional estimation techniques. The most common technique is the transmission of sequences of known training symbols (pilots) periodically [17, 132]. Alternatively, blind and semi-blind techniques have also been developed (see [133, 134] and the references therein). It is assumed that the users simultaneously transmit orthogonal pilot sequences to the relay, to avoid mutual interference. The set of orthogonal pilot sequences are given as $\Phi = \{\Phi_1, \dots, \Phi_K\}^T$ where Φ_k is the $1 \times \tau$ pilot sequence of the k -th user. Here, τ is the length of the pilot sequence used for channel estimation. Because Φ_k s are mutually orthogonal, the matrix is unitary ($\Phi\Phi^H = \mathbf{I}_K$). The received signal at the relay during the pilot transmission period is given as

$$\mathbf{Y}_p = \sqrt{P}\mathbf{H}\mathbf{A}\Phi + \mathbf{N}_p, \quad (5.3)$$

where \mathbf{N}_p is the AWGN at the relay with $\mathcal{CN}(0, \mathbf{I}_M)$ distribution and P is the transmit power of the users. Also $\mathbf{A} = \text{diag}(\sqrt{\alpha_{p,1}}, \dots, \sqrt{\alpha_{p,K}})$ is the power scaling coefficient matrix, where $\alpha_{p,k}$ corresponds to the coefficient used by S_k during the pilot transmission. Relay multiplies the above received signal by Φ^H to estimate the channels [132] and obtain

$$\mathbf{y}_k = [\mathbf{Y}_p\Phi^H]_k = \sqrt{P}\sqrt{\alpha_{p,k}\beta_k}\tilde{\mathbf{h}}_k + \mathbf{n}_{p,k}, \quad (5.4)$$

where $\mathbf{n}_{p,k} = [\mathbf{N}_p\Phi^H]_k \sim \mathcal{CN}(0, \mathbf{I}_M)$. Based on the above result and by using minimum mean square error (MMSE) criterion [132], the channel estimate for \mathbf{h}_k is given as [51]

$$\hat{\mathbf{h}}_k = \frac{\sqrt{P\alpha_{p,k}\beta_k}}{P\alpha_{p,k}\beta_k + 1}\mathbf{y}_k. \quad (5.5)$$

For beamforming purposes, the estimated channel $\hat{\mathbf{h}}_k$ is used. The true channel \mathbf{h}_k can be written in terms of its estimate by virtue of orthogonality principle of MMSE criterion as [132]

$$\mathbf{h}_k = \hat{\mathbf{h}}_k + \mathbf{e}_k, \quad (5.6)$$

where \mathbf{e}_k is the error vector, which is independent from $\hat{\mathbf{h}}_k$. The probability distributions of the k th element of estimate ($\hat{\mathbf{h}}_k$) and the error terms (\mathbf{e}_k) are $\mathcal{CN}\left(0, \frac{P\alpha_{p,k}\beta_k^2}{P\alpha_{p,k}\beta_k + 1}\right)$ and $\mathcal{CN}\left(0, \frac{\beta_k}{P\alpha_{p,k}\beta_k + 1}\right)$, respectively.

5.2.3 Signal model

Data transmission among the users requires two time slots. In the first time slot, all the users transmit to the relay R , which applies receive beamforming. Thus, for $k = 1, \dots, K$, user S_k transmits the signal

$$\bar{x}_k = \sqrt{\alpha_k P} x_k, \quad (5.7)$$

where x_k is the data symbol, P is the allowable maximum transmit power (assumed to be equal for all the users) and $0 < \alpha_k \leq 1$ is the power scaling factor of the k -th user. The received signal at the relay is the sum of all user signals and the additive noise, which is given as

$$\mathbf{y}_r = \sqrt{P} \mathbf{H} \alpha^{1/2} \mathbf{x} + \mathbf{n}_R, \quad (5.8)$$

where $\mathbf{x} = [x_1, \dots, x_K]^T$, $\alpha = \text{diag}(\alpha_1, \dots, \alpha_K)$, and \mathbf{n}_R is $M \times 1$ AWGN vector at the relay satisfying $\mathbb{E}[\mathbf{n}_R^H \mathbf{n}_R] = \mathbf{I}_M \sigma_R^2$. Next, the relay applies receive beamforming by multiplying by the matrix, \mathbf{W}_R and the processed signal can be given as

$$\mathbf{y}_p = \mathbf{W}_R \mathbf{y}_r = \mathbf{W}_R \left(\sqrt{P} \mathbf{H} \alpha^{1/2} \mathbf{x} + \mathbf{n}_R \right). \quad (5.9)$$

In the second time slot, the relay transmits the following superposition-coded signal to all the users:

$$\mathbf{y}_t = \Psi \mathbf{W}_T \mathbf{\Lambda} \mathbf{y}_p = \Psi \mathbf{W}_T \begin{bmatrix} \lambda_{1,1} & \lambda_{1,2} & \cdots \\ \lambda_{1,2} & \lambda_{2,2} & \cdots \\ \vdots & \ddots & \vdots \\ \lambda_{K,1} & \cdots & \lambda_{K,K} \end{bmatrix} \mathbf{W}_R \mathbf{y}_r, \quad (5.10)$$

where \mathbf{W}_T is the MRT precoder at the relay, $\mathbf{\Lambda}$ is the $K \times K$ power allocation matrix at the relay, and Ψ is the power control factor. The selection of MRC and MRT at the relay is due to the simplicity of those methods compared to other linear beamformers. The total power constraint at the relay is given as

$$P_R = \text{Tr}(\mathbf{y}_t \mathbf{y}_t^H) = \Psi^2 P \text{Tr}(\mathbf{V} \mathbf{H} \alpha \mathbf{H}^H \mathbf{V}^H) + \Psi^2 \sigma_R^2 \text{Tr}(\mathbf{V} \mathbf{V}^H), \quad (5.11)$$

where P_R is the transmit power at the relay and $\mathbf{V} = \mathbf{W}_T \mathbf{\Lambda} \mathbf{W}_R$ is the effective/cascaded detector/precoder at the relay, and the relay gain Ψ is computed to constrain the average transmit power as

$$\Psi = \sqrt{\frac{P_R}{P \mathbb{E}[\text{Tr}(\mathbf{V} \mathbf{H} \alpha \mathbf{H}^H \mathbf{V}^H)] + \sigma_R^2 \mathbb{E}[\text{Tr}(\mathbf{V} \mathbf{V}^H)]}}. \quad (5.12)$$

Based on (5.10), the intended transmit signal for S_k is given as $[\mathbf{y}_t]_k$, which is the k th row of \mathbf{y}_t . The received signal at S_k is given as

$$y_k = \Psi \mathbf{h}_k^T \mathbf{V} \sum_{m=1}^K \sqrt{P\alpha_m} \mathbf{h}_m x_m + \Psi \mathbf{h}_k^T \mathbf{V} \mathbf{n}_R + n_k, \quad (5.13)$$

where n_k is an AWGN at S_k with power σ_k^2 .

5.2.4 Imperfect SIC decoding at the user nodes

After receiving y_k , S_k will decode the symbols (i.e., all $x_{k'}$ for $k' \neq k$ and $k' \in \{1, \dots, K\}$) transmitted by all other users. It will use SIC decoding for this, which is an iterative process. In each step, a symbol x_l is decoded, and subtracted (i.e., cancelled) from y_k . Due to K users in total, $K - 1$ decoding steps occur at each user. The index of the user that has to be decoded at the n -th iteration at S_k is denoted by the function $f_k(n)$. Thus, $1 \leq n \leq K - 1$ and $1 \leq f_k(n) \leq K$ with $f_k(n) \neq k$.

At the n th decoding step, S_k will decode the signal transmitted by user $S_{f_k(n)}$, denoted as $x_{f_k(n)}$. The estimate of $x_{f_k(n)}$ at S_k is represented by $\hat{x}_{k,f_k(n)}$. Both of these are assumed to be jointly Gaussian distributed with a normalized correlation coefficient of $\rho_{k,f_k(n)}$ and may be expressed as [128]

$$\hat{x}_{k,f_k(n)} = \rho_{k,f_k(n)} x_{f_k(n)} + e_{k,f_k(n)}, \quad (5.14)$$

where the estimation error $e_{k,f_k(n)} \sim \mathcal{CN}\left(0, \sigma_{e_{f_k(n)}}^2 / \sqrt{1 + \sigma_{e_{f_k(n)}}^2}\right)$, the estimate $\hat{x}_{k,f_k(n)} \sim \mathcal{CN}(0, 1)$, and the correlation coefficient $\rho_{k,f_k(n)} = 1 / \sqrt{1 + \sigma_{e_{f_k(n)}}^2}$. Furthermore, the estimation error and the estimated value are statistically independent. The perfect SIC case is specified by the values $\rho_{k,f_k(n)} = 1$ and $\sigma_{e_{f_k(n)}}^2 = 0$. Next section investigates the effect of imperfect SIC on the system performance.

5.3 Achievable sum rate analysis

The achievable sum rate between each user pairs is derived by using the worst-case Gaussian technique [130]. The residual signal at S_k after decoding the cancelling signals belonging to $n - 1$ users can be written as

$$y_{k,n} = \Psi \sqrt{P\alpha_{f_k(n)}} \mathbb{E} \left[\mathbf{h}_k^T \mathbf{V} \mathbf{h}_{f_k(n)} x_{f_k(n)} \right] + \hat{\mathbf{n}}_{\mathbf{k},\mathbf{n}}, \quad (5.15)$$

where $\hat{\mathbf{n}}_{\mathbf{n},\mathbf{k}}$ is the effective noise and the first term is the desired signal. The function $f_k(n)$ is the index of the user that will be decoded in the n -th step. Moreover, the

noise term in (5.15) can be expressed as

$$\begin{aligned}
\hat{\mathbf{n}}_{\mathbf{k},\mathbf{n}} &= \underbrace{\Psi \sqrt{P\alpha_{f_k(n)}} \left(\mathbf{h}_k^T \mathbf{V} \mathbf{h}_{f_k(n)} x_{f_k(n)} - \mathbb{E} \left[\mathbf{h}_k^T \mathbf{V} \mathbf{h}_{f_k(n)} x_{f_k(n)} \right] \right)}_{\text{detection uncertainty}} \\
&+ \underbrace{\sum_{m'=1}^{n-1} \Psi \sqrt{P\alpha_{f_k(m')}} \left(\mathbf{h}_k^T \mathbf{V} \mathbf{h}_{f_k(m')} x_{f_k(m')} - \mathbb{E} \left[\mathbf{h}_k^T \mathbf{V} \mathbf{h}_{f_k(m')} \right] \hat{x}_{f_k(m'),k} \right)}_{\text{effect of imperfect SIC}} \\
&+ \underbrace{\Psi \mathbf{h}_k^T \mathbf{V} \sum_{m=n+1}^{K-1} \sqrt{P\alpha_{f_k(m)}} \mathbf{h}_{f_k(m)} x_{f_k(m)}}_{\text{interference from other users}} \\
&+ \underbrace{\Psi \mathbf{h}_k^T \mathbf{V} \mathbf{n}_R}_{\text{amplified noise}} + \underbrace{n_k}_{\text{AWGN noise at the receiver}}. \tag{5.16}
\end{aligned}$$

The matrix \mathbf{V} in (5.15) and (5.16) for MRC/MRT beamforming is given as

$$\mathbf{V} = \hat{\mathbf{H}}^* \mathbf{\Lambda} \hat{\mathbf{H}}^H. \tag{5.17}$$

Based on (5.15) and (5.16), and by assuming that the additive noise is independently distributed Gaussian noise having the same variance [130], a tight approximation for the achievable sum rate can be given as

$$\mathcal{R}_{k,f_k(n)} = \frac{(T_C - \tau)}{2T_C} \log \left(1 + \frac{\Psi^2 P \alpha_{f_k(n)} M_{k,n}^2}{\Psi^2 \mathcal{J} + \sigma_k^2} \right). \tag{5.18}$$

Here $\mathcal{J} = P \alpha_{f_k(n)} N_{k,n} + \sum_{m=n+1}^{K-1} P \alpha_{f_k(m)} P_{k,m} + \sum_{m'=1}^{n-1} P \alpha_{f_k(m')} R_{k,m'} + \sigma_R^2 Q_k$ and T_C is the coherence time of the channel. The pre-log factor $(T_C - \tau)/T_C$ accounts for the pilot overhead [17]. The two time-slots required for the data transmission between the users results in the pre-log factor of 1/2. The values of $M_{k,m}$, $N_{k,m}$, $P_{k,m}$, $R_{k,m}$, and Q_k in (5.18) are given as

$$M_{k,m} = \mathbb{E} \left[\mathbf{h}_k^T \hat{\mathbf{H}}^* \mathbf{\Lambda} \hat{\mathbf{H}}^H \mathbf{h}_{f_k(m)} \right], \tag{5.19a}$$

$$N_{k,m} = \mathbb{V} \left[\mathbf{h}_k^T \hat{\mathbf{H}}^* \mathbf{\Lambda} \hat{\mathbf{H}}^H \mathbf{h}_{f_k(m)} \right], \tag{5.19b}$$

$$P_{k,m} = \mathbb{E} \left[\left| \mathbf{h}_k^T \hat{\mathbf{H}}^* \mathbf{\Lambda} \hat{\mathbf{H}}^H \mathbf{h}_{f_k(m)} \right|^2 \right], \tag{5.19c}$$

$$R_{k,m} = \mathbb{E} \left[\left| \mathbf{h}_k^T \hat{\mathbf{H}}^* \mathbf{\Lambda} \hat{\mathbf{H}}^H \mathbf{h}_{f_k(m)} \right|^2 \right] + (1 - 2\rho_{k,f_k(m)}) \mathbb{E} \left[\mathbf{h}_k^T \hat{\mathbf{H}}^* \mathbf{\Lambda} \hat{\mathbf{H}}^H \mathbf{h}_{f_k(m)} \right]^2, \tag{5.19d}$$

$$Q_k = \mathbb{E} \left[\left\| \mathbf{h}_k^T \hat{\mathbf{H}}^* \mathbf{\Lambda} \hat{\mathbf{H}}^H \right\|^2 \right]. \tag{5.19e}$$

The value of $R_{k,m}$ can further be written as

$$R_{k,m} = P_{k,m} + (1 - 2\rho_{k,f_k(m)}) M_{k,m}^2. \tag{5.20}$$

The closed-form evaluations of (5.19) are provided in Appendix C.1. The value for Ψ is given as

$$\Psi = \sqrt{\frac{P_R}{P L_1 + \sigma_R^2 L_2}}, \quad (5.21)$$

where L_1 and L_2 are given as

$$L_1 = \text{Tr} \left(\mathbb{E} \left[\hat{\mathbf{H}}^* \boldsymbol{\Lambda} \hat{\mathbf{H}}^H \mathbf{H} \boldsymbol{\alpha} \mathbf{H}^H \hat{\mathbf{H}} \boldsymbol{\Lambda}^H \hat{\mathbf{H}}^T \right] \right), \quad (5.22a)$$

$$L_2 = \text{Tr} \left(\mathbb{E} \left[\hat{\mathbf{H}}^* \boldsymbol{\Lambda} \hat{\mathbf{H}}^H \hat{\mathbf{H}} \boldsymbol{\Lambda}^H \hat{\mathbf{H}}^T \right] \right), \quad (5.22b)$$

and are derived in Appendix C.2.

The rate of data transmission for each user is limited by the achievable sum rates among itself and all other users. Thus, the achievable transmission rate of S_m is given as

$$\mathcal{R}_m = \min_{k \in (1, \dots, K), k \neq m} (\mathcal{R}_{k,m}). \quad (5.23)$$

Based on this, the total achievable sum rate of the system is obtained as

$$\mathcal{R} = (K - 1) \sum_{m=1}^K \mathcal{R}_m. \quad (5.24)$$

Here in (5.24), the factor $(K - 1)$ represents the transmission of the data of each user to all the other users.

5.4 Asymptotic sum rate analysis

Asymptotic refers to the fact that the number of relay antennas M grows unbounded. The significance of this condition is that the relay can then scale down the transmit power inversely proportional to the number of antennas [135], which tends to improve the energy efficiency overall. Therefore, it is important to find the sum rate under this condition. The relay transmit power may thus be expressed as

$$P_R = E_R/M, \quad (5.25)$$

where E_R is fixed. First an asymptotic limit for Ψ is derived for the transmit power control factor at the relay. To obtain Ψ , the limits of L_1 and L_2 for extremely large values of M are needed. Generalized forms of the law of large numbers is utilized. The details are given in Appendix C.4 and by using the fact that $\hat{\mathbf{H}}$ and \mathbf{E} are independent of each other, the following asymptotic result is obtained:

$$\frac{L_1}{M^3} \xrightarrow[M \rightarrow \infty]{a.s.} \sum_{i=1}^K \sum_{j=1}^K \alpha_j \lambda_{i,j}^2 \hat{\beta}_i \hat{\beta}_j^2. \quad (5.26)$$

Here, the $\hat{\beta}_k$ values are defined as (C.5) in Appendix C.1. Similarly, an asymptotic limit for L_2 is derived as

$$\frac{L_2}{M^2} \xrightarrow[M \rightarrow \infty]{a.s.} \sum_{i=1}^K \sum_{j=1}^K \lambda_{i,j}^2 \hat{\beta}_i \hat{\beta}_j. \quad (5.27)$$

By using (5.26) and (5.27), the limit for Ψ can be obtained as

$$M^2 \Psi \xrightarrow[M \rightarrow \infty]{a.s.} \sqrt{\frac{E_R}{E \sum_{i=1}^K \sum_{j=1}^K \alpha_j \lambda_{i,j}^2 \hat{\beta}_i \hat{\beta}_j}} = \Psi^\infty. \quad (5.28)$$

By using the above asymptotic results and the value for \mathbf{V} (5.15) is written as

$$\begin{aligned} y_{k,n} &= M^2 \Psi \sqrt{P \alpha_{f_k(n)}} \frac{\mathbf{h}_k^T \hat{\mathbf{H}}^*}{M} \mathbf{\Lambda} \frac{\hat{\mathbf{H}}^H \mathbf{h}_{f_k(n)}}{M} x_{f_k(n)} \\ &+ M^2 \Psi \sum_{m=n+1}^{K-1} \sqrt{P \alpha_{f_k(m)}} \frac{\mathbf{h}_k^T \hat{\mathbf{H}}^*}{M} \mathbf{\Lambda} \frac{\hat{\mathbf{H}}^H \mathbf{h}_{f_k(m)}}{M} x_{f_k(m)} \\ &+ M^2 \Psi \sum_{m'=1}^{n-1} \sqrt{P \alpha_{f_k(m')}} \frac{\mathbf{h}_k^T \hat{\mathbf{H}}^*}{M} \mathbf{\Lambda} \frac{\hat{\mathbf{H}}^H \mathbf{h}_{f_k(m')}}{M} (x_{f_k(m')} - \hat{x}_{k,f_k(m')}) \\ &+ M^2 \Psi \frac{\mathbf{h}_k^T \hat{\mathbf{H}}^*}{M} \mathbf{\Lambda} \frac{\hat{\mathbf{H}}^H \mathbf{n}_R}{M} + n_k. \end{aligned} \quad (5.29)$$

The asymptotic results for each term in (5.29) are derived as follows:

$$\frac{\mathbf{h}_k^T \hat{\mathbf{H}}^*}{M} \mathbf{\Lambda} \frac{\hat{\mathbf{H}}^H \mathbf{h}_{f_k(n)}}{M} x_{f_k(n)} \xrightarrow[M \rightarrow \infty]{a.s.} \lambda_{k,f_k(n)} \hat{\beta}_k \hat{\beta}_{f_k(n)}, \quad (5.30)$$

$$\frac{\mathbf{h}_k^T \hat{\mathbf{H}}^*}{M} \mathbf{\Lambda} \frac{\hat{\mathbf{H}}^H \mathbf{h}_{f_k(m)}}{M} x_{f_k(m)} \xrightarrow[M \rightarrow \infty]{a.s.} \lambda_{k,f_k(m)} \hat{\beta}_k \hat{\beta}_{f_k(m)}, \quad (5.31)$$

$$\frac{\mathbf{h}_k^T \hat{\mathbf{H}}^*}{M} \mathbf{\Lambda} \frac{\hat{\mathbf{H}}^H \mathbf{n}_R}{M} \xrightarrow[M \rightarrow \infty]{a.s.} 0. \quad (5.32)$$

The asymptotic limit of the effect of imperfect SIC can be written as

$$\frac{\mathbf{h}_k^T \hat{\mathbf{H}}^*}{M} \mathbf{\Lambda} \frac{\hat{\mathbf{H}}^H \mathbf{h}_{f_k(m')}}{M} (x_{f_k(m')} - \hat{x}_{k,f_k(m')}) \xrightarrow[M \rightarrow \infty]{a.s.} (1 - \rho_{k,f_k(m')}) \lambda_{k,f_k(m')} \hat{\beta}_k \hat{\beta}_{f_k(m')}. \quad (5.33)$$

By using the aforementioned asymptotic results, the asymptotically achievable sum rate is given as

$$\mathcal{R}_{k,\hat{n}}^\infty = \frac{(T_C - \tau)}{2T_C} \log \left(1 + \frac{P (\Psi^\infty)^2 \alpha_{\hat{n}} \lambda_{k,\hat{m}}^2 \hat{\beta}_k^2 \hat{\beta}_{\hat{n}}^2}{P (\Psi^\infty)^2 \mathcal{H} + \sigma_k^2} \right), \quad (5.34)$$

where $\hat{m} = f_k(m)$ and $\mathcal{H} = \sum_{m'=1}^{n-1} (1 - \rho_{k,\hat{m}'})^2 \alpha_{\hat{m}'} \lambda_{k,\hat{m}'}^2 \hat{\beta}_k^2 \hat{\beta}_{\hat{m}'}^2 + \sum_{m=n+1}^{K-1} \alpha_{\hat{m}} \lambda_{k,\hat{m}}^2 \hat{\beta}_k^2 \hat{\beta}_{\hat{m}}^2$.

5.5 Energy efficiency of the system

This section analyzes the energy efficiency of the proposed MWRN. Energy efficiency, i.e., the number of information bits per unit of transmit energy, has been studied extensively [41]. It is defined as

$$\rho = \frac{\mathcal{R}}{P_{Tot}}, \quad (5.35)$$

where P_{tot} is the total power consumption and \mathcal{R} is the overall sum rate. The value for P_{tot} can be written as follows [125]

$$P_{Tot} = KP + P_R + P_{U,TC} + KP_{R,TC} + P_{C/D} + P_{R,LP}, \quad (5.36)$$

where $P_{U,TC}$ and $P_{R,TC}$ are the power consumed in the transceiver chains in each user and the relay, $P_{C/D}$ is the power consumed for coding and decoding, and $P_{R,LP}$ is the power consumed in the relay to perform for linear processing. The values for these power consumption components are given as follows [125]:

$$P_{U,TC} = 2P_{U,C} + 2P_{SYN} \quad P_{R,TC} = 2MP_{R,C} + 2P_{SYN}, \quad (5.37)$$

$$P_{C/D} = 2\mathcal{R}(P_{COD} + P_{DEC}), \quad (5.38)$$

$$P_{R,LP} = \frac{2BKM}{L_{BS}} + \frac{B}{U} \frac{3MK}{L_{BS}}, \quad (5.39)$$

where $P_{U,C}$ and $P_{R,C}$ are the powers required to run the circuit components at the users and the relay, P_{SYN} is the power of the local oscillator, P_{COD} and P_{DEC} is the coding and decoding power consumption, B is the bandwidth, L_{BS} is the computational efficiency (given in flops/W) of the end nodes, and U is the coherence block.

5.6 Comparison with other MWRN operations

The existing MWRN protocols in [40, 100] are analysed for comparison purposes. First, the MWRN in [40] requires K time slots to transmit the data of all the users to all other users. The use of time slots in [40] can be summarized as follows:

1. **Time-slot 1:** All the users transmit to the relay. This step is similar to the first step in the proposed NOMA-aided protocol.

2. **Time-slots 2 to K:** In these time slots, the relay transmits to all the users using beamforming. However, instead of sending a superposition-coded signal of all the other received signals, the data of a single user is transmitted to each user. Thus, $K - 1$ time slots are required to send the data of all the users to all other users.

The beamforming matrix at the relay for the j -th time slot ($2 \leq j \leq K$) is given as

$$\mathbf{V}_j = \mathbf{H}^* \mathbf{\Lambda}_j \mathbf{H}^H. \quad (5.40)$$

In (5.40), $\mathbf{\Lambda}_j$ is a permutation matrix in which each row consists of a single one and all zeros. The location of the number one, decides the transmitted signal of the initial set of users. The approximation for the end-to-end data rate between each pair of users can be obtained by using the same steps as the previous case. However, when calculating the achievable data rate, the pre-log factor $1/K$ is used, as K total time slots are required for the data transmission.

Secondly, the performance of the proposed MWRN is compared with that of [100], which utilizes $\lceil (K - 1)/2 \rceil + 1$ time-slots. The above two methods (i.e. [40] and [100]) are used as performance benchmarks.

5.7 Design of power allocation matrix

This section analyzes the design of $\mathbf{\Lambda}$ while satisfying the relay power constraints. By changing the values in $\mathbf{\Lambda}$, different sum rates can be achieved for different users in the system. Several optimization problems can be formulated to design $\mathbf{\Lambda}$ based on different criteria such as maximizing the minimum data rate (max-min fairness), maximizing the total sum rate, etc. The objective functions of these problems are constrained by three variables, namely $\mathbf{\Lambda}$, $\boldsymbol{\alpha}$, and $f_k(n)$. Here, $\mathbf{\Lambda}$ is a $K \times K$ matrix, while $\boldsymbol{\alpha} = \{\alpha_n, \dots, \alpha_K\}$ is a vector of length K . Furthermore, $f_k(n)$ is a $\mathbb{R}^2 \rightarrow \mathbb{R}$ function which can be represented by $K \times (K - 1)$ matrix.

The achievable rates that are used in the optimizing problems under the worst-case Gaussian technique can be written as

$$\mathcal{R}_{k,\hat{n}} = \frac{(T_C - \tau)}{2T_C} \log \left(1 + \frac{\Psi^2 \alpha_{\hat{n}} \left(\lambda_{k,\hat{n}}^2 A_{k,\hat{n}} + \lambda_{k,\hat{n}} \lambda_{\hat{n},k} B_{k,\hat{n}} + \lambda_{\hat{n},k}^2 C_{k,\hat{n}} \right)}{\Psi^2 \alpha_{\hat{n}} \mathcal{G} + \mathcal{Z}} \right). \quad (5.41)$$

Here, $\mathcal{G} = \sum_{i=1}^K \sum_{j=1}^K (\lambda_{i,\hat{n}} \lambda_{j,\hat{n}} D_{i,j,\hat{n}} + \lambda_{i,k} \lambda_{j,k} D_{i,j,k} + \lambda_{\hat{n},i} \lambda_{\hat{n},j} E_{i,j,\hat{n}} + \lambda_{k,i} \lambda_{k,j} E_{i,j,k})$ and \mathcal{Z}

is given as

$$\begin{aligned} \mathcal{Z} = & \sum_{m=n+1}^{K-1} \sum_{i=1}^K \sum_{j=1}^K (\lambda_{i,m} \lambda_{j,m} G_{i,j,m} + \lambda_{i,k} \lambda_{j,k} G_{i,j,k} \\ & + \lambda_{m,i} \lambda_{m,j} H_{i,j,m} + \lambda_{k,i} \lambda_{k,j} H_{i,j,k}) + \sigma_k^2, \end{aligned} \quad (5.42)$$

where $A_{k,\hat{n}}$, $B_{k,\hat{n}}$, $C_{k,\hat{n}}$, $D_{i,j,\hat{n}}$, $E_{i,j,\hat{n}}$, $G_{i,j,\hat{n}}$, and $H_{i,j,\hat{n}}$ are functions of M . As evident from (5.41), the sum rate between each pair is a complex function of α , Λ , and Ψ . Furthermore, the second degree terms of Λ components appear on the sum rate equation. Due to these reasons, solving optimizing problems involving (5.41) appear intractable.

To overcome this problem and to formulate solvable optimization problems, the following three steps are taken: (1) the above sum rates are replaced by their asymptotic values, (2) the asymptotic rates are simplified by assuming perfect SIC, and (3) the decoding order functions $f_k(n)$ are determined according to the large-scale fading coefficients. The simplified asymptotic sum rate is obtained as follows:

$$\mathcal{R}_{k,f_k(n)}^\infty = \frac{1}{2} \frac{(T_C - \tau)}{T_C} \log \left(1 + \frac{\lambda_{k,f_k(n)}^2 M_{k,n}}{\sum_{m=n+1}^{K-1} \lambda_{k,f_k(m)}^2 M_{k,m} + \sigma_k^2} \right), \quad (5.43)$$

where $M_{k,n} = E \alpha_{f_k(n)} \hat{\beta}_k^2 \hat{\beta}_{f_k(n)}^2$. This is obtained by removing the terms corresponding to the imperfect SIC in (5.34) and replacing Ψ with 1. The decoding order is defined as follows:

$$f_k(n) = \begin{cases} n & n < k \\ n+1 & n \geq k, \end{cases} \quad (5.44)$$

where the users are ordered according to the descending order of large-scale fading coefficients between them and the relay (i.e., $\beta_1 \geq \beta_2 \geq \dots \geq \beta_K$). Based on the above simplifications, the max-min fairness power allocation optimization problem is presented as follows:

$$\text{Maximize}_{\lambda} \quad \min \left(\frac{(T_C - \tau)}{2T_C} \log \left(1 + \frac{\lambda_{k,f_k(n)}^2 M_{k,f_k(n)}}{\sum_{m=n+1}^{K-1} \lambda_{k,f_k(m)}^2 M_{k,f_k(m)} + \sigma_k^2} \right) \right) \quad (5.45)$$

$$\text{subject to} \quad E \sum_{i=1}^K \sum_{j=1}^K \alpha_j \lambda_{i,j}^2 \hat{\beta}_i \hat{\beta}_j^2 \leq E_R. \quad (5.46)$$

The constraint (5.46) is the power constraint at the relay. It can be proven that the maximum value for (5.45) can be obtained when the data rates between all the users

are equal to each other and when the inequality (5.46) becomes an equality (i.e., when the relay uses the maximum available power for the transmission). Assuming that the signal-to-interference-plus-noise ratio (SINR) between each user reaches a common SINR \bar{t} , the optimal values for $\lambda_{k,f_k(n)}$'s are obtained as follows:

$$\lambda_{k,f_k(n)}^2 = \frac{\bar{t} \left(\sum_{m=n+1}^{K-1} \lambda_{k,f_k(m)}^2 M_{k,f_k(m)} + \sigma_k^2 \right)}{M_{k,f_k(n)}}, \quad n = K-1, \dots, 1. \quad (5.47)$$

The values for $\lambda_{k,f_k(n)}$ can be found by solving (5.47) starting from $n = K-1$ to $n = 1$ for each k value. By observing the pattern, a generalized expression for $\lambda_{k,f_k(n)}^2$ can be written as

$$\lambda_{k,f_k(n)}^2 = \frac{\bar{t} \sigma_k^2 (\bar{t} + 1)^{K-n}}{M_{k,f_k(n)}} = \frac{\bar{t} \sigma_k^2 (\bar{t} + 1)^{K-n}}{E \alpha_{f_k(n)} \hat{\beta}_k^2 \hat{\beta}_{f_k(n)}^2}. \quad (5.48)$$

Then by using those obtained values on (5.46), the following is obtained

$$\sum_{i=1}^K \frac{\sigma_i^2}{\hat{\beta}_i} \sum_{j=1}^{K-1} \bar{t} (\bar{t} + 1)^{K-j-1} = E_R. \quad (5.49)$$

This can be simplified as follows and the value for \bar{t} can be written as

$$\sum_{j=1}^{K-1} \bar{t} (\bar{t} + 1)^{K-j-1} = (\bar{t} + 1)^{K-1} - 1 = \frac{E_R}{\sum_{i=1}^K \frac{\sigma_i^2}{\hat{\beta}_i}}. \quad (5.50)$$

The non-negative real solution for the above polynomial can be derived as

$$\bar{t} = \sqrt[K-1]{1 + \frac{E_R}{\sum_{i=1}^K \frac{\sigma_i^2}{\hat{\beta}_i}}} - 1. \quad (5.51)$$

It can be seen that the asymptotically achievable sum rate under max-min fairness does not depend on the values of transmit power at the user nodes (i.e., P and α). Furthermore, the simulations show that it is independent of the decoding order at the user (i.e., $f_k(n)$).

5.8 Numerical results

This section investigates the performance gains of the proposed NOMA scheme via simulations. Apart from the power allocation obtained in Section 5.6, the following sub-optimal power allocation matrix is used for comparisons.

$$\mathbf{\Lambda} = \mathbf{D}^{-1} (\mathbf{B} + \mathbf{B}^T), \quad (5.52)$$

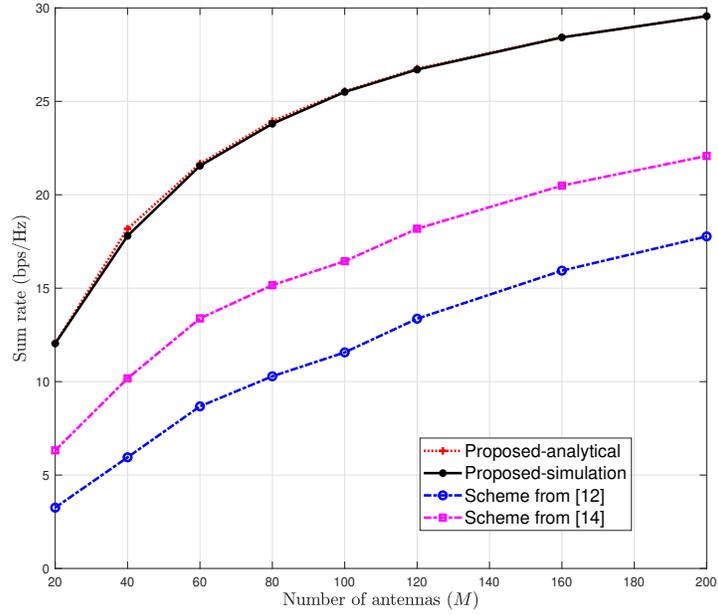


Figure 5.1: Total sum rate for $K = 8$ against the number of relay antennas.

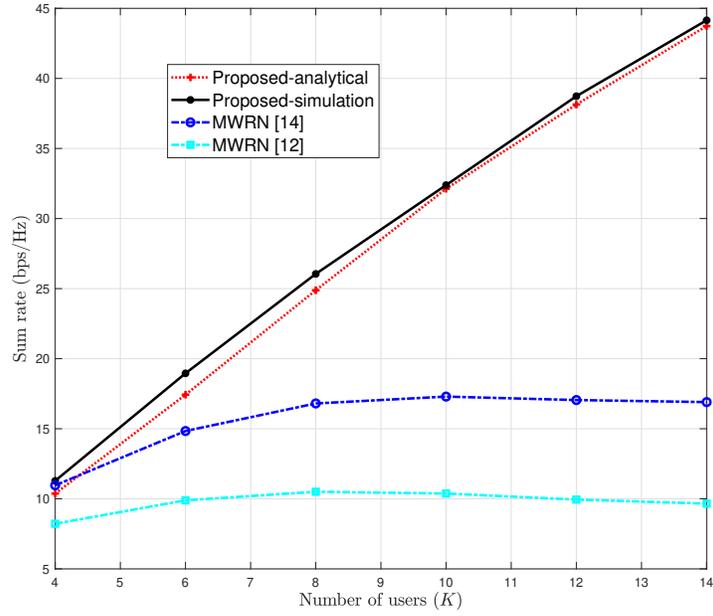


Figure 5.2: Average sum rate versus K for $M = 64$.

where \mathbf{B} is the $K \times K$ matrix with $[\mathbf{B}]_{i,j} = \sqrt{\frac{j-1}{K}}$ when $i < j$ and zero otherwise. This power allocation is based on the following observations.

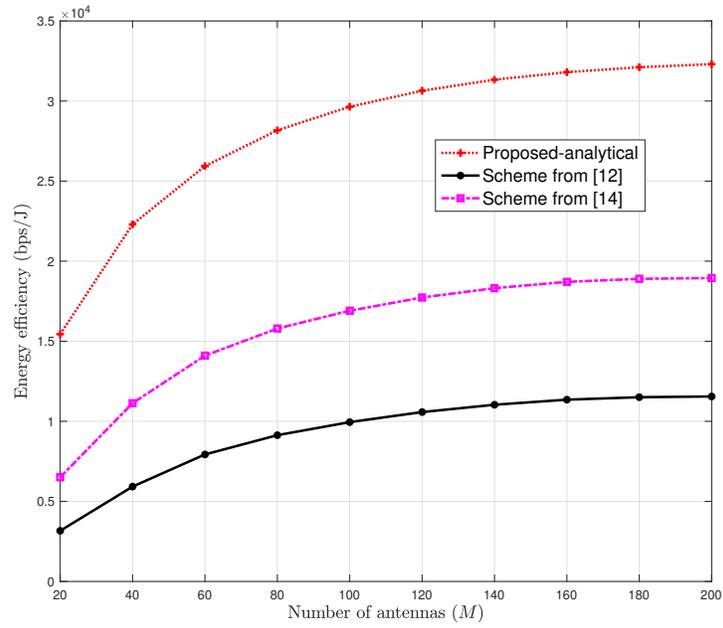


Figure 5.3: Average energy efficiency of the system versus M for $K = 12$.

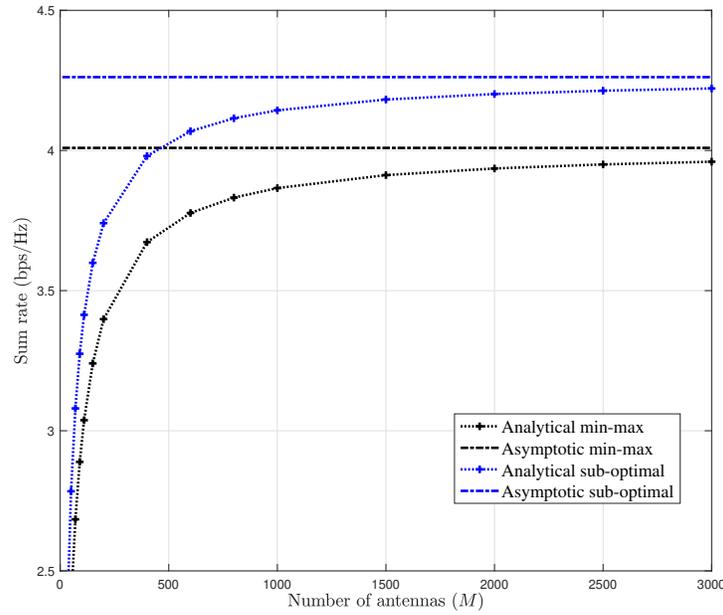


Figure 5.4: Average sum rate under different power allocation schemes.

- Each row in $\mathbf{\Lambda}$ corresponds to the power allocation factors to each user. Thus, to compensate for the downlink path-losses, more power is allocated to the

users which have the highest path-loss by multiplying each row of $\mathbf{\Lambda}$ by the inverse of the path-loss component of each user.

- The ratios between the non-zero coefficients in a single row determines the data rate between the users. Here, the ratios are designed in the form of $\sqrt{1/K}, \sqrt{2/K}, \dots, \sqrt{K-1/K}$.

Fig. 5.1 plots the achievable sum rate of the proposed MWRN system and those of [40] and [100] against the number of relay antennas. The power allocation matrix (52) with eight users ($K = 8$) and path-loss components $\mathbf{D} = \text{diag}(1, 0.875, 0.75, 0.625, 0.5, 0.375, 0.25, 0.125)$, and $\alpha_k = 1$ for $1 \leq k \leq K$ is considered. The proposed scheme clearly provides a higher achievable sum rate compared to the other two. For instance, with 100 relay antennas, it provides a sum rate of 25.2 bps/Hz while [40] and [100] achieve only 12.2 bps/Hz and 17.1 bps/Hz. This amounts to 106% and 47% gains respectively. Furthermore, it can be observed that more relay antennas increase the sum rate of the proposed system.

Fig. 5.2 plots the achievable sum rates for the proposed system and the two other MWRN schemes against the number of users (K). While the achievable sum rate reaches a constant value under [40] and [100], the proposed method provides constant sum rate improvements as the number of users are increased. For instance, both the proposed scheme and [100] has the same sum rate performance with 4 users. But with 10 users, the proposed scheme provides a sum rate of 33 bps/Hz, while [100] and [40] only provides 16 bps/Hz and 10 bps/Hz, respectively. Furthermore, the performance gap increases as the number of users are increased. As an example, with 8 users, the proposed system provides 47% and 127% increase of sum rate compared to [100] and [40] while this gain increases to 123% and 280% for 12 users.

Fig. 5.3 plots the energy efficiency (5.35) against the number of relay antennas M for $K = 12$ users. The values for $P_{U,C}$, $P_{R,C}$, P_{SYN} , P_{COD} , P_{DEC} , B , U , and L_{BS} are adopted from [125]. It can be seen that the proposed protocol provides higher energy efficiency compared to two other methods. As an example, with 100 relay antennas, the energy efficiency is 2.94×10^4 bps/J while the energy efficiency of [40] is only 1×10^4 bps/J. This is almost a 300% increase. This gain is both due to the sum rate increase as well as lower transmit powers due to the reduced number of time slots.

Fig. 5.4 plots the total sum rate of the system under the proposed two power

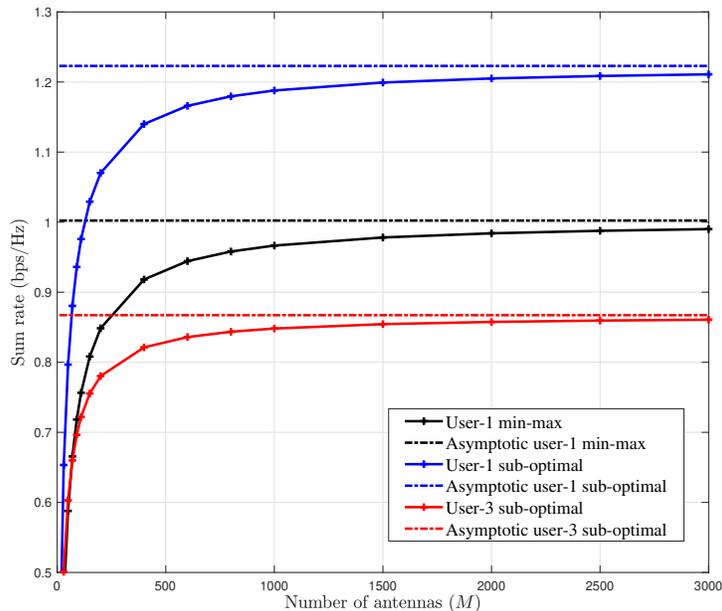


Figure 5.5: Average sum rate for individual users with different power allocation schemes.

allocation schemes namely the max-min power allocation and the sub-optimal power allocation for four users (i.e., $K = 4$ and $\mathbf{D} = \text{diag}(1, 0.75, 0.5, 0.25)$) with $\tau = 0.3$ and perfect SIC scenario. The sum rate from (5.18) for different M values and the asymptotic value from (5.34) are plotted. The figure shows that the sub-optimal power allocation results in slightly higher total sum rate than the min-max fairness power allocation. Also this shows that the asymptotic result (5.34) is accurate.

Fig. 5.5 plots the individual sum rate for user 1 and user 3 for the same system setup. It can be seen that the max-min power allocation scheme obtains 1 bps/Hz sum rate for all the users, while the sum rate obtained from the sub optimal power allocation scheme differs for different users. As an example, in this case user 1 (the user nearest to the relay) obtains 1.25 bps/Hz while the user 3 (user far away from the relay) only obtain 0.87 bps/Hz. Thus, this shows that the proposed power allocation can provide fairness to all users regardless of the distance between the relay and the users. This will be useful when all the users have to be treated equally.

Fig. 5.6 analyzes the effect of CSI availability on the performance of the system. Specifically, the achievable sum rate is plotted with 128 antennas and the asymptotic sum rate for three different K values (i.e., $K = 4$, $K = 8$, and $K = 12$) against the

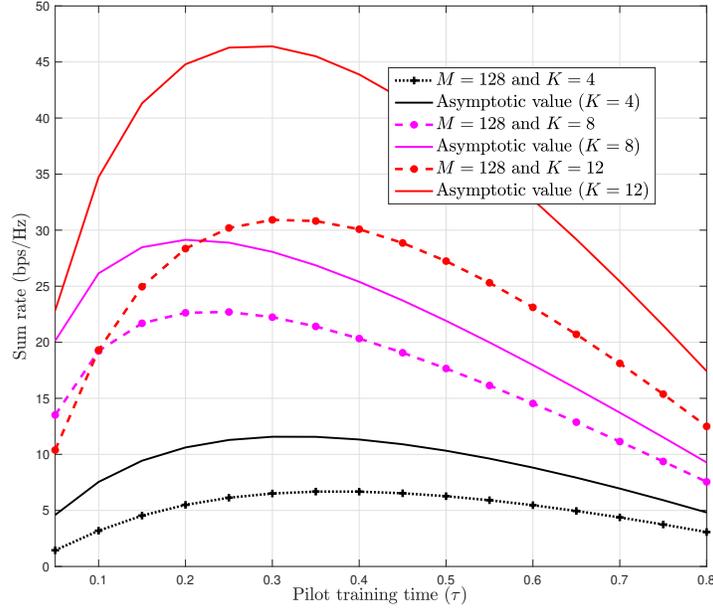


Figure 5.6: Average total sum rate for different CSI settings.

pilot training time (τ). Note that sum rates increase with τ up to a certain point and then starts to decrease. As an example, for $K = 8$, $\tau = 0.1$ provides 19 bps/Hz and increasing τ to 0.2 provides 22.3 bps/Hz. However after this, the achievable sum rate starts to decrease. A lower value for τ results in poorly estimated channel and lead to lower sum rates, while a higher value for τ will limit the time used for data transmission and also result in lower sum rate for the system. According to Fig. 5.6, the optimum τ values for $K = 4$, $K = 8$, and $K = 12$, are $\tau = 0.35$, $\tau = 0.2$, and $\tau = 0.3$ respectively for $M = 128$. This shows that the existence of optimum τ value based on the number of antennas, number of users, and other system parameters. However, this optimization of τ is not attempted and is left as a future research topic.

Next, in order to analyze the effect of imperfect SIC, the sum rate of user 1 against the number of relay antennas is plotted in Fig. 5.7. Here, it is assumed $K = 8$ and $\tau = 0.3$. The sum rate is plotted under four scenarios; namely perfect SIC, imperfect SIC with $\rho_{m,k} = 0.9$, $\rho_{m,k} = 0.7$, and $\rho_{m,k} = 0.5$ for all m, k values. As evident from this figure, when SIC is free from error propagation, the system sum rate increases significantly. As an example, with 500 antennas, user 1 obtains 1.4 bps/Hz with perfect SIC, but only 0.9 bps/Hz when $\rho = 0.9$. This shows the

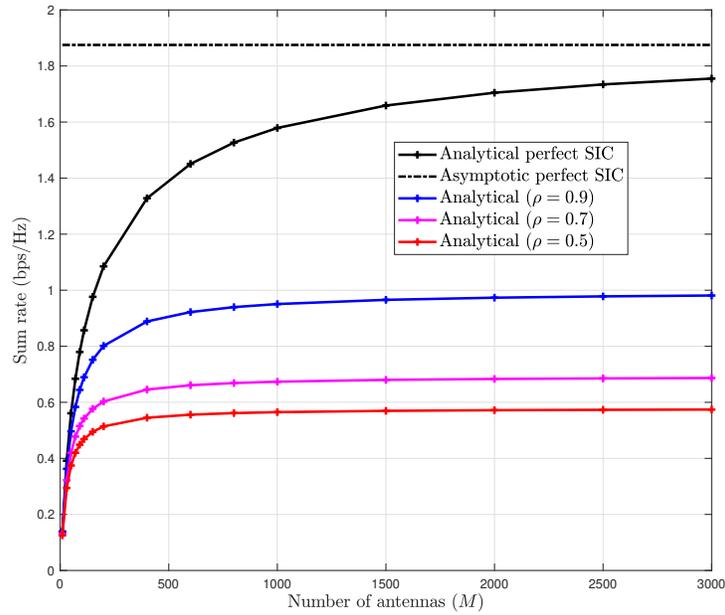


Figure 5.7: Sum rate of user-1 under different SIC conditions against the number of antennas.

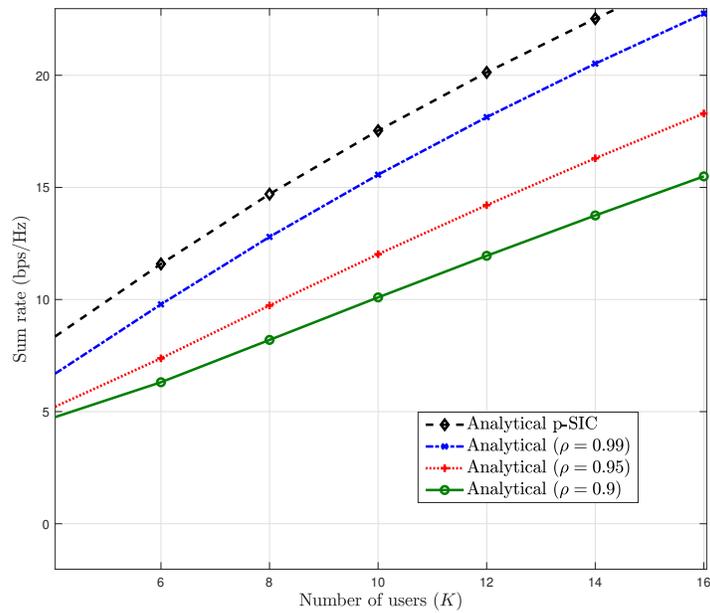


Figure 5.8: Average total sum rate for different SIC conditions against the number of users.

significant effect of SIC on NOMA systems.

To analyze this detrimental impact of imperfect SIC further, the total sum rate is plotted against the number of users K in Fig. 5.8. Here three different values for ρ are analysed along with the perfect SIC scenario. It can be seen that imperfect SIC reduces the achievable sum rate of a system. As an example with 6 users, the sum rate of the system will degrade to 7.5 bps/Hz from 11.9 bps/Hz when imperfect SIC is present with $\rho = 0.95$. This value further decreases to 6.8 bps/Hz when $\rho = 0.9$. This shows the importance of accurate SIC in NOMA systems.

5.9 Conclusion

A NOMA-aided mMIMO MWRN which enables data exchange between K users within only two time slots is proposed in this chapter. This is a drastic reduction compared to $\lceil (K + 1)/2 \rceil + 1$ time slots of the current state-of-the-art in MWRNs. In the first time slot, all user nodes transmit simultaneously to the relay, which in turn applies a linear MRC detector. In the second time slot, for each user, the relay constructs a superposition-coded signal consisting of data symbols belonging to all other users to be transmitted by using linear MRT precoding. Upon receiving this superposition-coded signal, users adopt SIC to decode the data from each user. The asymptotic sum rate is derived in closed-form. The proposed scheme provides a sum rate gain of $(1 - 4/K) \times 100\%$ over the current state-of-the-art MWRN. The gain is more significant when the number of users (K) is increased. Also, the use of two time slots enables the use of MWRNs in fast fading channels with small coherence times. Also, the proposed scheme provides significant energy efficiency gains due to the improved sum rate and the reduced number of time slots. Furthermore, a power allocation scheme, which improves user fairness in designed.

Chapter 6

Machine Learning for Multiple Relay Selection

This chapter proposes a multiple relay selection scheme based on machine learning. Specifically a cooperative wireless system where L amplify-and-forward (AF) relays enable the data transmission between a source and a destination node is analysed. Relays either support the transmission with their full power or do not transmit at all. Relay selection is modelled as a multi-class multi-label classification problem and a deep neural network (DNN) based solution is provided. Obtained results show that a DNN based linear method obtains 96% classification accuracy compared to the optimum relay selection method. Also, the proposed method achieves 99% of the sum rate performance of the optimum method, which has an exponential complexity with the number of relays. In comparison, our method has only a linear complexity with the number of relays and, through simulations it is shown that full diversity is achieved in 96% cases in our method.

6.1 Introduction

Multiple relay selection introduces several additional challenges compared to single relay selection schemes [42]. If the number of potential relays is L , there are $2^L - 1$ possible ways to select multiple relays. Thus, the relay selection complexity is increased due to the exponential number of possibilities. The work in [43] uses relay ordering proposed in [101] to design a linear complexity relay selection scheme. Furthermore, it proposes a quadratic complexity multiple relay selection scheme. Although the performance of the quadratic complexity relay selection achieves almost the same performance as the optimum relay selection, the performance of the

linear complexity relay selection method is lower. Thus, designing a multiple relay selection that achieves the performance of optimum method with a linear complexity with respect to the number of relays is a challenging problem.

Motivated by this, this chapter proposes a linear complexity relay selection based on machine learning. Multiple relay selection is formulated as a multi-class multi-label classification problem and DNNs are used to solve the problem.

The use of machine learning for wireless communication have been studied in [136–139]. Specifically, in [136], k-nearest neighbour and support vector machines are used for antenna selection in wireless systems by modelling it as multi-class classification problem. [137] uses DNNs to solve antenna selection and the multi-cast beamforming. Relay selection in the presence of an eavesdropper is solved using neural networks in [138]. In [139], DNNs are used for transmit antenna selection in a untrusted relay network. More recently, machine learning has being used for multiple relay selection in [140]. However, [140] looks at selecting a fixed number of relays and uses independent machine learning agents to decide the selected relays and their beamforming coefficients. In contrast, this work considers all the relays as a interdependent system and uses DNN to obtain the solution. Furthermore, [141] uses multi-armed bandit based relay selection for a power line communications system.

6.2 System, channel, and signal model

The system model consists of one source node (S), one destination node (D), and L relay nodes (R_l for $l \in \{1, 2, \dots, L\}$). All the nodes are single antenna nodes. The channel between S and R_l is represented by $h_{S,l}$ and the channel between R_l and D is given by $h_{l,D}$. The system model is shown in Fig. 6.1. For simulations, independent and identically distributed (i.i.d.) Rayleigh fading channels are assumed. The direct channel between S and D is assumed to be not available due to large pathloss and heavy shadowing effects [32].

Thus, the data transmission between S and D is supported by one or multiple relays. Each relay will either cooperate with it's full power or does not cooperate at all. The transmit power at S is P while the transmit power of R_l is P_l . The data transmission process takes two time slots. In the first time slot S transmits its data x to relays. And in the next time slot the selected relays will transmit the amplified version of the received signal to D . Binary variable a_l indicates whether relay R_l cooperates or not for the data transmission process.

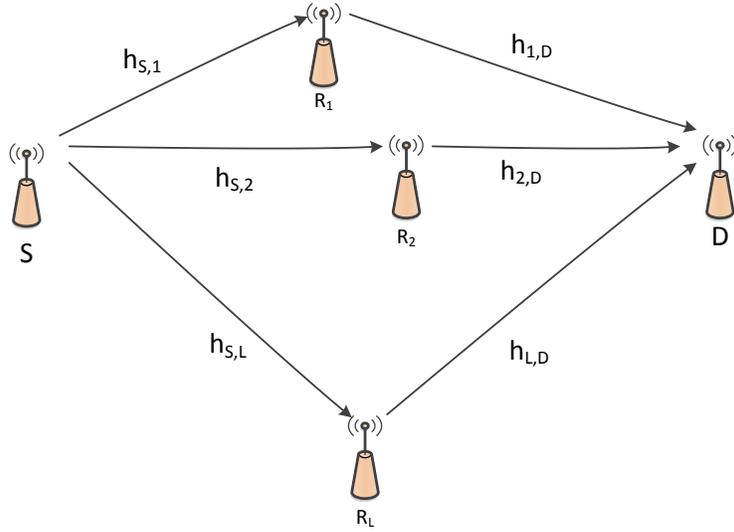


Figure 6.1: System model with L relays

The received signal at D at the end of data transmission can be written as

$$y = \sqrt{P} \sum_{l=1}^L \frac{a_l |h_{S,l} h_{l,D}| \sqrt{P_l}}{\sqrt{1 + |h_{S,l}|^2 P}} x + \sum_{l=1}^L \frac{a_l |h_{l,D}| \sqrt{P_l}}{\sqrt{1 + |h_{S,l}|^2 P}} n_l + n, \quad (6.1)$$

where n_l is the noise at R_l , and n is the noise at the receiver. Here, all the noises are assumed to be i.i.d. Gaussian random variables with zero mean and unit variances. In (6.1), the first term is the required signal, the second term is the amplified noise transmitted from relays, and the last term is the noise at the receiver itself. Based on this the received signal-to-noise ratio (SNR) can be written as [43]

$$\gamma = \frac{P \left(\sum_{l=1}^L \frac{a_l |h_{S,l} h_{l,D}| \sqrt{P_l}}{\sqrt{1 + |h_{S,l}|^2 P}} \right)^2}{1 + \sum_{l=1}^L \frac{a_l |h_{l,D}|^2 P_l}{1 + |h_{S,l}|^2 P}}. \quad (6.2)$$

Based on (4.23), the instantaneous sum rate of the system is defined as follows.

$$\mathcal{R} = \frac{1}{2} \log(1 + \gamma) \quad (6.3)$$

Furthermore, the average block error rate of the system is defined as

$$P_B = \mathbb{E} \left[\mathcal{Q}(\sqrt{k\gamma}) \right], \quad (6.4)$$

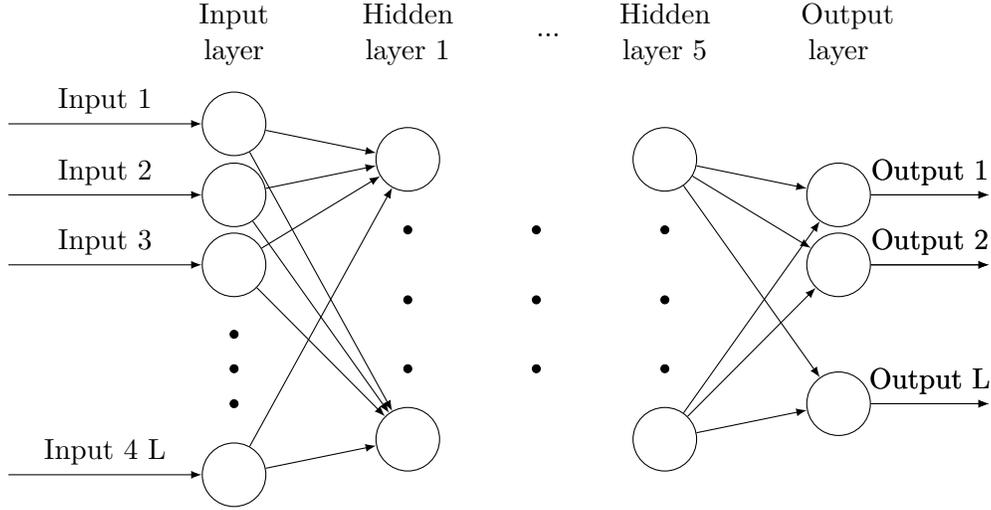


Figure 6.2: Deep Neural Network structure.

where k is a constant based on the modulation method and $\mathcal{Q}(x)$ is the Q function.

6.3 DNN solution

The optimal relay selection problem, under the given conditions (i.e., relays either use full power or no power), can be presented as follows.

$$\underset{a_1, \dots, a_L}{\text{maximize}} \quad \gamma \quad a_l \in \{0, 1\}. \quad (6.5)$$

The result of this maximization problem produces not only the highest SNR, but also lowest error rate and highest sum rate. However, this problem is a general non-linear 0 – 1 programming problem which is NP-hard. Furthermore, for L relays, there are $2^L - 1$ choices and thus the brute-force search is exponentially complex in the number of relays.

6.3.1 DNN structure

Input layer

The input layer consists of $4L$ nodes. These corresponds to the channel values of L relays. For R_l , $(\text{Re}(h_{S,l}), \text{Im}(h_{S,l}), \text{Re}(h_{l,D}), \text{Im}(h_{l,D}))$ are used as the inputs to the DNN. The training inputs are normalized and fed in to the DNN.

Hidden layers

Five dense hidden layers are used in the proposed DNN. The number of nodes in each layer is a hyper parameter that will be optimized during the training period. Furthermore, to mitigate the over-fitting in the system, random dropout of nodes in each layer is used. The dropout rate is another hyper parameter that will be optimized during the training of the system.

Output layer

The output layer consists of L nodes which outputs a binary value. The nodes are activated using the Sigmoid function. The l th output node indicates whether relay R_l cooperates or not.

6.3.2 Dataset

For simulations 100000 mmWave channel realizations each at P values of ranging from 0dB to 16dB with 2dB increments are generated. The relay powers (P_l s) are assumed to be proportional to the source power (i.e. $P_l = b_l P$ where b_l s are constants). For these channel realizations, the optimal relay selection solution is calculated by using exponentially complex brute force approach to be used for training and to check the accuracy of the proposed solution.

6.3.3 Training and validation

From the data set, 80000 channel realizations (and their corresponding outputs) at 6dB are selected and used to train the network for 300 epochs by using the Adam optimizer [142]. After the training, the model is validated on the remaining 20000 data. The results on this validation step is used to find the optimal hyper parameters for the system. Through this, it is shown that the optimum values for the number of nodes is 256 and the dropout rate is 10%.

6.3.4 Obtaining final results

Next, by using the model obtained at 6dB, the results for other channel realizations under different power values are obtained. Due to the use of normalization before feeding the data to the input layer, similar accuracy results can be obtained for the system. The classification accuracy (the correct number of predicted relays compared to the optimum relay selection criteria) and the sum rate accuracy (the

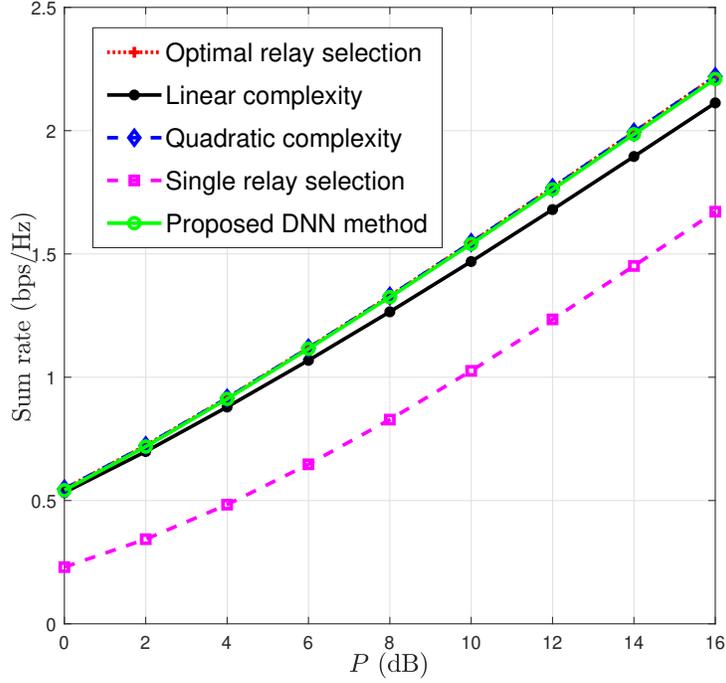


Figure 6.3: Sum rate versus SNR

percentage of the sum rate achieved by the DNN method compared to the optimum relay selection method) is measured for the system.

6.4 Numerical results

This section presents the numerical results for the proposed method. For comparison, following four methods are implemented. 1. Optimum multiple relay selection. 2. Linear complexity multiple relay selection based on SNR relay ordering [43]. 3. Quadratic complexity multiple relay selection [43]. 4. Single relay selection based on the best SNR. The numerical results with $L = 5$ and $L = 10$ are obtained. The powers of the relays are fixed as $P_1 = P_7 = 2P$, $P_2 = P_9 = 3/2P$, $P_3 = P_6 = P_8 = 5/4P$, $P_4 = P_{10} = 1/2P$, and $P_5 = 3/4P$.

In Fig. 6.3, the average sum rate, the average of (6.3), against P for different relay selection strategies is plotted. It can be seen that the proposed DNN-based relay selection achieves almost the same sum rate as the optimum relay selection method. Furthermore, it outperforms single relay selection and linear complexity relay selection. As an example, at 10dB, the proposed method obtains 1.6bps/Hz

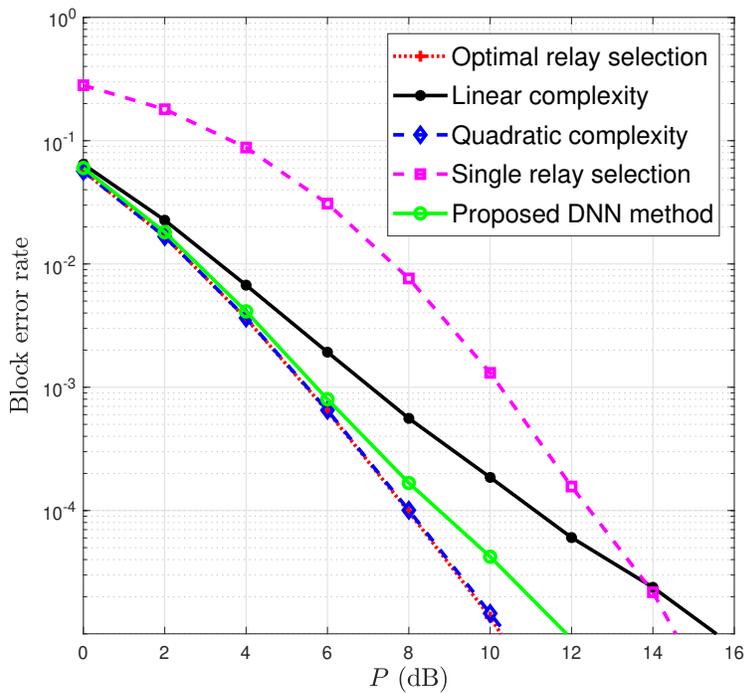


Figure 6.4: Block error rate versus SNR

while single relay selection obtains only 1bps/Hz and linear complexity algorithm obtains 1.5 bps/Hz. This is a gain of 60% and 6% respectively.

In Fig. 6.4, the average block error rate (6.4) under the considered relay selection schemes assuming BPSK is plotted. It can be seen that proposed method outperforms single relay selection and the linear complexity relay selection method. However, the performance of the proposed DNN method is only slightly lower than the optimum method and quadratic complexity method in [43]. In Fig. 6.5, accuracy of the proposed method is presented. It can be seen that from sum rate perspective, the DNN method achieves over 99% accuracy compared to the optimal method. However, the classification accuracy is lower than this value. The highest classification accuracy is obtained at 6 dB power value. This is expected as the initial training was done at $P = 6$ dB value. The lowest accuracy of 93.4% occurs at 0dB and the accuracy increases until 6 dB. Then it will go down to 95% when $P = 16$ dB.

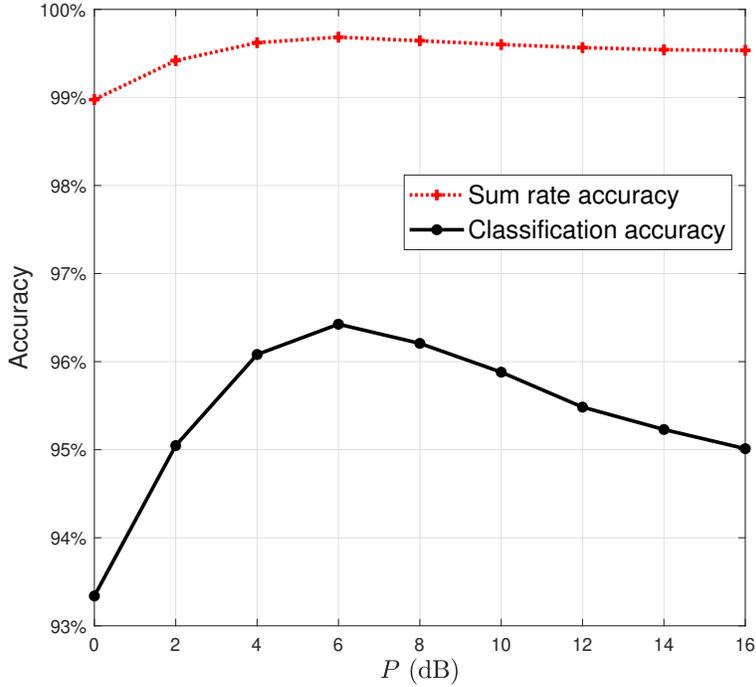


Figure 6.5: Accuracy of the system

6.4.1 Diversity order

The diversity order of a system is defined as the ratio between the logarithm value of the error rate and the logarithm value of the power at asymptotically large power domain [42]. It has been proven that, the multi relay selection methods proposed in [43] achieves diversity order of L . However, as with the use of DNN, obtaining performance guarantees via a closed form proof is impossible. However, the block error rate versus P value is plotted in Fig. 6.6 for different L values. It can be seen that as the number of relays in the system is increased, the slope of the block error rate reduces. Also, from Fig. 6.5 it is clear that the obtained relay selection is same as the optimum selection in 96% of the cases. Thus, it can be deduced that the diversity order of the proposed method should be L in at-least 96% of the time. Also, by comparing the slope of block error rate with other relay selection schemes in Fig. 6.4, it can be seen that the proposed method has the same slope as the optimum relay selection scheme.

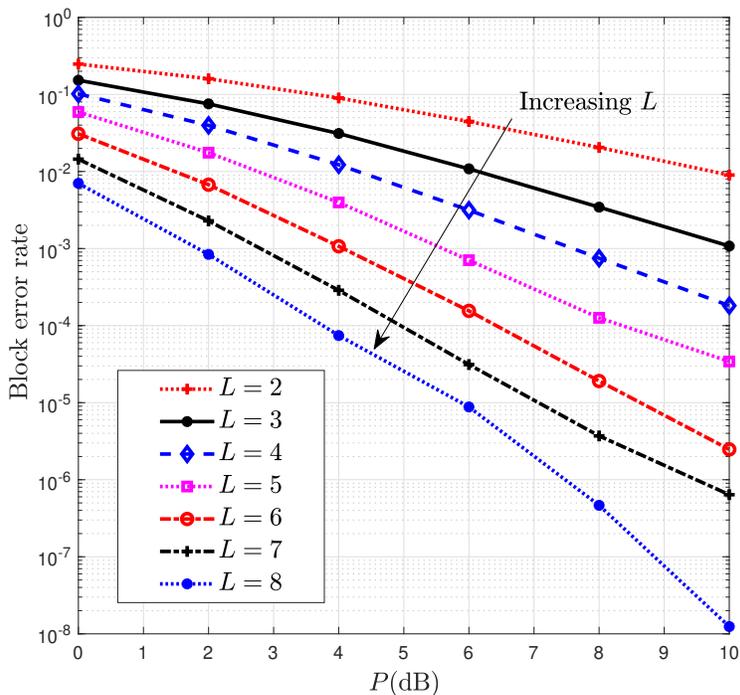


Figure 6.6: Block error rate of the system versus P

6.4.2 Complexity analysis

As mentioned in the introduction, computational complexity and feedback complexity are two important metrics for multi-relay selection algorithms. As no relay ordering is used, the feedback complexity of the proposed method is L . The predictions obtained through the DNN system are based on the calculation of L output nodes. The total number of calculations required for predicting the results is $4LN_1 + N_1N_2 + \dots + N_6L$, where N_i is the number of nodes in i th hidden layer. Thus the complexity of the proposed method is linear relative to L . Feedback and computational complexities along with the achievable diversity orders for the considered relay selection methods are highlighted in the following table.

6.5 Conclusion

This chapter proposes a DNN-based multi-relay selection approach for amplify and forward relays. Specifically, the multi relay selection problem is modelled as a multi-class multi-label classification problem. The resulting method has a linear complex-

¹For at-least 95% of the time a diversity order of L can be obtained in proposed DNN method.

Table 6.1: Complexity of the compared methods

Method	Feedback	Complexity	Diversity
Proposed DNN method	L	L	L^1
Single relay selection	1	L	L
Linear relay selection [43]	1	L	L
Quadratic relay selection [43]	L	L^2	L
Optimum relay selection	L	$2^L - 1$	L

ity in the number of relays and through numerical simulations it is shown that full diversity can be obtained for 96% of the time. Also, the sum rate and block error rate performance of the proposed method is better than single relay selection schemes. Classification accuracy of up to 96% and sum rate accuracy over 99% are obtained compared with optimum relay selection method which has an exponential complexity with the number of relays.

Chapter 7

Conclusion and Future Research Directions

7.1 Summary of contributions and conclusions

Cooperative communication expands the coverage region and reduces the energy consumption of wireless networks. It has been identified as a key component for Fifth Generation (5G) networks to accommodate exponentially increasing data rate and energy efficiency requirements [143]. Early work and surveys on the use of cooperative communications in 5G wireless systems have shown encouraging results [30,143]. Furthermore, the use of relay selection has shown promising results for 5G systems [144].

This thesis analyses the use of two-way relay networks (TWRNs), multi-way relay networks (MWRNs), and relay selection in different wireless settings. In summary, use of TWRNs, MWRNs, and relay selection resulted in improved sum rates, improved spectral efficiencies, and improved energy efficiencies. In summary, contributions of each chapter is listed as follows:

- In Chapter 3, a multi-cell two-way relay network (TWRN) consisting of single-antenna user nodes and amplify-and-forward (AF) relay nodes having very large antenna arrays is analysed. The combined impact of co-channel interference (CCI), imperfect channel-state information (CSI), pilot contamination, and the antenna correlation at the massive multiple-input multiple-output (mMIMO) node is analysed. By using a large number of antennas at the relay, it is shown that the effect of CCI can be mitigated. However, the effects of imperfect CSI and pilot contamination degrade the performance even with a large antenna array. Nevertheless, the use of mMIMO allows power scaling

at the user nodes and relay and thus, even with channel imperfections, the benefits of employing a mMIMO enabled relay on transmit power savings are significant. Furthermore, closed-form approximations for the sum rate were derived. This will help to decide on the required number of relay antennas to obtain a certain percentage of the asymptotic sum rate. Also, analysis on antenna correlation shows that it can be mitigated by using a large antenna array. The optimal pilot sequence length that maximizes the sum rate of the system was also derived.

- In Chapter 4, TWRNs in an underlay cognitive radio (CR) setting is analysed. Specifically, optimal selection of an AF two-way relay is considered to maximize the sum rate and to keep the interference on the primary user (PU) below an interference threshold. Asymptotic signal-to-interference-plus-noise ratio (SINR) values for two scenarios are obtained: (1) the relays and the two end nodes use transmit power scaling and (2) only the end nodes use transmit power scaling. For these two cases, optimal power allocations subject to the PU interference constraints are derived. With these optimal power allocations, the effects of relay selection on the outage, the sum rate, and the energy efficiency of the network are analysed. For the first scenario, the outage can be reduced to zero with appropriate power allocation and the relay selection can be done offline. For the second scenario, the outage will depend on the instantaneous channel state between the relays and the PU.
- Chapter 5 has introduced a MWRN transmission scheme which completes the data exchange among users in just two time slots. The so-called worst-case Gaussian approximation was used to derive the overall sum rate. The relay power allocation matrix is designed to maximize the fairness among all the users. Furthermore, the effects of imperfect successive interference cancellation (SIC) and imperfect CSI on the overall sum rate were analyzed.
- In Chapter 6, the use of DNNs for multiple relay selection in wireless has been proposed. Classification accuracy of up to 96% and sum rate accuracy of up to 99% are achieved. More importantly, the proposed method obtains the full diversity (at-least 96% time) and have a linear complexity with the number of relays in the system. In contrast, the optimum relay selection has an exponential complexity with the number of relays.

7.2 Future research directions

- In Chapter 3, multi-pair mMIMO TWRNs are analyzed where each cell has a single relay with a massive antenna array. It will be an interesting to study the effect of using cell-free mMIMO in such a system. Also, an analysis on relay selection for mMIMO cell free TWRNs will be an interesting future work.
- The analysis of CR relay selection in TWRNs in Chapter 4 is limited to underlay CR setting. Analysis of TWRNs and relay selection under overlay and interweave CR settings will be an useful future research direction.
- The power allocation scheme proposed in Chapter 5 is based on fairness among users. It will be interesting to see the ways to design the power allocation matrix to achieve different data rates for different users based on their requirements.
- In Chapter 6, the analysis is limited to the use of deep neural networks (DNNs). However, there are other machine learning techniques that have shown significant performance in multi-class multi-label classification problems. Utilization of such machine learning methods, including reinforcement learning, will be a future research direction. Furthermore, obtaining performance guarantees for DNN-based multiple relay selection scheme will be an interesting problem.
- The DNN approach used in Chapter 6 can be applied to other wireless communication problems that can be modelled as multi-class multi-label classification problems. Codebook-based hybrid beamforming is one such problem.

Bibliography

- [1] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Academic Press, 2007.
- [2] “The mobile economy 2018,” GSM Association, Tech. Rep., 2018. [Online]. Available: <https://www.gsma.com/mobileeconomy/wp-content/uploads/2018/05/The-Mobile-Economy-2018.pdf>
- [3] (2019) Cisco visual networking index: Global mobile data traffic forecast update, 2017–2022 white paper. Cisco Systems, Inc. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html>
- [4] D. S. Thomas, *The Role of Geographic Information Science & Technology in Disaster Management*. Springer International Publishing, 2018, pp. 311–330. [Online]. Available: https://doi.org/10.1007/978-3-319-63254-4_16
- [5] “IMT vision-framework and overall objectives of the future development of IMT for 2020 and beyond,” *Recommendation ITU*, pp. 2083–0, 2015.
- [6] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
- [7] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, “What will 5G be?” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [8] S.-Y. Lien, S.-L. Shieh, Y. Huang, B. Su, Y.-L. Hsu, and H.-Y. Wei, “5G new radio: Waveform, frame structure, multiple access, and initial access,” *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 64–71, Jun. 2017.

- [9] J. Lee, E. Tejedor, K. Ranta-aho, H. Wang, K.-T. Lee, E. Semaan, E. Molyeldin, J. Song, C. Bergljung, and S. Jung, "Spectrum for 5G: Global status, challenges, and enabling technologies," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 12–18, Mar. 2018.
- [10] 3GPP. Release 15. [Online]. Available: <http://www.3gpp.org/release-15>
- [11] ——. Release 16. [Online]. Available: <http://www.3gpp.org/release-16>
- [12] C. Bockelmann, N. K. Pratas, G. Wunder, S. Saur, M. Navarro, D. Gregoratti, G. Vivier, E. D. Carvalho, Y. Ji, C. Stefanovic, P. Popovski, Q. Wang, M. Schellmann, E. Kosmatos, P. Demestichas, M. Raceala-Motoc, P. Jung, S. Stanczak, and A. Dekorsy, "Towards massive connectivity support for scalable mMTC communications in 5G networks," *IEEE Access*, vol. 6, pp. 28 969–28 992, May 2018.
- [13] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 124–130, Jun. 2018.
- [14] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 2015.
- [15] V. Gazis, "A survey of standards for machine-to-machine and the Internet of Things," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 482–511, 2017.
- [16] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, "A survey on 5G networks for the Internet of Things: Communication technologies and challenges," *IEEE Access*, vol. 6, pp. 3619–3647, 2018.
- [17] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [18] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

- [19] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
- [20] "Quantifying the benefits of 64T64R massive MIMO with beamforming and multi-user MIMO capabilities," Signals Research Group, LLC, Tech. Rep., 2018. [Online]. Available: https://www.researchandmarkets.com/research/gg2bxp/quantifying_the?w=5
- [21] (2018) Ericsson and MTS to deliver superior mobile broadband experiences for football fans at tournament in Russia. [Online]. Available: <https://www.ericsson.com/en/press-releases/2018/6/ericsson-and-mts-to-deliver-superior-mobile-broadband-experiences-for>
- [22] (2019) Huawei launches 5G simplified solution. Huawei.com. [Online]. Available: <https://www.huawei.com/en/press-events/news/2019/2/huawei-5g-simplified-solution>
- [23] M. Sami, N. K. Noordin, M. Khabazian, F. Hashim, and S. Subramaniam, "A survey and taxonomy on medium access control strategies for cooperative communication in wireless networks: Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2493–2521, Fourthquarter 2016.
- [24] W. Zhuang and M. Ismail, "Cooperation in wireless communication networks," *IEEE Wireless Commun.*, vol. 19, no. 2, pp. 10–20, Apr. 2012.
- [25] Z. Sheng, K. K. Leung, and Z. Ding, "Cooperative wireless networks: from radio to network protocol designs," *IEEE Commun. Mag.*, vol. 49, no. 5, pp. 64–69, May 2011.
- [26] A. Nosratinia, T. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.
- [27] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. part i. system description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

- [28] V. Genc, S. Murphy, Y. Yu, and J. Murphy, “IEEE 802.16J relay-based wireless access networks: an overview,” *IEEE Wireless Commun.*, vol. 15, no. 5, pp. 56–63, Oct. 2008.
- [29] K. Loa, C.-C. Wu, S.-T. Sheu, Y. Yuan, M. Chion, D. Huo, and L. Xu, “IMT-advanced relay standards,” *IEEE Commun. Mag.*, vol. 48, no. 8, pp. 40–48, Aug. 2010.
- [30] R. Zhu, Y. E. Wang, Q. Xu, Y. Liu, and Y. D. Li, “Millimeter-wave to microwave MIMO relays (M4R) for 5G building penetration communications,” in *2018 IEEE Radio and Wireless Symposium (RWS)*, Jan. 2018, pp. 206–208.
- [31] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, “Cooperative diversity in wireless networks: Efficient protocols and outage behavior,” *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
- [32] B. Rankov and A. Wittneben, “Spectral efficient protocols for half-duplex fading relay channels,” *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 379–389, Feb. 2007.
- [33] Y. Han, S. H. Ting, C. K. Ho, and W. H. Chin, “Performance bounds for two-way amplify-and-forward relaying,” *IEEE Trans. Wireless Commun.*, vol. 8, pp. 432–439, Jan. 2009.
- [34] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*, 1st ed. Cambridge, UK ; New York, NY : Cambridge University Press, 2003.
- [35] G. Amarasuriya, C. Tellambura, and M. Ardakani, “Performance analysis of zero-forcing for two-way MIMO AF relay networks,” *IEEE Wireless Commun. Lett.*, vol. 1, no. 2, pp. 53–56, Apr. 2012.
- [36] L. Ong, S. J. Johnson, and C. M. Kellett, “The capacity region of multiway relay channels over finite fields with full data exchange,” *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3016–3031, May 2011.
- [37] D. Gunduz, A. Yener, A. Goldsmith, and H. V. Poor, “The multiway relay channel,” *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 51–63, Jan. 2013.

- [38] G. Amarasuriya, C. Tellambura, and M. Ardakani, “Sum rate analysis of two-way MIMO AF relay networks with zero-forcing,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4456–4469, Sep. 2013.
- [39] M. N. Hasan and K. Anwar, “Joint decoding for multiway multirelay networks with coded random access,” in *2016 22nd Asia-Pacific Conference on Communications (APCC)*, Aug. 2016, pp. 96–102.
- [40] D. P. Kudathanthirige and G. A. A. Baduge, “Multicell multiway massive MIMO relay networks,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 6831–6848, Aug. 2017.
- [41] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, “A simple cooperative diversity method based on network path selection,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 659–672, Mar. 2006.
- [42] S. Atapattu, Y. Jing, H. Jiang, and C. Tellambura, “Relay selection schemes and performance analysis approximations for two-way networks,” *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 987–998, Mar. 2013.
- [43] Y. Jing and H. Jafarkhani, “Single and multiple relay selection schemes and their achievable diversity orders,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1414–1423, Mar. 2009.
- [44] V. Sreng, H. Yanikomeroglu, and D. Falconer, “Relayer selection strategies in cellular networks with peer-to-peer relaying,” in *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 3, Oct. 2003, pp. 1949–1953 Vol.3.
- [45] S. Haykin, “Cognitive radio: brain-empowered wireless communications,” *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
- [46] (2018) Auction of residual spectrum licences in the 700 MHz, 2500 MHz, 2300 MHz and PCS-G bands. [Online]. Available: http://www.ic.gc.ca/eic/site/smt-gst.nsf/eng/h_sf11312.html
- [47] M. H. Islam, C. L. Koh, S. W. Oh, X. Qing, Y. Y. Lai, C. Wang, Y. Liang, B. E. Toh, F. Chin, G. L. Tan, and W. Toh, “Spectrum survey in Singapore:

- Occupancy measurements and analyses,” in *2008 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom 2008)*, 2008, pp. 1–7.
- [48] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, “Breaking spectrum gridlock with cognitive radios: An information theoretic perspective,” *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.
- [49] “IEEE standard for information technology - telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements - part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 5: Television white spaces (TVWS) operation,” *IEEE Std 802.11af-2013 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012, IEEE Std 802.11aa-2012, IEEE Std 802.11ad-2012, and IEEE Std 802.11ac-2013)*, pp. 1–198, Feb. 2014.
- [50] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, “Non-orthogonal multiple access (NOMA) for cellular future radio access,” in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, Jun. 2013, pp. 1–5.
- [51] H. V. Cheng, E. Björnson, and E. G. Larsson, “Performance analysis of NOMA in training-based multiuser MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 372–385, Jan. 2018.
- [52] D. Kudathanthirige and G. A. A. Baduge, “NOMA-aided multicell downlink massive MIMO,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 3, pp. 612–627, Jun. 2019.
- [53] Z. Ding and H. V. Poor, “Design of Massive-MIMO-NOMA With Limited Feedback,” *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 629–633, May 2016.
- [54] Z. Ding, L. Dai, R. Schober, and H. V. Poor, “NOMA Meets Finite Resolution Analog Beamforming in Massive MIMO and Millimeter-Wave Networks,” *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1879–1882, Aug. 2017.

- [55] X. Chen, Z. Zhang, C. Zhong, R. Jia, and D. W. K. Ng, “Fully Non-Orthogonal Communication for Massive Access,” *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1717–1731, Apr. 2018.
- [56] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, “A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.
- [57] L. Dai, B. Wang, Y. Yuan, S. Han, C. I, and Z. Wang, “Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends,” *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.
- [58] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. S. Kwak, “Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges,” *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, Secondquarter 2017.
- [59] J. Choi, “On the Power Allocation for a Practical Multiuser Superposition Scheme in NOMA Systems,” *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 438–441, Mar. 2016.
- [60] L. Song, Y. Li, Z. Ding, and H. V. Poor, “Resource Management in Non-Orthogonal Multiple Access Networks for 5G and Beyond,” *IEEE Network*, vol. 31, no. 4, pp. 8–14, Jul. 2017.
- [61] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, “On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users,” *IEEE Signal Processing Letters*, vol. 21, no. 12, pp. 1501–1505, 12 2014.
- [62] S. Timotheou and I. Krikidis, “Fairness for Non-Orthogonal Multiple Access in 5G Systems,” *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.
- [63] Y. Liu, M. ElKashlan, Z. Ding, and G. K. Karagiannidis, “Fairness of User Clustering in MIMO Non-Orthogonal Multiple Access Systems,” *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1465–1468, Jul. 2016.

- [64] S. Silva, G. Amarasuriya, M. Ardakani, and C. Tellambura, “NOMA-Aided Multi-Way massive MIMO relay networks,” in *Submitted to 2019 IEEE International Conference on Communications (ICC)*, Shanghai, P.R. China, May 2019.
- [65] H. Cui, M. Ma, L. Song, and B. Jiao, “Relay selection for two-way full duplex relay networks with amplify-and-forward protocol,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3768–3777, Jul. 2014.
- [66] M. Chen and A. Yener, “Multiuser two-way relaying: detection and interference management strategies,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4296–4305, Aug. 2009.
- [67] M. Iwamura, H. Takahashi, and S. Nagata, “Relay technology in LTE-advanced,” *NTT DOCOMO Techn. J.*, vol. 18, no. 2, pp. 31–36, Jul. 2010.
- [68] C. Perera, C. Liu, S. Jayawardena, and M. Chen, “A survey on internet of things from industrial market perspective,” *IEEE Access*, vol. 2, pp. 1660–1679, 2014.
- [69] H. Suraweera, H. Q. Ngo, T. Duong, C. Yuen, and E. Larsson, “Multi-pair amplify-and-forward relaying with very large antenna arrays,” in *Proc. IEEE Int. Conf. Commun. (ICC), 2013*, Jun. 2013, pp. 4635–4640.
- [70] G. Amarasuriya and H. V. Poor, “Multi-user relay networks with massive MIMO,” in *Proc. IEEE Int. Conf. on Commun. (ICC)*, London, UK, Jun. 2015, pp. 2017–2023.
- [71] G. Amarasuriya, “Sum rate analysis for multi-user massive MIMO relay networks,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, Dec. 2015, pp. 1–7.
- [72] G. Amarasuriya, C. Tellambura, and M. Ardakani, “Performance analysis of zero-forcing for two-way MIMO AF relay networks,” *IEEE Wireless Commun. Lett.*, vol. 1, no. 2, pp. 53–56, Apr. 2012.
- [73] D. P. Kudathanthirige and G. Amarasuriya Aruma Baduge, “Multicell multiway massive MIMO relay networks,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 6831–6848, Aug. 2017.

- [74] G. Amarasuriya and H. V. Poor, “Multi-way amplify-and-forward relay networks with massive MIMO,” in *Proc. IEEE Int. Symp. Personal, Indoor, and Mobile Radio Commun. (PIMRC)*, Washington, DC, USA, Sep. 2014, pp. 1–6.
- [75] S. Jin, X. Liang, K. K. Wong, X. Gao, and Q. Zhu, “Ergodic rate analysis for multipair massive MIMO two-way relay networks,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1480–1491, Mar. 2015.
- [76] C. Kong, C. Zhong, M. Matthaiou, E. Bjornson, and Z. Zhang, “Multi-pair two-way AF relaying systems with massive arrays and imperfect CSI,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 3651–3655.
- [77] J. Yang, H. Wang, J. Ding, B. Li, and C. Li, “Spectral and energy efficiency for massive MIMO multi-pair two-way relay networks with ZFR/ZFT and imperfect CSI,” in *2015 21st Asia-Pacific Conference on Communications (APCC)*, Oct. 2015, pp. 47–51.
- [78] Z. Zhang, Z. Chen, M. Shen, and B. Xia, “Spectral and energy efficiency of multipair two-way full-duplex relay systems with massive MIMO,” *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 848–863, Apr. 2016.
- [79] S. Silva, M. Ardakani, and C. Tellambura, “Relay selection for cognitive massive MIMO two-way relay networks,” in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2017, pp. 1–6.
- [80] X. Ge, J. Yang, H. Gharavi, and Y. Sun, “Energy efficiency challenges of 5G small cell networks,” *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 184–191, May 2017.
- [81] K. M. S. Huq, S. Mumtaz, J. Bachmatiuk, J. Rodriguez, X. Wang, and R. L. Aguiar, “Green hetnet CoMP: Energy efficiency analysis and optimization,” *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4670–4683, Oct. 2015.
- [82] S. Atapattu, Y. Jing, H. Jiang, and C. Tellambura, “Relay selection and performance analysis in multiple-user networks,” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 8, pp. 1517–1529, Aug. 2013.

- [83] G. Amarasuriya, M. Ardakani, and C. Tellambura, "Output-threshold multiple-relay-selection scheme for cooperative wireless networks," *IEEE Trans. Veh. Commun.*, vol. 59, no. 6, pp. 3091–3097, Jul. 2010.
- [84] S. Masrour, A. H. Bastami, and P. Halimi, "Spectrum sharing in cognitive radio networks using beamforming and two-path successive relaying," in *2017 Iranian Conference on Electrical Engineering (ICEE)*, May 2017, pp. 1810–1814.
- [85] G. Wang, Y. Zou, J. Lu, and C. Tellambura, "Cognitive transmission and performance analysis for amplify-and-forward two-way relay networks," in *2014 IEEE International Conference on Communications (ICC)*, Jun. 2014, pp. 5843–5848.
- [86] M. Shamailzadeh and A. H. Bastami, "Cooperative beamforming in cognitive radio network with network-coded bidirectional filter-and-forward relaying," in *2016 24th Iranian Conference on Electrical Engineering (ICEE)*, May 2016, pp. 1454–1459.
- [87] A. Advaita, M. M. Gali, T. M. C. Chu, and H. J. Zepernick, "Outage probability of MIMO cognitive cooperative radio networks with multiple AF relays using orthogonal space-time block codes," in *2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct. 2017, pp. 84–89.
- [88] J. Zou, W. Liu, M. Ding, H. Luo, and H. Yu, "Transceiver design for AF MIMO two-way relay systems with imperfect channel estimation," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, Dec. 2011, pp. 1–5.
- [89] P. Ubaidulla and S. Aissa, "Optimal relay selection and power allocation for cognitive two-way relaying networks," *IEEE Wireless Commun. Lett.*, vol. 1, no. 3, pp. 225–228, Jun. 2012.
- [90] S. Kusaladharma and C. Tellambura, "Massive MIMO based underlay networks with power control," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

- [91] N. Wang, E. Hossain, and V. Bhargava, “Downlink cell association for large-scale MIMO hetnets employing small cell wireless backhaul,” in *Electrical and Computer Engineering (CCECE), 2015 IEEE 28th Canadian Conference on*, May 2015, pp. 1042–1047.
- [92] H. Al-Hraishawi and G. Amarasuriya, “Sum rate analysis of cognitive massive MIMO systems with underlay spectrum sharing,” in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2016, pp. 1–7.
- [93] C. Yin, T. X. Doan, N. P. Nguyen, T. Mai, and L. D. Nguyen, “Outage probability of full-duplex cognitive relay networks with partial relay selection,” in *2017 International Conference on Recent Advances in Signal Processing, Telecommunications Computing (SigTelCom)*, Jan. 2017, pp. 115–118.
- [94] D. Wang, R. Zhang, X. Cheng, L. Yang, and C. Chen, “Relay selection in full-duplex energy-harvesting two-way relay networks,” *IEEE Transactions on Green Communications and Networking*, vol. 1, no. 2, pp. 182–191, Jun. 2017.
- [95] T. Ding, X. Yuan, and S. C. Liew, “Algorithmic beamforming design for MIMO multiway relay channel with clustered full data exchange,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 10 081–10 086, Oct. 2018.
- [96] S. Rahimian, W. Zhang, M. Noori, Y. Jing, and M. Ardakani, “Partial zero-forcing for multi-way relay networks,” *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4444–4456, Oct. 2018.
- [97] F. L. Duarte and R. C. de Lamare, “Buffer-aided max-link relay selection for multi-way cooperative multi-antenna systems,” *IEEE Commun. Lett.*, vol. 23, no. 8, pp. 1423–1426, Aug. 2019.
- [98] H. Wang and Q. Chen, “LDPC based network coded cooperation design for multi-way relay networks,” *IEEE Access*, vol. 7, pp. 62 300–62 311, May 2019.
- [99] C. D. Ho, H. Q. Ngo, M. Matthaiou, and T. Q. Duong, “On the performance of zero-forcing processing in multi-way massive MIMO relay networks,” *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 849–852, Apr. 2017.
- [100] C. D. Ho, H. Q. Ngo, M. Matthaiou, and L. D. Nguyen, “Power allocation for multi-way massive MIMO relaying,” *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4457–4472, Oct. 2018.

- [101] Y. Jing and H. Jafarkhani, “Network beamforming using relays with perfect channel information,” *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2499–2517, Jun. 2009.
- [102] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, “A comprehensive survey of pilot contamination in massive MIMO-5G system,” *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 905–923, Secondquarter 2016.
- [103] T. Shafique, A. M. Abdelhady, O. Amin, and M. S. Alouini, “Energy efficiency, spectral efficiency and delay analysis for selective ARQ multi-channel systems,” *IEEE Trans. Green Commun. and Netw.*, vol. 2, no. 3, pp. 612–622, Sep. 2018.
- [104] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, 1st ed. Prentice Hall, 1993.
- [105] J. Hoydis, S. ten Brink, and M. Debbah, “Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [106] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, “Cell-free massive MIMO versus small cells,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.
- [107] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, “Cell-free massive MIMO: A new next-generation paradigm,” *IEEE Access*, vol. 7, pp. 99 878–99 888, Jul. 2019.
- [108] Y.-C. Liang and R. Zhang, “Optimal analogue relaying with multi-antennas for physical layer network coding,” in *Communications, 2008. ICC '08. IEEE International Conference on*, May 2008, pp. 3893–3897.
- [109] S. Silva, G. Amarasuriya, C. Tellambura, and M. Ardakani, “Relay selection strategies for MIMO two-way relay networks with spatial multiplexing,” *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4694–4710, Dec. 2015.
- [110] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth, “On the Lambert-W function,” *Advances in Computational mathematics*, vol. 5, no. 1, pp. 329–359, 1996.

- [111] J. A. Tague and C. I. Caldwell, “Expectations of useful complex Wishart forms,” *Multidimensional Syst. Signal Process.*, vol. 5, no. 3, pp. 263–279, Jul. 1994. [Online]. Available: <http://dx.doi.org/10.1007/BF00980709>
- [112] D. Gore, J. Heath, R.W., and A. Paulraj, “On performance of the zero forcing receiver in presence of transmit correlation,” in *IEEE Intl. Symp. on Inf. Theory (ISIT)*, Lausanne, Switzerland, Jul. 2002, pp. 1–2.
- [113] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, “A coordinated approach to channel estimation in large-scale multiple-antenna systems,” *IEEE J. Sel. Areas Commun.*, vol. 2, no. 31, pp. 264—273, Feb. 2013.
- [114] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. Cambridge University Press, 2011.
- [115] H. Bolcskei, M. Borgmann, and A. J. Paulraj, “Impact of the propagation environment on the performance of space-frequency coded MIMO-OFDM,” *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 427–439, Apr. 2003.
- [116] A. A. Khan, M. H. Rehmani, and A. Rachedi, “Cognitive-radio-based internet of things: Applications, architectures, spectrum related functionalities, and future research directions,” *IEEE Wireless Communications*, vol. 24, no. 3, pp. 17–25, Jun. 2017.
- [117] T. T. Duy and H. Y. Kong, “Exact outage probability of cognitive two-way relaying scheme with opportunistic relay selection under interference constraint,” *IET Communications*, vol. 6, no. 16, pp. 2750–2759, Nov. 2012.
- [118] K. B. Fredj and S. Aissa, “Performance of spectrum-sharing constrained two-way relaying,” in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2014, pp. 845–850.
- [119] R. W. Heath, S. Sandhu, and A. Paulraj, “Antenna selection for spatial multiplexing systems with linear receivers,” *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 142–144, Apr. 2001.
- [120] G. Wang, Q. Liu, R. He, F. Gao, and C. Tellambura, “Acquisition of channel state information in heterogeneous cloud radio access networks: challenges

- and research directions,” *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 100–107, June 2015.
- [121] G. Wang, F. Gao, Y.-C. Wu, and C. Tellambura, “Joint CFO and channel estimation for OFDM-based two-way relay networks,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 456–465, 2010.
- [122] G. Wang, F. Gao, W. Chen, and C. Tellambura, “Channel estimation and training design for two-way relay networks in time-selective fading environments,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2681–2691, 2011.
- [123] S. Silva, G. Amarasuriya, C. Tellambura, and M. Ardakani, “Massive MIMO two-way relay networks with channel imperfections,” in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–7.
- [124] T. Pham-Gia, D. N. Thanh, and D. T. Phong, “Trace of the Wishart matrix and applications,” *Open Journal of Statistics*, vol. 5, no. 03, p. 173, 2015.
- [125] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, “Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?” *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3059–3075, Jun. 2015.
- [126] J. Zhu, J. Wang, Y. Huang, S. He, X. You, and L. Yang, “On optimal power allocation for downlink non-orthogonal multiple access systems,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2744–2757, Dec. 2017.
- [127] X. Chen, R. Jia, and D. W. K. Ng, “On the design of massive non-orthogonal multiple access with imperfect successive interference cancellation,” *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2539–2551, Mar. 2019.
- [128] Y. Li and G. A. Aruma Baduge, “NOMA-aided cell-free massive MIMO systems,” *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 950–953, Dec. 2018.
- [129] Y. Li and G. Amarasuriya, “NOMA-aided massive MIMO downlink with distributed antenna arrays,” in *Proc. IEEE International Conference on Communications (ICC)*, May 2019, pp. 1–7.
- [130] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

- [131] (2020) Cisco annual internet report (2018–2023) white paper. Cisco Systems, Inc. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>
- [132] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993, vol. 1.
- [133] A. Mezghani and A. L. Swindlehurst, “Blind estimation of sparse broadband massive MIMO channels with ideal and one-bit ADCs,” *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2972–2983, Jun. 2018.
- [134] K. Mawatwal, D. Sen, and R. Roy, “A semi-blind channel estimation algorithm for massive MIMO systems,” *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 70–73, Feb. 2017.
- [135] H. Q. Ngo, E. Larsson, and T. Marzetta, “Energy and spectral efficiency of very large multiuser MIMO systems,” *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [136] J. Joung, “Machine learning-based antenna selection in wireless communications,” *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2241–2244, Nov. 2016.
- [137] M. S. Ibrahim, A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, “Learning-based antenna selection for multicasting,” in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2018, pp. 1–5.
- [138] Z. Deng, Q. Sang, Y. Gao, and C. Cai, “Optimal relay selection for wireless relay channel with external eavesdropper: a NN-based approach,” in *2018 IEEE/CIC International Conference on Communications in China (ICCC)*, Aug. 2018, pp. 515–519.
- [139] R. Yao, Y. Zhang, S. Wang, N. Qi, N. I. Miridakis, and T. A. Tsiftsis, “Deep neural network assisted approach for antenna selection in untrusted relay networks,” *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1644–1647, Dec. 2019.

- [140] A. Gouisseem, L. Samara, R. Hamila, N. Al-Dhahir, L. Ben-Brahim, and A. Gastli, "Machine-learning based relay selection in AF cooperative networks," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2019, pp. 1–7.
- [141] B. Nikfar and A. J. Han Vinck, "Relay selection in cooperative power line communication: A multi-armed bandit approach," *Journal of Commun. and Netw.*, vol. 19, no. 1, pp. 1–9, Feb. 2017.
- [142] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [143] R. Elouafadi and M. Benjillali, "Cooperative NOMA-based D2D communications: A survey in the 5G/IoT context," in *2018 19th IEEE Mediterranean Electrotechnical Conference (MELECON)*, May 2018, pp. 132–137.
- [144] K. Shamganth and M. J. N. Sibley, "A survey on relay selection in cooperative device-to-device (D2D) communication for 5G cellular networks," in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, Aug. 2017, pp. 42–46.
- [145] H. Cramér, *Random variables and probability distributions*. Cambridge University Press, 1970.

Appendix A

Appendices for Chapter 3

A.1 Proof of limits

In this section, several important limit results are provided. Following three results [145] are used to obtain these limits. For two independent vectors, $\mathbf{p} \sim \mathcal{CN}_{N \times 1}(0, \sigma_p^2)$ and $\mathbf{q} \sim \mathcal{CN}_{N \times 1}(0, \sigma_q^2)$, the following identities are valid.

$$\mathbf{p}^H \mathbf{p} / N \xrightarrow[N \rightarrow \infty]{a.s.} \sigma_p^2 \quad \text{and} \quad \mathbf{p}^H \mathbf{q} / N \xrightarrow[N \rightarrow \infty]{a.s.} 0, \quad (\text{A.1})$$

$$\mathbf{p}^H \mathbf{q} / \sqrt{N} \xrightarrow[N \rightarrow \infty]{d} \mathcal{CN}(0, \sigma_p^2 \sigma_q^2), \quad (\text{A.2})$$

where subscripts *a.s.* and *d* stands for almost sure convergence and the convergence of distributions, respectively. By using the aforementioned identities, it can be shown that

$$\frac{\mathbf{G}_{jl}^H \mathbf{G}_{jl}}{N} = \mathbf{D}_{jl}^{\frac{1}{2}} \left(\frac{\mathbf{F}_{jl}^H \mathbf{F}_{jl}}{N} \right) \mathbf{D}_{jl}^{\frac{1}{2}} \xrightarrow[N \rightarrow \infty]{a.s.} \mathbf{D}_{jl}, \quad \text{and} \quad (\text{A.3})$$

$$\frac{\mathbf{G}_{jl}^H \mathbf{G}_{ml}}{N} = \mathbf{D}_{jl}^{\frac{1}{2}} \left(\frac{\mathbf{F}_{jl}^H \mathbf{F}_{ml}}{N} \right) \mathbf{D}_{ml}^{\frac{1}{2}} \xrightarrow[N \rightarrow \infty]{a.s.} \mathbf{0}_{2K}, \quad \text{for } j \neq m. \quad (\text{A.4})$$

Furthermore, by using the above two results, following limit can be obtained.

$$\begin{aligned} \frac{\hat{\mathbf{G}}_u^H \hat{\mathbf{G}}_u}{N} &= \frac{1}{N} \left[\left(\sum_{j=1}^L \mathbf{G}_{jl} + \frac{\mathbf{V}_l}{\sqrt{P_p}} \right) \tilde{\mathbf{D}}_u \right]^H \left(\sum_{j=1}^L \mathbf{G}_{jl} + \frac{\mathbf{V}_l}{\sqrt{P_p}} \right) \tilde{\mathbf{D}}_u \\ &= \tilde{\mathbf{D}}_u^H \left(\sum_{j=1}^L \sum_{m=1}^L \frac{\mathbf{G}_{jl}^H \mathbf{G}_{ml}}{N} + \frac{\mathbf{V}_l^H \sum_{j=1}^L \mathbf{G}_{jl}}{\sqrt{P_p} N} \right. \\ &\quad \left. + \frac{\sum_{j=1}^L \mathbf{G}_{jl}^H \mathbf{V}_l}{\sqrt{P_p} N} + \frac{\mathbf{V}_l^H \mathbf{V}_l}{N} \right) \tilde{\mathbf{D}}_u \\ &= \tilde{\mathbf{D}}_u^H \left(\sum_{j=1}^L \mathbf{D}_{jl} + \mathbf{I}_{2K} \right) \tilde{\mathbf{D}}_u \xrightarrow[N \rightarrow \infty]{a.s.} \mathbf{H}_u, \end{aligned} \quad (\text{A.5})$$

where \mathbf{H}_{ll} is a diagonal matrix with k th diagonal entry given by $\frac{P_P \left(\sum_{j=1}^L \eta_{j,l,k} \right)^2}{1 + P_P \sum_{j=1}^L \eta_{j,l,k}}$. Furthermore when transmit power scaling is done at the users (ie. $P_P = E_P / \sqrt{N}$), the following limit can be obtained.

$$\frac{\hat{\mathbf{G}}_{ll}^H \hat{\mathbf{G}}_{ll}}{\sqrt{N}} \xrightarrow[N \rightarrow \infty]{a.s.} \hat{\mathbf{H}}_{ll}, \quad (\text{A.6})$$

where $\hat{\mathbf{H}}_{ll}$ is a diagonal matrix with k th diagonal entry given by $E_P \left(\sum_{j=1}^L \eta_{j,l,k} \right)^2$. Furthermore, the following limit results are obtained using the same procedure.

$$\frac{\mathbf{G}_{jl}^H \hat{\mathbf{G}}_{ll}}{N} \xrightarrow[N \rightarrow \infty]{a.s.} E_P \text{diag} \left(\frac{\eta_{j,l,1} \sum_{j=1}^L \eta_{j,l,1}}{1 + E_P \sum_{j=1}^L \eta_{j,l,1}}, \dots, \frac{\eta_{j,l,2K} \sum_{j=1}^L \eta_{j,l,2K}}{1 + E_P \sum_{j=1}^L \eta_{j,l,2K}} \right), \quad P_P = E_P \quad (\text{A.7})$$

$$\frac{\mathbf{G}_{jl}^H \hat{\mathbf{G}}_{ll}}{\sqrt{N}} \xrightarrow[N \rightarrow \infty]{a.s.} E_P \text{diag} \left(\eta_{j,l,1} \sum_{j=1}^L \eta_{j,l,1}, \dots, \eta_{j,l,2K} \sum_{j=1}^L \eta_{j,l,2K} \right), \quad P_P = \frac{E_P}{\sqrt{N}} \quad (\text{A.8})$$

A.2 Proof of limits for power scaling at the user nodes

This section provides a sketch of the proof of signal-to-interference-plus-noise ratio (SINR) for the transmit power scaling scenario. First limits for the power normalizing factor $\hat{\beta}_l$ is obtained. The first term in the denominator in (3.10) can be written as

$$\begin{aligned} & \text{Tr} \left(\mathbf{G}_{jl} \mathbf{G}_{jl}^H \hat{\mathbf{G}}_{ll} \left[\hat{\mathbf{G}}_{ll}^H \hat{\mathbf{G}}_{ll} \right]^{-1} \mathbf{P} \left[\hat{\mathbf{G}}_{ll}^T \hat{\mathbf{G}}_{ll}^* \right]^{-1} \mathbf{P} \left[\hat{\mathbf{G}}_{ll}^H \hat{\mathbf{G}}_{ll} \right]^{-1} \hat{\mathbf{G}}_{ll}^H \right) \\ &= \frac{1}{\sqrt{N}} \text{Tr} \left(\frac{\mathbf{G}_{jl}^H \hat{\mathbf{G}}_{ll}}{\sqrt{N}} \left[\frac{\hat{\mathbf{G}}_{ll}^H \hat{\mathbf{G}}_{ll}}{\sqrt{N}} \right]^{-1} \mathbf{P} \left[\frac{\hat{\mathbf{G}}_{ll}^T \hat{\mathbf{G}}_{ll}^*}{\sqrt{N}} \right]^{-1} \mathbf{P} \left[\frac{\hat{\mathbf{G}}_{ll}^H \hat{\mathbf{G}}_{ll}}{\sqrt{N}} \right]^{-1} \frac{\hat{\mathbf{G}}_{ll}^H \mathbf{G}_{jl}}{\sqrt{N}} \right) \end{aligned} \quad (\text{A.9})$$

By using the limit results (A.5) and (3.58) given in Appendix A.1 on each term in the above equation, following result is formulated as:

$$\begin{aligned} & \sqrt{N} \text{Tr} \left(\mathbf{G}_{jl} \mathbf{G}_{jl}^H \hat{\mathbf{G}}_{ll} \left[\hat{\mathbf{G}}_{ll}^H \hat{\mathbf{G}}_{ll} \right]^{-1} \mathbf{P} \left[\hat{\mathbf{G}}_{ll}^T \hat{\mathbf{G}}_{ll}^* \right]^{-1} \mathbf{P} \left[\hat{\mathbf{G}}_{ll}^H \hat{\mathbf{G}}_{ll} \right]^{-1} \hat{\mathbf{G}}_{ll}^H \right) \\ & \xrightarrow[N \rightarrow \infty]{a.s.} \frac{1}{E_P} \sum_{i=1}^K \left(\frac{\eta_{j,l,2i-1}^2}{\hat{\eta}_{l,2i-1}^2 \hat{\eta}_{l,2i}} + \frac{\eta_{j,l,2i}^2}{\hat{\eta}_{l,2i-1} \hat{\eta}_{l,2i}^2} \right) \end{aligned} \quad (\text{A.10})$$

Here in (A.10), $\hat{\eta}_{l,k} = \sum_{j=1}^L \eta_{j,l,k}$. Similarly, the limit of the second term in the denominator in (3.10) is derived as

$$N \text{Tr} \left(\left[\hat{\mathbf{G}}_{ll}^H \hat{\mathbf{G}}_{ll} \right]^{-1} \mathbf{P} \left[\hat{\mathbf{G}}_{ll}^T \hat{\mathbf{G}}_{ll}^* \right]^{-1} \mathbf{P} \right) \xrightarrow[N \rightarrow \infty]{a.s.} 2 \sum_{i=1}^K \frac{1}{\left(E_P \hat{\eta}_{l,2i-1} \hat{\eta}_{l,2i} \right)^2}. \quad (\text{A.11})$$

By using the above two results, (3.13) is obtained.

A.3 Proof of limits for power scaling at the relay nodes

This section provides a sketch of the proof of SINR for the transmit power scaling at the relay. The first term in the denominator in (3.10) can be written in a similar way by replacing \sqrt{N} by N in (A.9) in Appendix A.2. By using the limit results (A.5) and (A.7) given in Appendix A.1,

$$\begin{aligned}
N \quad & \text{Tr} \left(\mathbf{G}_{jl} \mathbf{G}_{jl}^H \hat{\mathbf{G}}_u \left[\hat{\mathbf{G}}_u^H \hat{\mathbf{G}}_u \right]^{-1} \mathbf{P} \left[\hat{\mathbf{G}}_u^T \hat{\mathbf{G}}_u^* \right]^{-1} \mathbf{P} \left[\hat{\mathbf{G}}_u^H \hat{\mathbf{G}}_u \right]^{-1} \hat{\mathbf{G}}_u^H \right) \\
& \xrightarrow[N \rightarrow \infty]{a.s.} \sum_{i=1}^K \frac{(1 + P_P \hat{\eta}_{l,2i-1}) (1 + P_P \hat{\eta}_{l,2i-1})}{P_P^2 \hat{\eta}_{l,2i-1} \hat{\eta}_{l,2i}} \\
& \times \left[\frac{\eta_{j,l,2i-1} (1 + P_P \hat{\eta}_{l,2i-1})}{P_P \hat{\eta}_{l,2i-1}} + \frac{\eta_{j,l,2i} (1 + P_P \hat{\eta}_{l,2i})}{P_P \hat{\eta}_{l,2i}} \right]. \tag{A.12}
\end{aligned}$$

Similarly, the limit of the second term in the denominator in (3.10) is derived as

$$N^2 \text{Tr} \left(\left[\hat{\mathbf{G}}_u^H \hat{\mathbf{G}}_u \right]^{-1} \mathbf{P} \left[\hat{\mathbf{G}}_u^T \hat{\mathbf{G}}_u^* \right]^{-1} \mathbf{P} \right) \xrightarrow[N \rightarrow \infty]{a.s.} 2 \sum_{i=1}^K \left(\frac{1 + P_P \hat{\eta}_{l,2i-1}}{P_P \hat{\eta}_{l,2i-1}} \right)^2 \left(\frac{1 + P_P \hat{\eta}_{l,2i}}{P_P \hat{\eta}_{l,2i}} \right)^2 \tag{A.13}$$

By using the above two results, (3.23) is obtained.

Appendix B

Appendices for Chapter 4

B.1 Asymptotic signal-to-interference-plus-noise ratio (SINR) with power scaling at all nodes

This Appendix provides the sketch of the proof for SINR values (4.23) and (4.24).

Based on (A.1) and (A.2), the following limit results are obtained:

$$\frac{\mathbf{H}_{i,k}\mathbf{H}_{i,k}^H}{N_i} = \mathbf{D}_{i,k}^{\frac{1}{2}} \left(\frac{\tilde{\mathbf{H}}_{i,k}\tilde{\mathbf{H}}_{i,k}^H}{N_i} \right) \mathbf{D}_{i,k}^{\frac{1}{2}} \xrightarrow[N_i \rightarrow \infty]{a.s.} \mathbf{D}_{i,k}, \quad (\text{B.1})$$

$$\frac{\mathbf{G}_k\mathbf{G}_k^H}{N_i} = \mathbf{D}_k^{\frac{1}{2}} \left(\frac{\tilde{\mathbf{G}}_k\tilde{\mathbf{G}}_k^H}{N_i} \right) \mathbf{D}_k^{\frac{1}{2}} \xrightarrow[N_i \rightarrow \infty]{a.s.} \mathbf{0}_{N \times N}, \quad (\text{B.2})$$

$$\frac{\mathbf{F}_i^H\mathbf{F}_i}{N_i} = \hat{\mathbf{D}}_i^{\frac{1}{2}} \left(\frac{\tilde{\mathbf{F}}_i^H\tilde{\mathbf{F}}_i}{N_i} \right) \hat{\mathbf{D}}_i^{\frac{1}{2}} \xrightarrow[N_i \rightarrow \infty]{a.s.} \hat{\mathbf{D}}_i. \quad (\text{B.3})$$

The asymptotic limit of (B.1) shows that the left hand side converges to the diagonal matrix. Since the trace of the inverse of a diagonal matrix can be readily determined, the following limits are obtained:

$$\lim_{N_i \rightarrow \infty} N_i \text{Tr} \left(\left[\mathbf{H}_{i,k}\mathbf{H}_{i,k}^H \right]^{-1} \right) = \lim_{N_i \rightarrow \infty} \text{Tr} \left(\left[\frac{\mathbf{H}_{i,k}\mathbf{H}_{i,k}^H}{N_i} \right]^{-1} \right) = N_{R_k} \eta_{i,k}^{-1}. \quad (\text{B.4})$$

Furthermore, (B.2) and (B.3) immediately lead to the following limits:

$$\lim_{N_i \rightarrow \infty} \frac{\text{Tr} \left(\mathbf{G}_k^H\mathbf{G}_k \right)}{N_i} = 0 \quad \text{and} \quad \lim_{N_i \rightarrow \infty} \frac{\text{Tr} \left(\mathbf{F}_i^H\mathbf{F}_i \right)}{N_i} = \hat{\eta}_i N, \quad (\text{B.5})$$

Now the proofs of (4.23) and (4.24) can be outlined as follows. First, scaled transmit powers (4.25) are substituted on the SINR expression (4.18) to yield the following:

$$\gamma_{S_{1,k}^{(l)} | O_{S_k} = 0} = \frac{\frac{E_{R_k} E_{2,k}}{N_2 \text{Tr}([\mathbf{H}_{2,k} \mathbf{H}_{2,k}^H]^{-1})}}{E_{R_k} \sigma_{R_k}^2 + \left(\frac{E_{1,k}}{N_1 \text{Tr}([\mathbf{H}_{1,k} \mathbf{H}_{1,k}^H]^{-1})} + \frac{E_{2,k}}{N_2 \text{Tr}([\mathbf{H}_{2,k} \mathbf{H}_{2,k}^H]^{-1})} + \sigma_{R_k}^2 \right) \frac{\sigma_1^2}{\eta_{1,k}} \Psi}, \quad (\text{B.6})$$

where $\Psi = \left[N_1 \left(\tilde{\mathbf{H}}_{k,1}^H \tilde{\mathbf{H}}_{k,1} \right)^{-1} \right]_{l,l}$. Here, the limit of the matrix $\left(\tilde{\mathbf{H}}_{k,1}^H \tilde{\mathbf{H}}_{k,1} \right) / N_1$ as N_1 tends to infinity is the identity matrix of size $N_{R_k} \times N_{R_k}$, which follows from (A.1) and (A.2). By using this result and by replacing the trace terms in (B.6) by (B.4), (4.23) is obtained. With similar steps, (4.24) can be obtained. Also, by using the second limit of (B.5) on (4.22), the interference constraint (4.27) can be obtained.

Appendix C

Appendices for Chapter 5

C.1 Expected value results for Imperfect channel-state information (CSI)

C.1.1 Derivation of $M_{k,m}$

First, by using (C.14b), the term inside the expected value of $M_{k,m}$ is rewritten as

$$\mathbf{h}_k^T \hat{\mathbf{H}}^* \mathbf{\Lambda} \hat{\mathbf{H}}^H \mathbf{h}_{f_k(m)} = \sum_{i=1}^K \sum_{j=1}^K \lambda_{i,j} \mathbf{h}_k^T \hat{\mathbf{h}}_i^* \hat{\mathbf{h}}_j^H \mathbf{h}_{f_k(m)}. \quad (\text{C.1})$$

By substituting the value for \mathbf{h}_k , (C.1) is written as

$$M_{k,m} = \sum_{i=1}^K \sum_{j=1}^K \lambda_{i,j} \left(\hat{\mathbf{h}}_k^T + \mathbf{e}_k^T \right) \hat{\mathbf{h}}_i^* \hat{\mathbf{h}}_j^H \left(\hat{\mathbf{h}}_{f_k(m)} + \mathbf{e}_{f_k(m)} \right). \quad (\text{C.2})$$

As \mathbf{e}_k 's are independent from $\hat{\mathbf{h}}_k$'s, only the first term of the above summation can have a non-zero expectation. Next, by considering different i and j combinations (i.e., $i = k, j = f_k(m)$ and $j = k, i = f_k(m)$ and using the fact that $k \neq f_k(m)$) in the double summation in (C.1), the expected value can be written as

$$\mathbb{E} \left[\sum_{i=1}^K \sum_{j=1}^K \lambda_{i,j} \mathbf{h}_k^T \hat{\mathbf{h}}_i^* \hat{\mathbf{h}}_j^H \mathbf{h}_{f_k(m)} \right] = \lambda_{k,f_k(m)} \|\mathbf{h}_k\|^2 \|\mathbf{h}_{f_k(m)}\|^2 + \lambda_{f_k(m),k} |\mathbf{h}_k \mathbf{h}_{f_k(m)}|^2. \quad (\text{C.3})$$

Finally, by using the expected value results given in Appendix C.3, the value for $M_{k,m}$ is obtained as

$$M_{k,m} = M \left(\lambda_{k,f_k(m)} M + \lambda_{f_k(m),k} \right) \hat{\beta}_k \hat{\beta}_m, \quad (\text{C.4})$$

where $\hat{\beta}_k$ is given as

$$\hat{\beta}_k = \frac{P \alpha_{p,k} \beta_k^2}{(P \alpha_{p,k} \beta_k + 1)}. \quad (\text{C.5})$$

Here, $\hat{\beta}_k$ acts as a correction for β_k values due to the imperfect CSI. When perfect CSI is available, $\hat{\beta}_k$ will become β_k .

C.1.2 Derivation of $N_{k,m}$

In order to compute $N_{k,m}$, the value of $\bar{N}_{k,m} = \mathbb{E} \left[|\mathbf{h}_k^T \mathbf{H}^* \mathbf{\Lambda} \mathbf{H}^H \mathbf{h}_{f_k(m)}|^2 \right]$ is derived by using the same procedure as in $M_{k,m}$ and obtained as

$$\begin{aligned}
\bar{N}_{k,\hat{m}} &= M(M+1) \left(\lambda_{k,k}^2 \hat{\beta}_k^2 \left(\hat{\beta}_{\hat{m}} + \hat{\eta}_{\hat{m}} \right) \left((M+2)\hat{\beta}_k + \hat{\eta}_k \right) \right. \\
&\quad + \lambda_{\hat{m},\hat{m}}^2 \hat{\beta}_{\hat{m}}^2 \left(\hat{\beta}_k + \hat{\eta}_k \right) \left((M+2)\hat{\beta}_{\hat{m}} + \hat{\eta}_{\hat{m}} \right) \\
&\quad + M\hat{\beta}_{\hat{m}}\hat{\beta}_k \left((M+1)\hat{\beta}_{\hat{m}}\hat{\beta}_k \left(\lambda_{k,\hat{m}}^2 M(M+1) + 2\lambda_{\hat{m},k}^2 \right) \right. \\
&\quad + (M+1) \left(\hat{\eta}_{\hat{m}}\hat{\beta}_k + \hat{\beta}_{\hat{m}}\hat{\eta}_k \right) \left(\lambda_{k,\hat{m}}^2 M + \lambda_{\hat{m},k}^2 \right) \\
&\quad + M\hat{\eta}_{\hat{m}}\hat{\eta}_k \left(\lambda_{k,\hat{m}}^2 + \lambda_{\hat{m},k}^2 \right) \\
&\quad + 2\lambda_{k,\hat{m}}\lambda_{\hat{m},k}M\hat{\beta}_{\hat{m}}\hat{\beta}_k \left((M+1)\hat{\beta}_k + \hat{\eta}_k \right) \left((M+1)\hat{\beta}_{\hat{m}} + \hat{\eta}_{\hat{m}} \right) \\
&\quad + \sum_{\substack{i=1 \\ i \neq k, \hat{m}}}^K \sum_{\substack{j=1 \\ j \neq k, \hat{m}, i}}^K \lambda_{i,j}^2 M^2 \hat{\beta}_i \hat{\beta}_j \beta_k \beta_{\hat{m}} + \sum_{\substack{i=1 \\ i \neq k, \hat{m}}}^K \lambda_{i,i}^2 M(M+1) \hat{\beta}_i^2 \beta_k \beta_{\hat{m}} \\
&\quad + 2 \sum_{\substack{j=1 \\ j \neq k, \hat{m}}}^K \lambda_{k,j} \lambda_{j,k} M \hat{\beta}_k \hat{\beta}_j \left(\hat{\beta}_{\hat{m}} + \hat{\eta}_{\hat{m}} \right) \left((M+1)\hat{\beta}_k + \hat{\eta}_k \right) \\
&\quad + 2 \sum_{\substack{j=1 \\ j \neq k, \hat{m}}}^K \lambda_{\hat{m},j} \lambda_{j,\hat{m}} M \hat{\beta}_{\hat{m}} \hat{\beta}_j \left(\hat{\beta}_k + \hat{\eta}_k \right) \left((M+1)\hat{\beta}_{\hat{m}} + \hat{\eta}_{\hat{m}} \right) \\
&\quad + \sum_{\substack{j=1 \\ j \neq k, \hat{m}}}^K \lambda_{k,j}^2 M^2 \hat{\beta}_k \hat{\beta}_j \left(\hat{\beta}_{\hat{m}} + \hat{\eta}_{\hat{m}} \right) \left((M+1)\hat{\beta}_k + \hat{\eta}_k \right) \\
&\quad + \sum_{\substack{j=1 \\ j \neq k, \hat{m}}}^K \lambda_{\hat{m},j}^2 M \hat{\beta}_{\hat{m}} \hat{\beta}_j \left(\hat{\beta}_k + \hat{\eta}_k \right) \left((M+1)\hat{\beta}_{\hat{m}} + M\hat{\eta}_{\hat{m}} \right) \\
&\quad + \sum_{\substack{i=1 \\ i \neq k, \hat{m}}}^K \lambda_{i,k}^2 M \hat{\beta}_k \hat{\beta}_i \left(\hat{\beta}_{\hat{m}} + \hat{\eta}_{\hat{m}} \right) \left((M+1)\hat{\beta}_k + M\hat{\eta}_k \right) \\
&\quad + \sum_{\substack{i=1 \\ i \neq k, \hat{m}}}^K \lambda_{i,\hat{m}}^2 M^2 \hat{\beta}_{\hat{m}} \hat{\beta}_i \left(\hat{\beta}_k + \hat{\eta}_k \right) \left((M+1)\hat{\beta}_{\hat{m}} + \hat{\eta}_{\hat{m}} \right).
\end{aligned} \tag{C.6}$$

Here, $\hat{m} = f_k(m)$ and $\hat{\eta}_k$ is defined as

$$\hat{\eta}_k = \frac{\beta_k}{(P\alpha_{p,k}\beta_k + 1)}. \tag{C.7}$$

Here, $\hat{\eta}_k$ acts as an error term due to imperfect CSI. When perfect CSI is available, this term vanishes. The value of $N_{k,m}$ is then derived as

$$N_{k,m} = \bar{N}_{k,m} - M_{k,m}^2. \tag{C.8}$$

C.1.3 Derivation of Q_k

Similar to the previous cases, the value of Q_k is written as

$$\begin{aligned}
Q_k &= \lambda_{k,k}^2 M(M+1) \hat{\beta}_k^2 \left((M+2) \hat{\beta}_k + \hat{\eta}_k \right) + \sum_{i=1, i \neq k}^K \sum_{j=1, j \neq k}^K \lambda_{i,j} M \hat{\beta}_i \beta_k \hat{\beta}_j (\lambda_{j,i} + \lambda_{i,j} M) \\
&+ \sum_{i=1, i \neq k}^K M \hat{\beta}_k \hat{\beta}_i \left(2\lambda_{k,i} \lambda_{i,k} \left((M+1) \hat{\beta}_k + \hat{\eta}_k \right) + \lambda_{k,i}^2 M \left((M+1) \hat{\beta}_k + \hat{\eta}_k \right) \right) \\
&+ \lambda_{i,k}^2 \left((M+1) \hat{\beta}_k + M \hat{\eta}_k \right). \tag{C.9}
\end{aligned}$$

C.2 Derivation of Ψ for imperfect CSI

This section computes the average value of Ψ , which will be used for the signal-to-interference-plus-noise ratio (SINR) calculations.

C.2.1 Computation of L_1

This section derives L_1 . By using matrix multiplication identities, L_1 can be simplified as follows:

$$L_1 = \sum_{i=1}^K \sum_{j=1}^K \sum_{k=1}^K \sum_{l=1}^K \sum_{m=1}^K \lambda_{i,j} \lambda_{k,l} \alpha_{m,m} \left(\mathbb{E} \left[\hat{\mathbf{h}}_i^T \hat{\mathbf{h}}_k^* \hat{\mathbf{h}}_l^H \hat{\mathbf{h}}_m \hat{\mathbf{h}}_m^H \hat{\mathbf{h}}_j \right] + \mathbb{E} \left[\hat{\mathbf{h}}_i^T \hat{\mathbf{h}}_k^* \hat{\mathbf{h}}_l^H \mathbf{e}_m \mathbf{e}_m^H \hat{\mathbf{h}}_j \right] \right) \tag{C.10}$$

Then, L_1 can be derived in closed-form by looking at all the possibilities for used variables as

$$\begin{aligned}
L_1 &= \sum_{i=1}^K \lambda_{i,i}^2 M(M+1) \hat{\beta}_i^2 \left((M+2) \hat{\beta}_i + \hat{\eta}_i \right) + \sum_{i=1}^K \sum_{j=1, j \neq i}^K \sum_{m=1, m \neq i, j}^K \lambda_{i,j}^2 M^2 \hat{\beta}_i \hat{\beta}_j \beta_m \\
&+ \sum_{i=1}^K \sum_{j=1, j \neq i}^K M \hat{\beta}_i \left(\lambda_{i,i}^2 (M+1) \hat{\beta}_i \beta_j + \lambda_{i,j}^2 (M+1) \hat{\beta}_j \left(\hat{\beta}_i + M \hat{\beta}_j \right) \right) \\
&+ \lambda_{i,j} \lambda_{j,i} (M+1) \hat{\beta}_j \left(\hat{\beta}_i + \hat{\beta}_j \right) + \lambda_{i,j}^2 M \hat{\beta}_j \left(\hat{\eta}_i + \hat{\eta}_j \right) \\
&+ \lambda_{i,j} \lambda_{j,i} \hat{\beta}_j \left(\hat{\eta}_i + \hat{\eta}_j \right). \tag{C.11}
\end{aligned}$$

C.2.2 Computation of L_2

By using matrix multiplication theories, L_2 can be simplified as follows.

$$L_2 = \sum_{i=1}^K \sum_{j=1}^K \lambda_{i,j} \mathbb{E} \left[\hat{\mathbf{h}}_i^T \hat{\mathbf{H}}^* \mathbf{\Lambda}^H \hat{\mathbf{H}}^H \hat{\mathbf{h}}_j \right]. \tag{C.12}$$

By using the results in (C.4), (C.12) is written as

$$L_2 = \sum_{i=1}^K \sum_{j=1}^K (\lambda_{i,j}M + \lambda_{j,i}) M \lambda_{i,j} \hat{\beta}_i \hat{\beta}_j. \quad (\text{C.13})$$

C.3 Important Results

The following results are used to derive (5.19a):

$$\mathbf{H}^* \mathbf{H}^H = \mathbf{H} \mathbf{H}^T = \sum_{n=1}^K \mathbf{h}_n^* \mathbf{h}_n^H, \quad (\text{C.14a})$$

$$\mathbf{H}^* \mathbf{\Lambda} \mathbf{H}^H = \mathbf{H} \mathbf{\Lambda} \mathbf{H}^T = \sum_{i=1}^K \sum_{j=1}^K \lambda_{i,j} \mathbf{h}_i^* \mathbf{h}_j^H. \quad (\text{C.14b})$$

Next, some of the important expected value results that were used in Appendices C.1 and C.2 are listed.

$$\mathbb{E} [\hat{\mathbf{h}}_k^H \hat{\mathbf{h}}_k] = \mathbb{E} [\|\hat{\mathbf{h}}_k\|^2] = M \frac{P \alpha_{p,k} \beta_k^2}{(P \alpha_{p,k} \beta_k + 1)} = M \hat{\beta}_k. \quad (\text{C.15a})$$

$$\mathbb{E} [\hat{\mathbf{h}}_k^H \hat{\mathbf{h}}_j] = 0 \quad k \neq j. \quad (\text{C.15b})$$

$$\mathbb{E} [|\hat{\mathbf{h}}_k^H \hat{\mathbf{h}}_j|^2] = M \hat{\beta}_k \hat{\beta}_j \quad k \neq j. \quad (\text{C.15c})$$

$$\mathbb{E} [|\hat{\mathbf{h}}_k^H \hat{\mathbf{h}}_k|^2] = (M + 1) \hat{\beta}_k^2. \quad (\text{C.15d})$$

$$\mathbb{E} [\mathbf{e}_k^H \mathbf{e}_k] = \mathbb{E} [\|\mathbf{e}_k\|^2] = M \frac{\beta_k}{(P \alpha_{p,k} \beta_k + 1)} = M \hat{\eta}_k. \quad (\text{C.15e})$$

$$\mathbb{E} [|\mathbf{e}_k^H \mathbf{e}_j|^2] = M \hat{\eta}_k \hat{\eta}_j \quad k \neq j. \quad (\text{C.15f})$$

$$\mathbb{E} [|\mathbf{e}_k^H \mathbf{e}_k|^2] = M(M + 1) \hat{\eta}_k^2. \quad (\text{C.15g})$$

C.4 Asymptotic SINR with power scaling

Based on (A.1) and (A.2), the following asymptotic limit results can be established [135, 145]:

$$\frac{\mathbf{H}^H \mathbf{H}}{M} = \mathbf{D}^{\frac{1}{2}} \left(\frac{\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}}{M} \right) \mathbf{D}^{\frac{1}{2}} \xrightarrow[M \rightarrow \infty]{a.s.} \mathbf{D}, \quad (\text{C.16a})$$

$$\frac{\mathbf{V}^H \mathbf{H}}{M} \xrightarrow[M \rightarrow \infty]{a.s.} \mathbf{0}_K, \quad (\text{C.16b})$$

$$\frac{\mathbf{H}^H \mathbf{V}}{M} \xrightarrow[M \rightarrow \infty]{a.s.} \mathbf{0}_K, \quad (\text{C.16c})$$

$$\frac{\mathbf{V}^H \mathbf{H}}{M} \xrightarrow[M \rightarrow \infty]{a.s.} \mathbf{I}_K, \quad (\text{C.16d})$$

By using (C.16a), (C.16b), (C.16c) and (C.16d), the asymptotic limit for $\hat{\mathbf{H}}$ can be derived as follows:

$$\begin{aligned} \frac{\hat{\mathbf{H}}^H \hat{\mathbf{H}}}{M} &= \frac{1}{M} \left[\left(\mathbf{H} + \frac{\mathbf{V}}{\sqrt{P_p}} \right) \tilde{\mathbf{D}} \right]^H \left(\mathbf{H} + \frac{\mathbf{V}}{\sqrt{P_p}} \right) \tilde{\mathbf{D}} \\ &\xrightarrow[M \rightarrow \infty]{a.s.} \tilde{\mathbf{D}}^H (\mathbf{D} + \mathbf{I}_K) \tilde{\mathbf{D}} = \text{diag} \left(\frac{P_p \beta_k^2}{1 + P_p \beta_k} \right). \end{aligned} \quad (\text{C.17})$$

Similarly, when the pilot power P_p is scaled as $P_p = E_p/\sqrt{M}$ the limit results can be written as

$$\frac{\hat{\mathbf{H}}^H \hat{\mathbf{H}}}{\sqrt{M}} \xrightarrow[M \rightarrow \infty]{a.s.} \text{diag} \left(E_p \beta_k^2 \right). \quad (\text{C.18})$$