Improving Performance of NOMA and RSMA Systems with Improper Gaussian Signaling

by

ISLAM ABU MAHADY

Graduate Program

in

Department of Electrical and Computer Engineering

A thesis submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy (Ph.D.)

The Faculty of Graduate Studies

Lakehead University

Thunder Bay, Ontario, Canada

April 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: **Akram Bin Sediq**
Adjunct Research Professor, Carleton University, ON, Canada,
Principle Developer at Ericsson Inc., Canada.

Internal Members: **Maysa Yaseen**
Assistant Professor, Dept. of Electrical and Computer Engineering,
Lakehead University, Canada

Malek Alsmadi
Adjunct Assistant Professor, Dept. of Electrical Engineering,
Lakehead University, Canada

Supervisors: **Salama Ikki**
Professor, Dept. of Electrical and Computer Engineering,
Lakehead University, Canada
Ebrahim Bedeer
Assistant Professor, Dept. of Electrical and Computer Engineering,
University of Saskatchewan, Canada

I hereby confirm that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated).

I understand that my thesis may be made electronically available to the public.

Abstract

With the new wave of beyond fifth generation (B5G) and sixth generation (6G) communication systems, there is a perpetual demand for more wireless services with higher data rates, lower latency, and greater connectivity. In order to meet these growing expectations, new candidate technologies (e.g., small cells, millimeter wave, and massive multiple-input multiple-output) have been introduced. Non-orthogonal multiple access (NOMA) has been presented among the most promising strategies for wireless applications due to its effectiveness in supporting heavily-loaded systems by serving users with diverse channel conditions in the same time-frequency resources. NOMA makes it possible to allocate one resource (frequency, time, code, or spatial) to serve multiple users at once by employing superposition coding at the transmitter side and successive interference cancellation (SIC) at the receiver side, resulting in more spectral-efficient and energy-efficient systems.

Recently, rate-splitting multiple access (RSMA) has emerged as a more generalized multiple access technique than NOMA which can serve various under-loaded and over-loaded wireless applications by taking advantage of the common streams to better manage the interference. RSMA offers a flexible interference management technique by enabling an intelligent combination of transmitter-side and receiver-side interference mitigation rather than fully mitigating the interference at the receiver side as in NOMA. The RSMA strategy involves splitting user messages and employing a non-orthogonal transmission scheme, where common messages are decoded by multiple users, and private messages are decoded by their respective users. This approach enhances performance across a broader range of network loads, improving spectral and energy efficiency as well as user fairness.

Improper Gaussian signalling (IGS) has emerged as a signal processing tool and potential alternative to the proper Gaussian signalling (PGS) to improve the spectral efficiency of interference-limited 5G and beyond networks. IGS achieves higher degrees of freedom than PGS due to its capability to control the interference signal dimension. IGS can be viewed as a type of interference alignment, where interference is effectively eliminated by confining it to a single orthogonal signal space dimension, allowing the desired signal to be decoded from the remaining orthogonal dimension.

In this thesis, we investigate the potential performance merits of using IGS in the downlink

interference-limited NOMA systems assuming practical scenarios including SIC imperfection in point-to-point NOMA systems and imperfect self-interference cancellation in cooperative full-duplex NOMA systems etc. We also investigate the potential performance merits of using IGS in RSMA as a generalization scheme of NOMA system.

In the first part of this thesis, a point-to-point downlink NOMA system is studied, where the IGS strategy is adopted to compensate for the performance loss caused by imperfect SIC. New closed-form expressions for achievable user rates are derived when users employ the IGS strategy. Joint optimization problems are then formulated to maximize the overall spectral efficiency and energy efficiency of a two-user NOMA system, subject to minimum user-rate requirements and total power constraints. Sub-optimal solutions for the IGS circularity coefficients and power allocation are derived for the formulated problems. Additionally, improper constellation diagrams are designed using widely linear transformation (WLT) with the predefined optimized IGS coefficients to analyze the impact of IGS on throughput and error performance.

In the second part of this thesis, a downlink cooperative full-duplex NOMA (FD-NOMA) system employing IGS under imperfect self-interference cancellation is analyzed. Optimization problems are formulated and solved to maximize the sum rate, achieve max-min rate fairness, and enhance energy efficiency. These problems involve the joint optimization of the circularity coefficients of the IGS and the power allocation at the base station, subject to each user's rate constraints. We propose iterative algorithms based on solving the Karush-Kuhn-Tucker (KKT) conditions to derive sub-optimal solutions to the formulated problems. Additionally, we illustrate the impact of the IGS circularity coefficient on the constellation diagram of each user.

In the third part of this thesis, we consider a downlink cellular system using RSMA transmission scheme at the base-station with IGS to serve multiple users. We first derive the achievable user private rate and user common rate considering IGS is used for the common message. Then, we maximize the private sum rate of the users' private rates subject to certain minimum users' common rate constraint. In this optimization problem, we optimize the IGS circularity coefficients and power allocation.

The thesis results show that the performance of IGS-based NOMA/RSMA system outperforms its counterpart PGS-based NOMA/RSMA system under the realistic hardware imperfections.

Dedication

To my dearest parents, for their support and encouragement

To my family, my husband Ali and our adorable six kids, Tasnim, Mohamed, Sarah, Israa, Saja, and Sena for their patience.

To the soul of the martyrs, my parents, MoM (Najah), Dad (Abd-Elfattah), my brothers, Mohamed, Ahmed, Khaled, Mahmoud, Saleh, Anas, and my sister, Israa. May Allah shower his mercy on them.

Acknowledgements

First and foremost, all praise and thanks are due to Allah (God) for giving me the power to believe in my passion and pursue my dreams. I could never have finished this thesis without the faith I have in Him and depending on Him.

I would like to express my gratitude and thanks to my supervisors Prof. Salama Ikki and Prof. Ebrahim Bedeer for their roles as great advisors and all the support they provided me both technically and socially. Their support and insightful directions were crucial to my academic success and my development as a researcher.

I am also thankful to Prof. Abdelhamid Tayebi, Prof. Sameh Sorour (May Allah shower his mercy on him), Assistant Prof. Maysa Yaseen, Assistant Prof. Malek Alsmadi, and Research Prof. Akram Bin Sediq, for serving on my thesis committee and for providing constructive comments that made this thesis better.

Finally, I would like to express my greatest and deepest appreciation to my family. My sincere Duaa to my parents, the martyrs, my mother and my father whose endless love and sincere prayers have always been with me (May Allah accept them from Shuhdaa'). Special thanks to Ali, my husband and the best friend, for all the tremendous love and all the support at all the times he gives to me. Thank you Ali for being a steadfast source of encouragement and inspiration to me. It is no exaggeration to say that without your sacrifice and infinite patience, this thesis would never have existed. Thanks to my brothers and sister who are alive or left this life (the martyrs, May Allah shower his mercy on them) for the constant encouragement they always give to me. Thanks also to my father-in-law and mother-in-law for their continuous support as my parents. Last but not least, thanks to our kids, Tasnim, Mohamed, Sarah, Israa, Saja, Sena for being so cute and filling our life with joy and hope.

Table of Contents

List of Tables

List of Figures

List of Symbols

$\mathbf{x}$	vector $\mathbf{x}$
$\mathbf{X}$	matrix $\mathbf{X}$
n_i	thermal noise for terminal i
σ_i^2	variance of noise level at node i
$(\mathbf{x})^T$	transpose of vector $\mathbf{x}$
$(\mathbf{x})^H$	conjugate Transpose of vector $\mathbf{x}$
$(\mathbf{x})^*$	complex conjugate of vector $\mathbf{x}$
$\|\mathbf{x}\|$	euclidean norm for vector $\mathbf{x}$
min	minimization
max	maximization
$\mathcal{E}(a)$	expectation of random variable a
$\log(a)$	natural logarithm of a
$(\mathbf{X})^\dagger$	Pseudo-inverse of matrix $\mathbf{X}$
Pr(a)	Probability of event a
$K(x)$	Modified Bessel function
$F(x)$	Cumulative distribution function
$f(x)$	Probability density function
$\mathcal{CN}$	complex Gaussian random variable
$\in$	within the group
P_{out}	Outage probability
P_e	Probability of error

Abbreviations

1G first generation 1

2G second generation 1

3G third generation 1

3GPP 3rd Generation Partnership Project 14, 29, 30, 42

4G fourth generation 1

5G fifth generation 1, 29

6G sixth generation 1, 3, 4, 7, 8, 126

AWGN additive white Gaussian noise 3, 29

B5G beyond fifth generation xiii, xiv, 1, 3, 4, 7, 8, 12, 17, 27, 126, 130, 131

BS base station 4, 8, 18, 19, 23, 27, 35, 42–44, 49, 50, 57, 58, 63, 70, 75–81, 83–85, 127, 128, 131

CDMA code division multiple access 1

CR cognitive radio 27, 30, 43, 76

CSCG circularly symmetric complex Gaussian 3, 29, 41

CSI channel state information 29, 108

DoF degrees of freedom 26, 29

eMBB enhanced mobile broadband 29

FD-NOMA full-duplex NOMA 4, 5, 76–79, 82, 90, 96–98, 106, 127, 129

FDMA frequency division multiple access 1

GSM Global System for Mobile communication 30, 42

HD-NOMA half-duplex relaying NOMA 19, 76

IGS improper Gaussian signaling ix, 3–5, 9, 28, 30–33, 39, 42–44, 68, 69, 74, 76–79, 81, 97, 98, 107, 126–131

IoT Internet of Things 7

KKT Karush-Kuhn-Tucker 4, 37, 38, 44, 52, 55, 126, 127

LTE long term evolution 1, 8, 14

MIMO multiple-input multiple-output 1, 7, 13, 30, 42, 128, 130

MISO multiple input multiple output 27, 29, 30, 128

MMF maximin fairness 15

mmW Millimeter Wavelength 1, 7, 14, 27, 130, 131

NOMA non-orthogonal multiple access xiv, 2–5, 8–19, 22, 24–28, 30, 31, 35, 39, 40, 42–44, 57, 75–77, 126–132

NR new radio 1

OFDMA orthogonal frequency division multiple access xiv, 1, 11, 12

OMA orthogonal multiple-access 1, 2, 12, 18, 25, 28, 31, 39, 42, 129

PGS proper Gaussian signaling 3, 42, 44, 68, 69, 77, 78, 85, 97, 98, 108, 127, 128, 152

QoS quality-of-service 7, 8, 43

RB resource block 1, 2

RS rate splitting 28, 29

RSMA rate splitting multiple access 2–5, 8, 9, 22–29, 35, 39, 40, 107, 126, 128, 130–132

SDMA space division multiple access 2, 27, 29

SIC successive interference cancellation 2–5, 9, 10, 12–15, 17, 18, 22, 25, 26, 30, 40, 42–44, 57, 76, 79, 80, 126–130

SISO single input single output 22, 26, 28, 108, 128, 131

UAV unmanned aerial vehicle 131

UAVs unmanned aerial vehicles 27, 131

URLLC ultra reliability and low-latency communications 29

WLT widely linear transformation 4, 44, 126

Chapter 1

Introduction

1.1 Background and Motivation

For the past four generations of wireless cellular systems, the enabling technologies named as orthogonal multiple-access (OMA) have been the fundamental access techniques [1], [2]. In principle, frequency division multiple access (FDMA) was adopted in the first generation (1G) of analog technology. Time division multiple access (TDMA) was employed in the second generation (2G) of digital communications technology. Then, code division multiple access (CDMA) became the powerful multiple access technique in the third generation (3G) systems [3]. Recently, OFDMA was standardized for the fourth generation (4G), known as long term evolution (LTE) systems and for the fifth generation (5G), known as new radio (NR) [4]. However, all OMA techniques serve only a single user in each orthogonal time/frequency/code resource block (RB), which leads to spectral inefficiency.

With the new wave of beyond 5G (B5G) communication systems [5] and sixth generation (6G) mobile communication systems [6], [7], there is a perpetual demand for more wireless services with higher data rates, lower latency, and greater connectivity. In order to meet these growing expectations, new candidate technologies (e.g. small cells, Millimeter Wavelength (mmW), and massive multiple-input multiple-output (MIMO)), new

signaling, and multiple access techniques have been introduced [9], [10]. Recently, NOMA has emerged as one of the most promising strategies for wireless applications due to its effectiveness in supporting heavily loaded systems with limited resources and enhancing overall spectral efficiency [11]. Traditional OMA techniques, however, are less suited to accommodate massive connectivity, as they allocate orthogonal resources to different users to avoid intra-channel interference, leading to inefficient use of scarce resources.

As a potential technique to overcome the shortcomings of OMA techniques, the NOMA strategy is introduced with the concept of serving more than one user on the same RB. NOMA can be implemented in various domains, such as power, code, and others [12], [13], [14]. The basic idea in power-domain NOMA is to multiplex users' signals on the same RB by superimposing them at different power levels. Then, the successive interference cancellation (SIC) technique is used at the receiver to differentiate the signals of different users [15]. This makes NOMA a more spectrally efficient technique compared to OMA systems. It is being proposed as an emerging solution for next-generation wireless communication systems due to its strong ability to achieve higher rates and fairness among users.

Rate splitting multiple access (RSMA) has been introduced as a novel technology for optimizing multiple access and interference management strategies for future wireless systems [16–19]. In particular, RSMA is considered a generalization of the existing four special cases: OMA, physical-layer multicasting, precoded space division multiple access (SDMA), and NOMA. The RSMA strategy relies on splitting user messages, as well as the non-orthogonal transmission of common messages decoded by multiple users and private messages decoded by their corresponding users [9]. By doing so, RSMA overcomes two extreme obstacles in interference management strategies: fully decoding interference (as in NOMA) and treating interference as noise (as in SDMA).

NOMA and SDMA are known to be well-suited for strong and weak interference levels, respectively. However, both schemes still do not perform well for medium interference

levels [9]. The RSMA scheme is proposed as a flexible approach that can adapt to different interference levels, rather than operating at the extremes of weak and strong interference. It has been proven that RSMA also offers performance improvements in terms of spectral efficiency, energy efficiency, and user fairness for a wider range of network loads, including both under-loaded and over-loaded systems [19].

Proper Gaussian signaling (PGS) has been widely used in communication systems due to its entropy-maximizing property, which is known to achieve maximum throughput in an additive white Gaussian noise (AWGN) channel, but not in the case of an interference channel [20–22]. PGS relies on proper circularly symmetric complex Gaussian (CSCG) signals, with independent and equal variance for the real and imaginary components [23–25]. If the real and imaginary components of the complex Gaussian signals have non-identical variances and/or are correlated, the signals are considered improper [23, 25].

It has been shown that IGS achieves higher degrees of freedom than PGS due to its ability to identify the interference signal dimension. In fact, it can be described as a form of interference alignment technique, as it mitigates interference by aligning it in only one orthogonal signal space dimension and extracting the desired signal from the other orthogonal space [27]. Due to these impropriety characteristics, IGS has been identified as a potential candidate for improving spectral efficiency in interference-limited systems [24], [25].

Since many practical scenarios in B5G and 6G systems are interference-limited, enabling IGS in such systems will play a vital role in enhancing their spectral and energy efficiencies. Cooperative NOMA under imperfect self-interference cancellation, as well as non-cooperative NOMA and RSMA systems under imperfect SIC, are considered interference-limited systems. To mitigate the impact of self-interference cancellation and SIC imperfections, and to improve the overall performance of these interference-limited systems, IGS is adopted in this research work.

1.2 Objectives and Contributions of the Thesis

The main objective of this thesis is to explore the potential performance benefits of using IGS in downlink interference-limited NOMA and RSMA systems as viable candidates for B5G and 6G networks. This thesis addresses several practical challenges in NOMA and RSMA systems, aiming to improve the overall system performance. Additionally, proposed solutions and designs are presented to enhance spectral and energy efficiencies by adopting IGS for both cooperative and non-cooperative relaying NOMA and RSMA systems. The main contributions of the thesis can be summarized as follows:

1. The performance of using IGS in point-to-point interference-limited NOMA systems is investigated under the practical assumption of imperfect SIC. New closed-form expressions for achievable user rates are derived when users adopt IGS. Joint optimization problems are formulated to maximize the overall spectral efficiency and energy efficiency of a two-user NOMA system, subject to minimum user-rate requirements and total power constraints. Sub-optimal solutions for IGS circularity coefficients and power allocation are derived for the formulated problems. Furthermore, improper constellation diagrams are designed using widely linear transformation (WLT) with predefined optimized IGS coefficients to analyze the impact of IGS on throughput and error performance.

2. We consider a cooperative full-duplex NOMA (FD-NOMA) system employing IGS as another interference-limited system under imperfect self-interference cancellation. We formulate and solve optimization problems for sum rate maximization, max-min rate fairness, and energy efficiency maximization, where the circularity coefficients of the IGS and power allocation at the base station (BS) are jointly optimized under each user's rate constraint. We propose iterative algorithms based on solving the Karush-Kuhn-Tucker (KKT) conditions to obtain sub-optimal solutions to the formulated problems, and we illustrate the effect of the IGS circularity coefficient

on the constellation diagram of each user. Additionally, we design improper constellations using the optimized IGS coefficients to study the impact of IGS on error performance.

3. As an advanced newly emerged multiple access technique, RSMA in conjunction with IGS is investigated under the impact of imperfect SIC. We derive new expressions for the achievable user rates (common rate and private rate), assuming IGS is used for common message signaling. Using these derived expressions, we formulate a sum rate maximization problem for the users' private rates, with the constraint of maintaining the minimum users' common rate above a certain threshold. We determine optimal solutions for the optimized IGS circularity coefficients and power allocation at the BS.

1.3 Thesis Outline

The structure of the thesis is organized as follows:

Chapter 2 introduces relevant background on NOMA, RSMA, and IGS. In particular, the definition of IGS and preliminary analysis are explained.

Chapter 3 addresses resource allocation in point-to-point downlink NOMA systems using IGS under imperfect SIC. The chapter begins with an introduction, followed by a review of related work. Next, the system model and problem formulation are presented. The chapter concludes with the proposed solutions and a discussion of their performance.

Chapter 4 focuses on resource allocation in cooperative FD-NOMA systems using IGS. This chapter starts with an introduction, followed by a discussion of related works and the contributions of the research. The system model and problem formulation are then proposed, followed by an illustration of the proposed solutions. Finally, simulation results are presented to evaluate the proposed solutions.

Chapter 5 discusses the RSMA scenario in a point-to-point system using IGS. This

chapter begins with an introduction, followed by the presentation of the system model and the problem of maximizing sum rate. Numerical results are then provided to demonstrate the performance of the proposed algorithms.

In Chapter 6, we provide a summary of our research contributions and conclusions. Additionally, we discuss open challenges for future research.

1.4 List of Publications

1. I. Abu Mahady, E. Bedeer, S. Ikki, and H. Yanikomeroglu, "Sum rate maximization of NOMA systems under imperfect successive interference cancellation," *IEEE Communications Letters,* vol. 23, no. 3, pp. 474–477, Mar. 2019. (***+156 citations***)

2. I. Abu Mahady, E. Bedeer, S. Ikki, and H. Yanikomeroglu, "NOMA Spectral Efficiency Maximization with Improper Gaussian Signaling and SIC Imperfection" *IEEE International Conference on Communications (ICC),* Montreal, June, 2021.

3. I. Abu Mahady, E. Bedeer, and S. Ikki, "Non-orthogonal multiple access: The case of improper Gaussian signaling and imperfect successive interference cancellation," *Frontiers in Communications and Networks- Communications Theory Journal,* vol. 3, March 2022. (***Invited Journal Paper***).

4. I. Abu Mahady, E. Bedeer, S. Ikki, and H. Yanikomeroglu, "Energy efficiency maximization of full-duplex NOMA systems with improper Gaussian signaling under imperfect self-interference cancellation", *IEEE Communications Letters,* vol. 26, no. 7, pp. 1613-1617, July 2022.

Chapter 2

Background and Preliminaries

The B5G and 6G cellular communication systems attract significant attention from academia and industry. The new requirements of B5G and 6G are highly demanding since they will need to support the Internet of Things (IoT), provide wireless services with much higher data rate (ultra-high bandwidth with Gb/s applications 3D video), ultra-high reliability and low latency (e.g. autonomous cars, surgery operations), massive connectivity (e.g. massive IoT ultra-low power requirements) as depicted in Fig. 2.1, beside, heterogeneous quality-of-service (QoS), very accurate localization, and computing.

Based on mobile operators data, it is predicted that by 2030, the demand for mobile traffic data volume will be ten thousands times more than that in 2020. Figure 2.2 shows total global monthly data (ExaBytes per month) from quarter 3 (Q3) 2017 to Q3 2024 [26], along with the year-on-year percentage change for mobile network data traffic. It demonstrates a continued huge growth in data traffic. It also shows that data traffic grew around 15% quarter-on-quarter and around 75% year-on-year. This growth is driven by the exponential increase in smart phone subscriptions and average data volume per subscription, which has been boosted primarily by more use for video content.

To achieve the requirements for the enhanced capacity, new technologies such as mmW frequency bands, massive MIMO, and small cells (densification) are proposed. On the

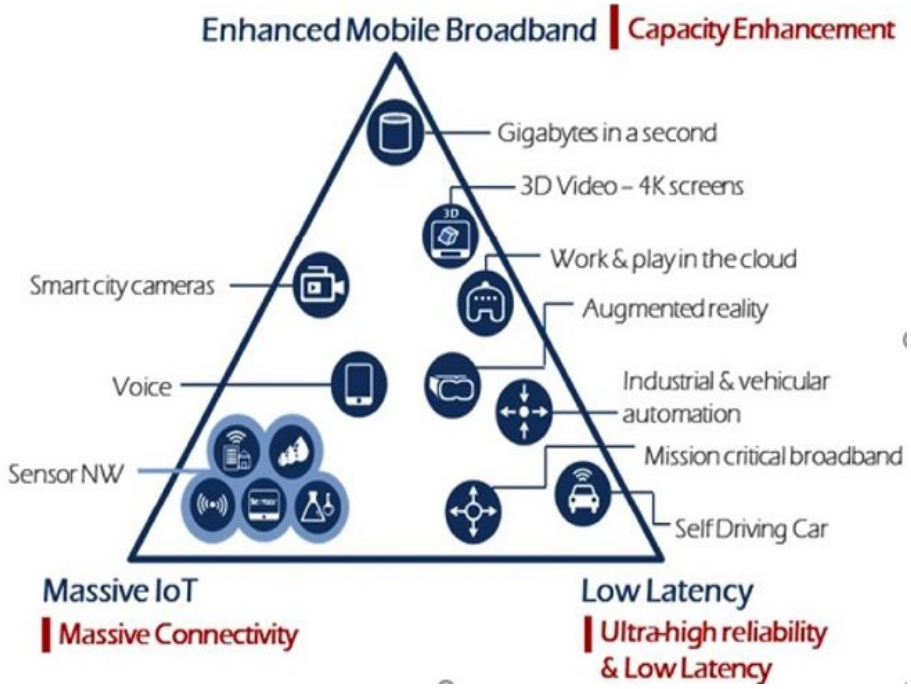

Figure 2.1: B5G applications requirements (source: ITU-R IMT 2020 requirements).

other hand, for ultra-high reliability and low latency applications, the latency in the current LTE networks is around 15 ms, and the target is to make it less than 5 ms. To achieve this goal, one of the potential solutions is to use shorter symbol periods, in order of 4 μs rather than 74 μs in LTE. Another solution for ultra-low latency is to use flexible transmission time interval (TTI) schemes, where these schemes allow for slot sizes that can vary according to the length of the packet or transport block. Machine-to-machine (M2M) and vehicle-to-vehicle (V2V) communications without the need for a BS to control are typical low-latency applications.

As for the massive connectivity, B5G networks need to connect billions of sensors and machines, from watches to refrigerators, to parking meters and cars. New multiple access techniques e.g., NOMA and RSMA, to achieve the requirements of connecting large number of applications with different traffic and power requirements are proposed in the literature. However, it is still challenging to achieve different QoS, latency, and power requirements.

Therefore, B5G and 6G networks require a more efficient utilization of the wireless

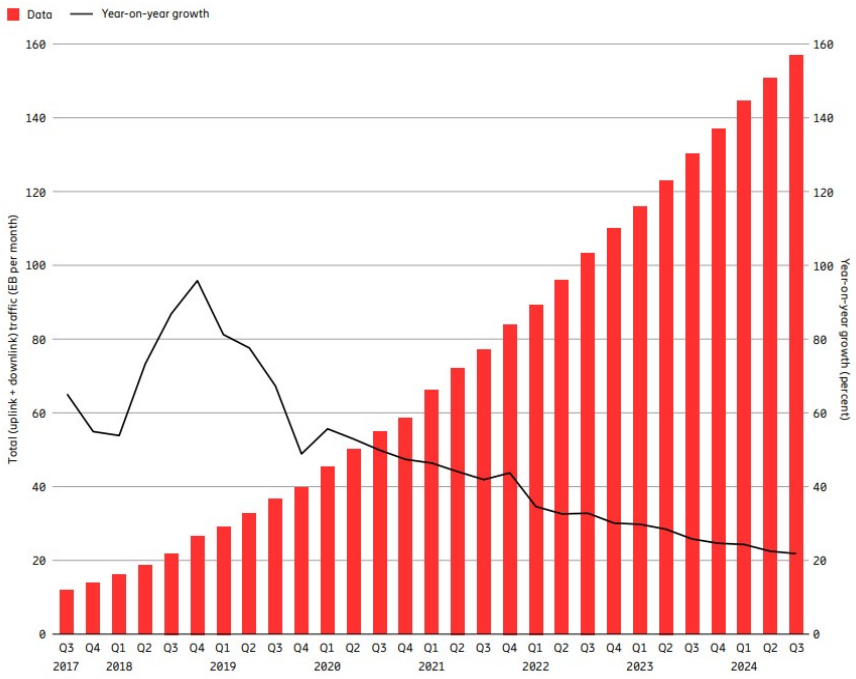

Figure 2.2: Mobile data traffic growth predicted by Ericsson [26].

scarce resources and more powerful means to manage interference. This has triggered a rethinking of proposing new signalling of physical layer and multiple access techniques for wireless communication systems [2, 9]. In this research work, we focus on NOMA, RSMA, and IGS as promising multiple access techniques and new signalling technique, respectively.

2.1 NOMA Systems

NOMA has been introduced as a potential candidate for enabling heavily-loaded systems using limited resources while improving the spectral efficiency [12]. NOMA proposes the adoption of power/code domain to multiplex signal streams from multiple users together and allow them to transmit simultaneously using the same frequency/time/code resources [14]. The popular power-domain NOMA employs the superposition coding at the transmitter side to superimpose the signals of multiple users together by differentiating them in the power domain. At the receiver side, SIC is used to separate multiplexed users' signals. As illustrated in Fig. 2.3 [15], the superposition coding scheme and SIC

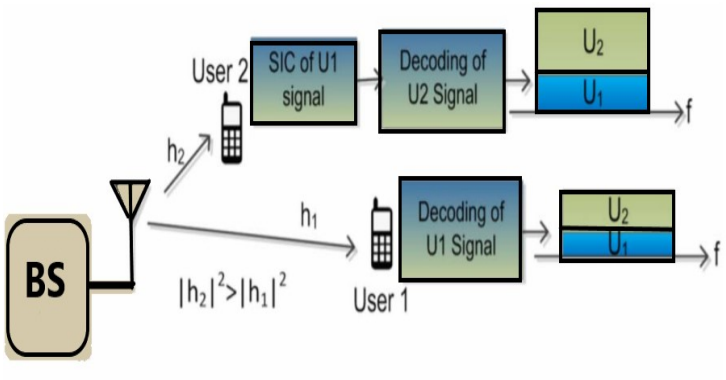

Figure 2.3: Fundamental structure of power-domain NOMA system [15].

process for two users scenario can be implemented by the following steps

- At the BS, the transmitted signal is a combination of superimposed coded signals from the two users.
- At the receiver side, U1 with the weaker channel treats U2's signal as noise and decodes its message from its received signal.
- U2, with the better channel, performs SIC, i.e., it first decodes U1's message and subtracts U1's signal from its received signal. Afterward, U2 can decode its own message.

2.1.1 Advantages of NOMA over OMA

Consider a downlink NOMA system where a BS transmits superimposed signal $x = x_1 + x_2$ to two users (U1 and U2). The channel gains from the BS to the two users are h_1 and h_2, respectively. Assume $|h_2| > |h_1|$, and both h_1 and h_2 are perfectly known at the BS and receiver sides. Therefore, the received signal at each user can be written as

$$y_i = h_i x + n_i, \forall i = 1, 2, \tag{2.1}$$

where $n_i \sim \mathcal{CN}(0, \sigma_n^2)$ is independent and identically distributed (iid) complex additive white Gaussian noise (AWGN) with zero mean and variance σ_n^2.

According to the Shannon's capacity formula, the data rate of U1 and U2 in a NOMA setup can be computed in bits/sec/Hz, respectively, as

$$R_1 = \log_2\left(1+\frac{\alpha_1 P|h_1|^2}{\alpha_2 P|h_1|^2+\sigma_n^2}\right) \quad (2.2)$$

$$R_2 = \log_2\left(1+\frac{\alpha_2 P|h_2|^2}{\sigma_n^2}\right), \quad (2.3)$$

where P is the assigned power on each channel. $\alpha_i, \forall i = 1, 2$ is the power allocation factor. Therefore, the sum rate of a NOMA system can be written as

$$R_{\text{NOMA}} = \log_2\left(1+\frac{\alpha_1 P|h_1|^2}{\alpha_2 P|h_1|^2+\sigma_n^2}\right)+\log_2\left(1+\frac{\alpha_2 P|h_2|^2}{\sigma_n^2}\right). \quad (2.4)$$

In OFDMA systems, we assume OFDM with orthogonal user multiplexing. The data rates of U1 and U2 in OFDMA systems can be, respectively, written as

$$R_1 = \frac{1}{2}\log_2\left(1+\frac{P|h_1|^2}{\sigma_n^2}\right), \quad (2.5)$$

$$R_2 = \frac{1}{2}\log_2\left(1+\frac{P|h_2|^2}{\sigma_n^2}\right). \quad (2.6)$$

Therefore, the sum rate of the two users in OFDMA system can be presented as

$$R_{\text{OFDMA}} = \frac{1}{2}\log_2\left(1+\frac{P|h_1|^2}{\sigma_n^2}\right)+\frac{1}{2}\log_2\left(1+\frac{P|h_2|^2}{\sigma_n^2}\right). \quad (2.7)$$

Using the simulation setup values in [28], we set $\alpha_1 = 0.8$, $\alpha_2 = 0.2$, $\frac{P|h_1|^2}{\sigma_n^2} = 0$ dB, and $\frac{P|h_2|^2}{\sigma_n^2} = 20$ dB. Assume each user has the same equal bandwidth = 1 Hz. Therefore, $R_{\text{OFDMA}} = 3.38$ bits/sec and $R_{\text{NOMA}} = 4.53$ bits/sec. Hence, the gain of the sum rate of the NOMA scheme is around 35% more than that of the OFDMA scheme. If we consider a better channel condition with $\frac{P|h_1|^2}{\sigma_n^2} = 10$ dB, and $\frac{P|h_2|^2}{\sigma_n^2} = 30$ dB, the $R_{\text{OFDMA}} = 6.72$ bits/sec and $R_{\text{NOMA}} = 9.52$ bits/sec. Hence, the gain of the sum rate of the NOMA scheme is around 42% more than that of the OFDMA scheme.

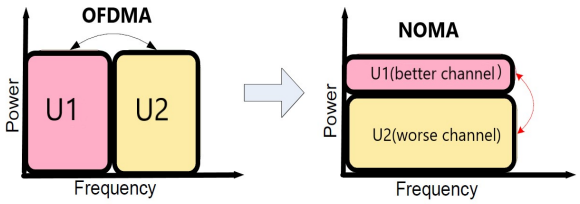

Figure 2.4: OFDMA versus power-domain NOMA system.

Based on the above numerical example, it can be concluded that NOMA systems can outperform OFDMA systems when there is a channel gain difference between the user channels. Additionally, the sum rate gain of NOMA over OFDMA increases as the channel gain difference between the two users becomes larger.

Exploiting both the superposition coding and the SIC technology makes NOMA a more spectral efficient technique over OMA systems. Thus, NOMA is being considered as an emerging solution for future B5G systems due to its great capability of achieving a higher achievable rates and supporting a large number of connected users over the same channel resources. One of the advantages of NOMA systems is that when the available resource blocks are assigned to weak channel users, they can still be accessed by other strong channel users, which qualifies NOMA techniques to achieve a higher overall spectral efficiency [29, 30]. Fig. 2.4 illustrates the comparison between the NOMA system and the OFDMA system. It represents how the OFDMA system divides the frequency resources between the two users while the NOMA system allows the two users to share the frequency with different power levels.

Moreover, NOMA is capable of achieving a better tradeoff between system overall rate and user fairness than OMA. This can be obtained by allocating less power to users with strong channel conditions and more power to users with weak channel conditions. This is quite opposite to the conventional water-filling power allocation scheme that is widely used in OMA, where good channels get more power and vice versa. However, NOMA techniques achieve this potential higher spectral efficiency considering perfect SIC (see, e.g., [14, 29, 30], and the references therein). In practice, SIC technique is imperfect. So

it is desirable to investigate the impact of imperfect SIC on NOMA systems performance and find signal processing ways to reduce that severe impact. This is one of our objectives in this research work.

2.1.2 Successive Interference Cancellation (SIC)

In real scenarios, the assumption of perfect SIC at the receiver might not be practical, since there still remain several serious implementation problems by using SIC, e.g., error propagation and complexity scaling [14]. In [31], a unified framework is presented assuming imperfect SIC, which shows that the performance converges to an error floor at the high signal-to-noise ratio (SNR) region and obtain a zero diversity order. Hence it is of great interest to compensate the impact of imperfect SIC for the NOMA systems.

SIC is widely known as an interesting type of multi-user detector. The key procedure of SIC is that users are decoded sequentially, with the receiver canceling interference after each user. It allows the user having the strongest signal to be detected first, i.e., the user with the least interference signal, then, the strongest user remodulates its signal, which is then subtracted from the superimposed signal. The same operation is executed by the second strongest user and so on until the weakest user decodes its information without suffering from any interference at all [12]. Since the power-domain NOMA supports the users via different power levels, SIC is considered a suitable multi-user detector [12].

Furthermore, SIC has also been used in various practical technologies, such as multi-user MIMO systems [32], cooperative relaying systems [33], and in large-scale systems modeled by stochastic geometry [34] etc. Moreover, SIC has been implemented in industry, such as IEEE 802.15.4. As a practical tool, SIC has to be considered in a realistic assumption, e.g., imperfect SIC. This issue will be one of the considered practical scenarios in our research work.

2.1.3 Related Works on General NOMA Systems

NOMA has received much attention recently and has been identified as a working item in 3rd Generation Partnership Project (3GPP) Release 15 (Rel. 15) [35], where NOMA is referred to as multi-user superposition transmission. Another version of NOMA, named as layer division multiplexing, has been proposed for next generation digital TV standard advanced television systems committee. Comprehensive research studies for NOMA systems have been conducted in the literature [11]. In [36] and [37], the authors discussed a combination of NOMA with MIMO technologies. NOMA has also been introduced to be used with other technologies, e.g., visible light communication (VLC) [20] and mmW communication [38]. In the following, we will discuss the related works according to the following categories: performance analysis, resource allocation, cooperative NOMA, and full-duplex NOMA, etc.

2.1.4 Related Works on Performance Analysis in NOMA Systems

Several valuable works have investigated the performance of NOMA in terms of downlink transmission [39–44]. In [39], a two-user NOMA downlink transmission employing SIC technique was introduced. A more general NOMA transmission scheme was presented in [40], which examined a BS communicating with multiple randomly deployed users. It was shown that NOMA is capable of achieving much better performance compared to OMA in terms of both its outage probability and its ergodic rate. In [41], the fairness issues were considered by developing suitable power allocation coefficients for the multiple users in a general NOMA downlink transmission scenario. Considering the practical link-adaptation implementations of the LTE, the system-level performance was evaluated in [42] and [43]. In [44], this work focused on a simple two-user scenario, and developed analytical and simulation results to reveal that for this considered scenario, conventional hybrid NOMA is still an optimal transmission strategy.

Motivated by minimizing the signaling overhead required for CSI training, some works

investigated the performance of downlink NOMA transmissions using partial CSI at the transmitter [45–47]. More specifically, authors in [45] addressed the outage probability of NOMA by considering either imperfect CSI or second-order statistics- based CSI. In [46], the outage performance of NOMA was investigated assuming the knowledge of statistical CSI and considering jointly both the decoding order selection and the power allocation of the users. Assuming that only the average CSI was available at the BS, [47] addressed both the optimal decoding order as well as the optimal power allocation of the users in downlink NOMA systems, where both the transmit power of the BS and rate fairness of users were optimized. By assuming only a single-bit feedback of the CSI from each user to the BS, the outage performance of a downlink NOMA system was addressed by [48]. Based on the analytical expressions derived, the proposed power allocation optimization problem was solved by reducing the outage probability.

2.1.5 Related Works on Resource Allocation in NOMA Systems

In NOMA systems, resource allocation has been studied for different performance measures. In the literature, the sum rate maximization is the most commonly adopted objective (see, e.g., [50–54], and the references therein), where the focus is on the assumption of perfect SIC. Fairness is also an important issue in NOMA systems, where the most common fairness indication is the maximin fairness (MMF). Therefore, a number of works has studied resource allocation for MMF, e.g., [53]. The MMF power allocation problem enhances the rate of the weakest of all NOMA users [41, 47, 55]. In [55], the power allocation method was tackled under the assumptions of knowing either the instantaneous CSI or the average CSI. Note that by adopting the MMF rate as our objective function, we can satisfy a predefined level of rate fairness, but at the expense of sacrificing the system's sum rate.

In NOMA scenarios, a feasible proportional fairness strategy is to schedule users based on the instantaneous user rates, while satisfying a certain average target rate [56, 57]. The

weighted sum rate problem is studied in [57] where an additional positive weighing factor for each user's achievable rate is considered, which reflects the priority of each user in the context of resource allocation. By adopting fractional power allocation among users' channels and equal power allocation across sub-channels, the works in [58, 59] compared the system-level performance of the NOMA system with its counterpart OMA system, and demonstrated that the degree of proportional fairness, cell-edge user throughput, and overall cell throughput of NOMA scheme are all superior to their counterpart in the OMA scheme.

With the higher desire for green communications in 5G networks, minimizing energy consumption in NOMA systems has become of great importance for both academia and industry An energy-efficient two-user single-cell NOMA was investigated in [60]. It showed that for a given SE for each user, maximum energy efficiency performance can be attained. In [49], an energy-efficiency optimization problem was developed in downlink NOMA systems under different data rate requirements of the users. It was proved that NOMA outperforms OMA in terms of its energy efficiency.

Table 2.1 lists the works related to the enabling technologies with NOMA systems, including back-scatter communications [61–63], integrated terrestrial satellite networks [64–66], mobile edge computing and caching [67, 68], and intelligent reflecting surfaces [69,70]. Design metric and main observations of the related work are listed in this table 2.1. For full details on existing related works on NOMA systems, the reader is referred to [12, 13, 15], and references therein.

2.2 Cooperative NOMA Systems

Cooperative NOMA is introduced to enhance the overall system coverage and reliability in addition to spectral efficiency [72], [73]. In cooperative NOMA, the near users (users close to the base BS) act as relays for the far users (users far from the BS). The combination of cooperative relaying and NOMA is of interest since the information for the far

Table 2.1: Literature review on using NOMA in B5G technologies

Technology	**Ref.**	**Design Objective**	**Main Finding**
Backscatter Communications	[61–63]	Max. sum rate, Maxmin throughput	NOMA outperforms TDMA
Terrestrial- Satellite Networks	[64, 65]	Max. sum rate	The sum-rate performance of the proposed optimized UAV-location NOMA scheme outperforms an equivalent scheme with OMA deployment
Mobile Edge Computing	[67, 68]	Max. weighted sum rate	The WSR performance of the proposed NOMA scheme outperforms OMA scheme
Intelligent Reflecting Surfaces	[69, 70]	Max. sum rate, Maxmin SINR	The performance of the proposed IRS-NOMA outperforms three benchmark schemes including the conventional NOMA, IRS-OMA, and OMA
mmWave-NOMA	[38]	Beamforming efficiency, power allocation, and user fairness	Hybrid mmWave-NOMA and massive MIMO architectures improve user connectivity and system throughput significantly.
VLC-NOMA	[20]	Spectral efficiency, illumination constraints, and data rate maximization.	VLC-NOMA effectively supports high-speed communication while maintaining proper illumination conditions.

users is known to the near users owing to SIC. Although extra time slot is still needed, cooperative NOMA can serve the data transmission of both near and far users using NOMA. Cooperative NOMA can be classified based on two relaying strategies, namely user relaying (i.e., having a user act as a relay) as in Figs. 2.5[a-b], and dedicated relaying

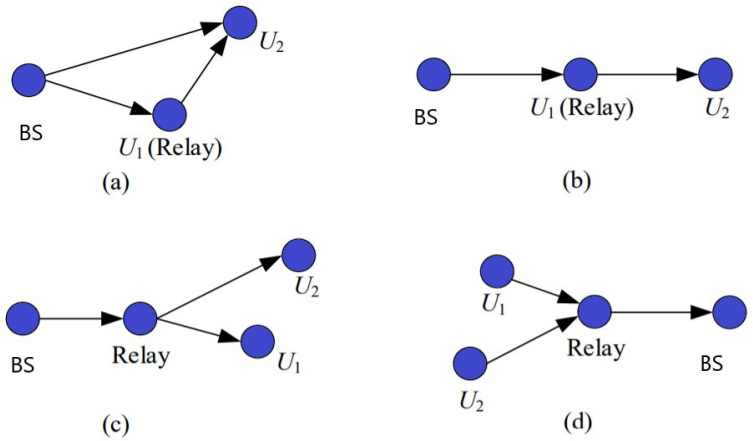

Figure 2.5: Cooperative NOMA structures, user relaying (a,b), and dedicated relaying (c,d).

(i.e., employing an external entity as a relay) as presented in Figs. 2.5[c-d].

In NOMA systems that adopt the dedicated relaying technique, the BS transmits signals to the two users via an external relay. Alternatively, the two users send data to a common destination via an external relay [75]. With user relaying, the BS transmits two superimposed signals to both the strong user U1 and the weak user U2. Afterwards, U1 uses SIC to decode U2's signal, forwarding it to U2. Depending on whether or not a direct link exists between the BS and U2, two scenarios can exist when it comes to user relaying in NOMA systems. The first scenario occurs when U2 receives its signal from both relaying phases and employs a combining method to improve its reception reliability. The second scenario involves U2 receiving its message via U1 [77].

This technique would be especially useful for short-range communication technologies like wireless applications, where users with better channel conditions can act as relays and forward the messages to those with poor channel conditions [72]. Similar to the cooperative relaying in traditional OMA technique, U2 receives two versions of desired signals, one is from the direct link (from BS) and the other one is from relaying link (from the U1). Combining the two versions at U2 leads to what is known as spatial diversity gain, and hence, improves the reliability.

A cooperative relaying NOMA transmission scheme was proposed in [72], [73], to

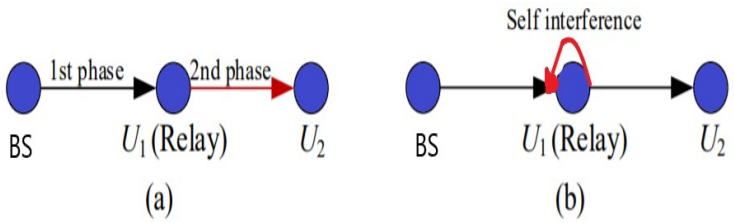

Figure 2.6: Cooperative (a)- HD-NOMA and (b)- FD-NOMA structures.

improve spectral efficiency, where a few users have beforehand information of the other users' messages. In [74], a novel device-to-device (D2D)-assisted cooperative NOMA model with a two-phase transmission scenario was proposed. A cooperative relay NOMA-based system was emerged in [75], where the enhancement of the spectral efficiency was proved by numerical results. A sub-carrier channel assignment and power allocation algorithm was introduced for the cooperative NOMA system in [76]. More works are listed in the survey papers [11, 77].

2.3 Cooperative Full-Duplex NOMA Systems

To further improve the spectral efficiency of half-duplex relaying NOMA (HD-NOMA) systems that require two time slots for receiving and relaying data, full-duplex (FD) relaying has been recently investigated. In particular, the relay user U1 receives the superimposed signals from the BS and concurrently performs relay transmission in the same time slot as illustrated in Fig. 2.6. This overcomes the drawback of HD relaying, which requires an extra time slot for the relaying phase. Both HD and FD relaying strategies are considered in NOMA systems. NOMA with full-duplex relaying (FD-NOMA) further improves the spectral efficiency. In the literature, there is an extensive research regarding HD-NOMA systems, and many facets have been studied including system performance, resource allocation optimization, multiple antennas, etc. (see [72] – [77] and references therein).

Nonetheless, FD-NOMA suffers from self-interference at the FD-relay. To realize FD relaying, practical self-interference cancellation techniques are required to mitigate the

self-interference to tolerable limits. This, however, adds more complexity when compared with half-duplex relaying. Therefore, from practical implementation perspective, whether selecting FD-NOMA or HD-NOMA relaying relies on the signal processing capabilities of the relaying nodes. For instance, HD-NOMA may be suitable for the user relaying with limited capabilities in hardware and signal processing. In the other hand, FD-NOMA can be used for the dedicated relaying with strong signal processing capability.

The major challenge in FD-NOMA is self-interference, which arises from simultaneous transmission and reception at the same node. The work in [78] discusses various self-interference cancellation (SIC) techniques, including analog and digital domain processing, to minimize its impact. The study in [79] explores energy-efficient designs for FD-NOMA, emphasizing the role of power control and relay selection. It highlights research efforts in designing low-power FD transceivers and integrating energy harvesting mechanisms such as Simultaneous Wireless Information and Power Transfer (SWIPT) to improve sustainability in FD-NOMA networks. The work in [80] cites recent studies demonstrating that cooperative FD-NOMA outperforms traditional half-duplex NOMA in terms of outage probability and achievable data rates. Security is a critical aspect of FD-NOMA due to the increased risk of eavesdropping from simultaneous transmissions. [81] reviews physical layer security techniques such as jamming-based security and artificial noise generation to protect user transmissions. It also suggests the use of intelligent reflecting surfaces to enhance security and coverage. The integration of multiple antennas and MIMO techniques in FD-NOMA is another important area covered in [82]. The study highlights beamforming and antenna selection methods to improve signal quality while minimizing interference.

Table 2.2 presents the main works recently handled the cooperative FD-NOMA, where it demonstrates the relaying mode, the design objective, the system model used, and the main finding from the proposed scheme. In common finding in all listed works is that the performance of FD-NOMA is superior to the performance of FD-OMA and TDMA schemes in terms of spectral efficiency [87].

Table 2.2: Literature review on cooperative full-duplex NOMA

Relaying Mode	**Ref.**	**Design Metric**	**System Model**	**Main Finding**
FD-DF (UL-DL)	[82]	maxmin sum of UL+DL rates	Relay-assisted: one BS + one relay + K users, two-antenna nodes	The performance of the proposed asymptotic optimal algorithm outperforms the original standard solution
FD-DF (UL-DL)	[83]	weighted sum rate	Relay-assisted: A multi-carrier NOMA scheme, single-antenna nodes	The sum-rate performance of the proposed scheme is superior to its counterparts with either HD-NOMA and HD/FD-OMA schemes
HD/FD-DF	[84]	maxmin user-rate	Relay-assisted: Two NOMA-based cooperative broadcasting/multicasting schemes, single-antenna nodes	Both proposed schemes offer better performance than a fixed power allocation NOMA scheme as well as an optimized TDMA scheme
FD-DF	[85]	max the strong user-rate	User-assisted: Downlink SWIPT cooperative relaying NOMA scheme, strong user: multiple antennas, BS/weak user: single antenna	The proposed scheme achieves higher data rate performance for the strong user compared to FD-NOMA without SWIPT, HD-NOMA with SWIPT, and conventional OMA with SWIPT schemes
HD/FD-DF, and hybrid-DF	[86]	maxmin user-rate	User-assisted: Hybrid HD/FD cooperative NOMA scheme with transmit power adaptation, BS: single antenna, users: two antennas	The proposed hybrid scheme outperforms HD-cooperative NOMA, FD-cooperative NOMA, and conventional NOMA schemes

2.4 Rate-Splitting Multiple Access (RSMA)

By dividing users in the power domain, power-domain NOMA can simultaneously serve multiple users at the same resource block. Consequently, NOMA can achieve higher spectral efficiency than conventional OMA. However, using SIC in NOMA, the users must decode all of the interfering signals as they receive their own messages [28], which extremely increases the computational complexity required for signal processing. To solve this burden, the novel idea of RSMA is proposed in [16–19].

In RSMA, the signal transmitted to the users is divided into a common message and a private message. The common part is a message decoded by multiple users and the private part is a message that only intended to be received by specific user. At each receiver, the common message is first decoded while treating all private signals as noise. Once the common message is successfully decoded, it is removed from the received signal using SIC, allowing the receiver to then decode its own private message. The original information is successfully delivered provided that each receiver correctly decodes both the common and its corresponding private message. Therefore, adapting the split of common and private signals can compromise the signal processing complexity and the data rate attained by RSMA [19]. However, RSMA still has various challenges in term of implementation in wireless systems such as the split of common and private signals, and synchronization of signal transmission.

A two user RSMA scenario for a single input single output (SISO) system is presented in Fig. 2.7 [9]. The transmitter splits the message W_i of the ith user into a common message $W_{c,i}$ and a private message $W_{p,i}$, $\forall i = 1, 2$, and combines $W_{c,1}$ and $W_{c,2}$ into a common message W_c. The three generated $W_c, W_{p,1}$ and $W_{p,2}$ messages resulted from W_1 and W_2 are independently encoded and linearly precoded at the transmitter side.

At the receiver side, each user first decodes the common stream s_c by treating all private streams as noise. After removing the decoded common stream from the received

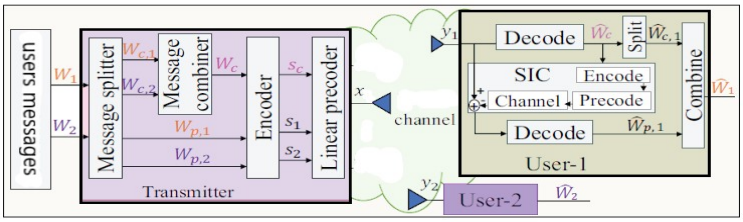

Figure 2.7: Two-user RSMA transmission framework [9].

signal, each user decodes the intended private stream $s_k, \forall k = 1, 2$, by treating the private stream of the other user as noise. As shown in Fig. 2.7, RSMA makes it possible to adaptively manage the interference by fractionally decoding the interference through the common stream decoding and fractionally treating the interference as noise when decoding the intended private stream at each user [19]. Consider a downlink single-cell system that consists of a BS serving two users using RSMA as depicted in Fig. 2.7. The transmitted signal x of the BS can be expressed as

$$x = \sqrt{p_c} s_c + \sum_{k=1}^{2} \sqrt{p_k} s_k, \tag{2.8}$$

where p_c is the transmit power of the common stream s_c (i.e., the common message W_c) and p_k is the transmit power of the private stream s_k (i.e., the private message Wp_k) transmitted to user $k, \forall k = 1, 2$.

The received signal at user k can be written as

$$\begin{aligned} y_k &= \sqrt{h_k} x + n_k, \\ &= \sqrt{p_c h_k} s_c + \sum_{j=1}^{2} \sqrt{h_k p_j} s_j + n_k, \end{aligned} \tag{2.9}$$

where h_k refers to the channel gain between user k and the BS and n_k refers to the AWGN with variance σ^2. Under the assumption of proper Gaussian signaling, the instantaneous

rates for decoding the common and private streams at user k are given as follows

$$R_{c,k} = \log_2 \left(1 + \frac{h_k p_c}{h_k \sum_{j=1}^{2} p_j + \sigma^2} \right), \tag{2.10}$$

$$R_k = \log_2 \left(1 + \frac{h_k p_k}{h_k \sum_{j=1, j\neq k}^{2} p_j + \sigma^2} \right). \tag{2.11}$$

To ensure s_c is successfully decoded by both users, its rate can not exceed

$$R_c = \min\{R_{c,1}, R_{c,2}\}. \tag{2.12}$$

As s_c includes sub-messages $W_{c,1}, W_{c,2}$ of both users, the rate R_c split between the two users is adapted to the portion of sub-messages that each user contributed. Let C_k denotes the amount of rate R_c allocated to user k for $W_{c,k}$ such that $C_1 + C_2 = R_c$. Then the overall achievable rate of user k is given as

$$R_{k,tot} = C_k + R_k. \tag{2.13}$$

Apparently, the rate of each user is split into two parts, namely, relevant part of the rate of s_c (a.k.a. the common rate) and the rate of s_k (a.k.a. the private rate).

2.4.1 Advantages of RSMA over NOMA and OMA

RSMA is a novel multiple access technique that has emerged to overcome the disadvantages of other orthogonal and non-orthogonal multiple access techniques. Here, we list the advantages of RSMA over NOMA and OMA schemes.

- RSMA is a generalized multiple access scheme that achieves equal or better performance compared to NOMA, OMA, and other multicasting schemes [9]. RSMA is a generalized multiple access scheme because it unifies and extends the capabilities of both NOMA and OMA by enabling dynamic message splitting, partial decoding, and interference management, offering greater flexibility and performance in diverse

channel conditions. Therefore, the performance of other OMA and NOMA schemes can be investigated through a single unified RSMA framework.

- RSMA is a good fit for various systems loads including underloaded and overloaded and diverse user deployments by taking advantage of using common messages to mitigate the interference. RSMA offers a flexible interference management technique by enabling an intelligent combination of transmitter-side and receiver-side interference mitigation rather than fully mitigating the interference at the receiver side as in NOMA and using single user transmission to avoid interference (as in OMA).

- RSMA is proved to show better spectral and energy efficiencies performance gain than NOMA and OMA schemes under different channel conditions including user mobility [9, 17, 19].

- RSMA requires a lower transceiver complexity than NOMA scheme since the latter needs to perform multiple SIC processes at each user to detect each of other users' signals. However, RSMA requires only one SIC process at each user. RSMA is also less sensitive to difference channel conditions than NOMA, which results in a simpler scheduler.

However, implementing RSMA in wireless systems also faces various challenges [19] such as ability to split the common and private messages, resource allocation for efficient private message transmission, and synchronization of message transmission.

Table 2.3 summarizes the comparison between different multiple access techniques, OMA, NOMA, and RSMA in terms of their access principles, the receiver architecture, the user deployment scenarios and the network loads they are best useful for. Apparently, the most interesting characteristic that distinguishes RSMA from the other multiple access techniques is its flexible interference management strategy of partially decoding interference and partially treating residual interference as noise. This allows RSMA to include and comprise NOMA, SDMA, OMA, and multicasting as sub-schemes. Fundamentally,

RSMA smoothly bridges different sub-schemes without the need to switch between them in a hybrid mode.

Table 2.3: Comparison between different multiple access techniques: OMA, NOMA, RSMA

Multiple access technique	**OMA**	**NOMA**	**RSMA**
Access principle	Orthogonal resource assignment to avoid interference	Non-orthogonal resource assignment and fully decode interference	Partially decode interference and partially treat interference as noise
Receiver architecture	Treat interference as noise	SIC	SIC
Typical user deployment scenario	Any user channel strengths (similar or different)	Users with large different channel strengths (weak and strong)	Any user channel strengths
Network load	Only one active user in each resource block	Suitable for over-loaded networks	Suitable for under-loaded and over-loaded networks

2.5 Related Works on RSMA

In the literature, a number of existing works investigated interesting problems related to RSMA. The survey in [19] summarized the challenges and opportunities of using RSMA for MIMO wireless networks. RSMA was studied in various single and multiple antenna networks. Works in [17, 88, 89] considered RSMA in SISO systems including interference and broadcasting channels, where the rate regions were defined. The existing literature on RSMA in multi-antenna networks has investigated different metrics such as the degrees of freedom (DoF) region [93], max-min fairness [94], and capacity region [95]. In [93], obvious symmetric DoF gains over NOMA are obtained. Authors in [94] showed that RSMA fully exploits the multi-antenna DoF and the benefits of SIC receivers, and outperforms NOMA in both under-loaded and over-loaded systems. In [95], RSMA has been proved to achieve the sum capacity within a constant gap for the two-user case.

The first generalized RSMA framework was introduced in [96] which used the terminology "RSMA" for the multi-antenna broadcasting. Authors in this work compared the performance of RSMA with NOMA and dirty paper coding (DPC) and illustrated that RSMA is a general scheme that include NOMA and SDMA as sub-schemes. Later, the work in [97] showed that the performance of RSMA outperforms NOMA, and SDMA for a two-user multiple input multiple output (MISO) broadcasting using low-complexity precoding and closed-form optimal power allocation scheme.

The physical layer architecture of RSMA was presented in [98] considering finite constellation modulation schemes and finite length polar codes. The link-level simulations conducted in this work demonstrated that RSMA can achieve significant throughput gains over existing multiple access techniques. Authors in [102] exploited RSMA to tackle user mobility in modern wireless cellular systems. In their link-level simulation setup, authors used a realistic 3GPP channel model and OFDM waveforms. This work showed the efficacy of using RSMA in such mobility networks.

The promising advantages of RSMA in the conventional multi-antenna systems have led to an explosion of applications of RSMA in emerging B5G system architectures as listed in 2.4. RSMA was studied in massive MIMO [102,103], mmW [104], cognitive radio (CR) [105], cooperative relay networks [106], Intelligent reconfigurable surface [107], and unmanned aerial vehicles (UAVs) [108, 117] etc.

Appealing resource allocation methods using RSMA were proposed in the literature. In [124], the authors introduced a distributed RSMA method to maximize the data rates of the users. In [125], the authors made use of RSMA to maximize the rate of all users in downlink multi-user MISO systems under imperfect channel state information (CSI) at the BS. The energy efficiency problem for RSMA was initially studied in [126]. The work in [104] studied the energy efficiency of the RSMA and NOMA schemes in a mmW downlink transmission. In [127], the spectral and energy efficiencies were optimized using RSMA in non-orthogonal unicast and multicast transmissions. In all previous works,

Table 2.4: Literature review on the applications of RSMA

Application	Reference	Sum Rate	Max-Min Fair	Energy Efficiency
Massive MIMO	[102, 103, 106]	✔		
Cognitive D2D	[105, 118]	✔		✔
Multigroup multicast	[119, 120]	✔	✔	
millimeter-wave	[104, 121]	✔		✔
Cloud-Radio Access Networks	[109, 112–114]	✔		✔
Unmanned Aerial Vehicles	[108, 115, 116]	✔		
Visible light communication	[122, 123]	✔		
Intelligent Reconfigurable Surface	[110, 111]		✔	✔

RSMA offers potential spectral and energy efficiency gains over NOMA and OMA systems. Table 2.4 summarizes the applications that consider RSMA and the resource allocation methods used in them. However, all aforementioned works considered PGS as its signaling scheme. In this research, we aim to investigate the potential merit of using IGS to further improve the spectral efficiency of RSMA systems.

2.5.1 Fundamental Research Works in Histroy of RSMA

The idea of rate splitting (RS) was originally introduced in [88] for SISO interference channel where capacity region was determined. The Han and Kobayashi (HK) scheme was proposed in [89] as a further improvement for the work in [88]. Uplink RS was introduced in [90] for K-user SISO multiple access channel, and was proved to achieve the capacity region without time sharing. In [91], a simplified version of HK scheme was presented to achieve rates within 1 bit/s/Hz of the capacity of the SISO Gaussian interference channel. The multi-antenna RS scheme was introduced for the first time in [18] for multi-antenna networks, where it was named as 1-layer RS scheme. In [92],

this 1-layer RS scheme was proved to achieve the sum-DoF of the K-user MISO broadcast channel with imperfect channel state information (CSI). In [92], a precoding scheme for RS was optimized to achieve a better rate region than SDMA.

A generalized RSMA framework was introduced in [96] to compare the performance of RSMA with NOMA and DPC in multi-antenna networks. The work in [97] is the first work which proposed a low-complexity precoding and closed-form optimal power allocation schemes and compared RSMA, SDMA, and NOMA for a two-user MISO system. The design of physical layer signaling of RSMA was presented for the first time in [98] with different constellation modulation schemes and fixed length polar codes. Works provided in [98] used link-level simulator which showed that RSMA achieves considerable throughput gains over existing multiple access techniques. The work in [102] adopted RSMA in a practical 3GPP channel modeling and signaling to handle user mobility and hand-over in 5G wireless networks.

The work in [99] summarized the gains of RSMA over SDMA and NOMA in terms of DoF in both over-loaded and under-loaded multi-antenna networks with perfect and imperfect CSI. The potential merits of using RSMA to meet the 3GPP standard requirements of 6G was addressed in [100]. To discuss the challenges of ultra reliability and low-latency communications (URLLC) and enhanced mobile broadband (eMBB) core applications, work in [101] investigated RSMA in short-packet communications and outdated CSI scenarios.

2.6 Improper Gaussian Signaling

Proper Gaussian signals are CSCG signals with independent and equal variance on the real and imaginary components [23–25]. If the real and imaginary components of the complex Gaussian signal either have non-identical variance or are correlated, then the signals are considered improper [23, 25]. The most significant property of PGS is that it achieves the maximum throughput in an AWGN channel. However, this is not the case

in an interference channel [20–22].

Majority of the existing works assumed proper signal model for the underlying CSCG signals, which is in contrast with the many of the real-world models. The improper characteristics with non-identical variance on the real and imaginary components are known to exist in some well-known constellation diagrams such as continuous phase modulation [24]. In practical systems, improper signaling techniques are widely adopted in linear receivers for Global System for Mobile communication (GSM) and 3GPP networks [23].

Compared to the PGS scheme which assumes independent real and imaginary signal components with equal power, the IGS scheme loosens these constraints introducing a more general Gaussian signaling scheme [25]. By relaxing the constraints of PGS (i.e. considering a correlation and/or unequal powers between real and imaginary parts), IGS [24], [25], was shown to offer significant improvement in several limited-interference communication scenarios, e.g. NOMA systems under imperfect SIC. Optimizing the covariance and pseudo-covariance coefficients make it possible to improve the system performance by either maximizing the average achievable rate and achievable sum rate, minimizing the maximum achievable rate, and maximizing the energy efficiency.

2.6.1 Related Works-IGS

IGS has been considered for communication systems that are subject to interference in order to enhance the achievable throughput performance of two user SISO systems [128], K-user MISO systems [129], K-user MIMO systems [130], Z-interference channels [131], and MIMO systems with Z-interference channel [132]. IGS scheme is also employed in different interference scenarios such as full duplex relaying [133], alternate relaying [134] and underlay CR [38].

In CR systems, it was shown that the entropy loss due to IGS is the least compared with PGS, which makes it a suitable transmission scheme for the cognitive secondary user (SU) as the improper interference on the primary user (PU) is much less harmful compared

to proper interference [135]. Thus, IGS is the potential choice for SU to maximize its achievable rate while keeping the interference reflected on the PU limited.

Motivated by the promising merits of IGS in interference limited environments, wide range of contributions have investigated IGS benefits in FD/HD cooperative decode and-forward (DF) relaying systems by effectively compensating residual self-interference (RSI), inter-relay interference (IRI), and/or hardware impairments (HWI) [134, 136, 137]. IGS notability is also proven in multi-antenna scenario to mitigate different HWIs [138, 139]. The adoption of IGS yields up to ten percentage increase in average achievable rate of MIMO relative to PGS even in low SNR region [138].

Interestingly, IGS makes it possible to achieve the QoS constraints of wireless networks while consuming less power at the the BSs owing to the additional IGS degrees of freedom. Various energy-efficient solutions for multi-user, multi-antenna, and cooperative relaying scenarios are proposed [140, 141], where significant gain of energy efficiency performance is achieved. IGS is used to enhance the energy efficiency performance of one direction DF FD-MIMO relay [142] and bi-direction amplify-and-forward (AF) FD-MIMO relay [143] for multi-user interference systems. Although IGS demonstrates significant energy efficiency gains over PGS, the relative energy efficiency gains are less than the achievable rate gains [144]. A comprehensive literature review on the journey of IGS can be found in [24]. All the aforementioned contributions are in the OMA domain. In this research work, we investigate the performance merit of IGS in NOMA systems.

2.6.2 Practical Implementation Challenges of IGS

Signal Generation and Modulation: Generating IGS signals with the necessary correlation between real and imaginary components is complex and requires specialized hardware, such as precise digital-to-analog converters (DACs).

Synchronization and Calibration: Accurate timing synchronization and frequent calibration are essential to ensure proper signal reception and avoid performance degradation

due to hardware imperfections.

Non-Linearities and Hardware Imperfections: Non-linearities in hardware components, such as power amplifiers and mixers, can distort IGS signals, reducing their quality. Linearization techniques are required but add complexity.

Complexity of Receiver Design: IGS receivers need advanced signal processing algorithms to detect and decode the non-circular signals, which are computationally intensive and demand specialized hardware like FPGAs.

2.7 Preliminary: IGS Definition and Generation

2.7.1 Improper Random Variables

For a scalar complex RV x, we use φ_x and $\hat{\varphi}_x$ to denote the variance and pseudo-variance of x, respectively. For a zero-mean Gaussian RV x, we define $\varphi_x = \mathcal{E}[xx^*]$, which is a real value and equivalent to the power of the transmit signal, where the superscript $(\cdot)^*$ refers to the complex conjugation operation. We also define $\hat{\varphi}_x = \mathcal{E}[xx]$, which is typically a complex value. Here, $\mathcal{E}[\cdot]$ refers to the expectation of a random variable.

To illustrate the difference between IGS and PGS schemes, we introduce the following definitions [130].

Definition 1: A signal is called proper if it is uncorrelated with its complex conjugate and has a zero pseudo-variance, i.e., $\hat{\varphi}_x = 0$, otherwise it is called an improper signal.

Definition 2: The impropriety degree (or the IGS circularity coefficient) of x is given as

$$\xi_x \quad = \quad \frac{|\hat{\varphi}_x|}{\varphi_x}, \tag{2.14}$$

where $0 \leq \xi_x \leq 1$. If $\xi_x = 0$, then x is proper signal, and if $\xi_x = 1$, we have a maximally improper signal.

Definition 3: The achievable rate expression for a point-to-point IGS system is given as [21]

$$\mathcal{R}_i = \frac{1}{2}\log_2\left(\frac{\Omega_{y_i}^2 - |\hat{\Omega}_{y_i}|^2}{\Omega_{z_i}^2 - |\hat{\Omega}_{z_i}|^2}\right), \tag{2.15}$$

where Ω_{y_i} and $\hat{\Omega}_{y_i}$ refer respectively to the covariance and pseudo-covariance components of the received signal and Ω_{z_i} and $\hat{\Omega}_{z_i}$ refer to covariance and pseudo covariance components of interference plus noise signal, respectively.

In the case of vector complex RV $\mathbf{x}$, we have the following definitions:

Definition 4: A complex RV vector $\mathbf{x}$ is called proper if its pseudo-covariance matrix $\hat{\mathbf{C}}_{\mathbf{x}}$ vanishes to a zero matrix; otherwise it is called improper [130].
Define $\mathbf{J}_{\mathbf{x}}$ as the covariance matrix of the augmented vector $[\mathbf{x}^T \ \ \mathbf{x}^{*T}]^T$ as

$$\mathbf{J}_{\mathbf{x}} = \mathcal{E}\left(\begin{bmatrix}\mathbf{x}\\ \mathbf{x}^*\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{x}^*\end{bmatrix}^H\right) = \begin{bmatrix}\mathbf{C}_{\mathbf{x}} & \hat{\mathbf{C}}_{\mathbf{x}}\\ \hat{\mathbf{C}}_{\mathbf{x}}^* & \mathbf{C}_{\mathbf{x}}^*\end{bmatrix}.$$

Definition 5: $\mathbf{C}_{\mathbf{x}}$ and $\hat{\mathbf{C}}_{\mathbf{x}}$ are a valid pair of covariance and pseudo-covariance matrices, i.e., there exists a RV $\mathbf{x}$ with covariance and pseudo-covariance matrices given by $\mathbf{C}_{\mathbf{x}}$ and $\hat{\mathbf{C}}_{\mathbf{x}}$ respectively, if and only if, the augmented covariance matrix $\mathbf{J}_{\mathbf{x}}$ is positive semidefinite [25].

Definition 6: The differential entropy of a complex Gaussian RV $\mathbf{x}$ with augmented covariance matrix $\mathbf{J}_{\mathbf{x}}$ is given by [25]

$$I(\mathbf{x}) = \frac{1}{2}\log_2[(\pi e)^2|\mathbf{J}_{\mathbf{x}}|], \tag{2.16}$$

where $|\cdot|$ is the determinant of the matrix. Using (2.16), achievable user rates can be computed. For more details about improper RVs, we refer the reader to [25].

2.7.2 Improper Constellation Design

Widely linear transformation (WLT) is considered one of the most popular ways of transforming proper signals to improper ones. A simple design involving the generation of improper discrete constellations was introduced in [145, 146] in which a WLT of a standard unit energy proper M-QAM (quadrature amplitude modulation) constellation was used. In particular, an improper constellation design, with a complex symbol x and predefined circularity coefficient ξ_x, can be obtained from a proper constellation design with complex symbol s with the following unit energy:

$$x = \epsilon s + \varepsilon s^*, \tag{2.17}$$

where ϵ and ε are complex-valued quantities such that $\epsilon = \sqrt{0.5(1+\sqrt{1-\xi_x^2})}$ and $\varepsilon = \sqrt{0.5(1-\sqrt{1-\xi_x^2})}e^{j\phi}$, $|\epsilon|^2 + |\varepsilon|^2 = 1$, $\phi \in [0, \pi/2]$. An optimal ϕ is selected to maximize the minimum distance between constellation points. For $0 \leq \xi_x \leq 0.5$, the optimal ϕ for an arbitrary M is found to be $\phi = \pi/2$. On the other hand, when $0.5 \leq \xi_x \leq 2/\sqrt{5}$, the optimal ϕ is found by solving the nonlinear equation $\sin(\phi) - \cos(\phi) = 1/\xi_x$, which is equal to $\phi = \arcsin(\frac{\sqrt{5}}{5\xi_x}) + \arctan\frac{1}{2}$. Note that the optimal ϕ for the region of $2/\sqrt{5} \leq \xi_x \leq 1$ is shown to be the same as the one in the solution for $0.5 \leq \xi_x \leq 2/\sqrt{5}$ at high signal-to-noise ratio [146]. Fig. 4.2 shows the improper constellations with 16-QAM that are designed based on WLT. It can be seen that, as ξ_i increases, the minimum Euclidean distance between the constellations points is changed among the improper constellations. Due to this impropriety, the throughput enhancement is achieved as proven in this work.

2.8 Resource Allocation and Convex Optimization Approaches

Resource allocation plays an important role to improve the spectral and energy efficiency in wireless communication systems. The most valuable resources in wireless communications are frequency, time, and power optimization. There are various mechanisms for

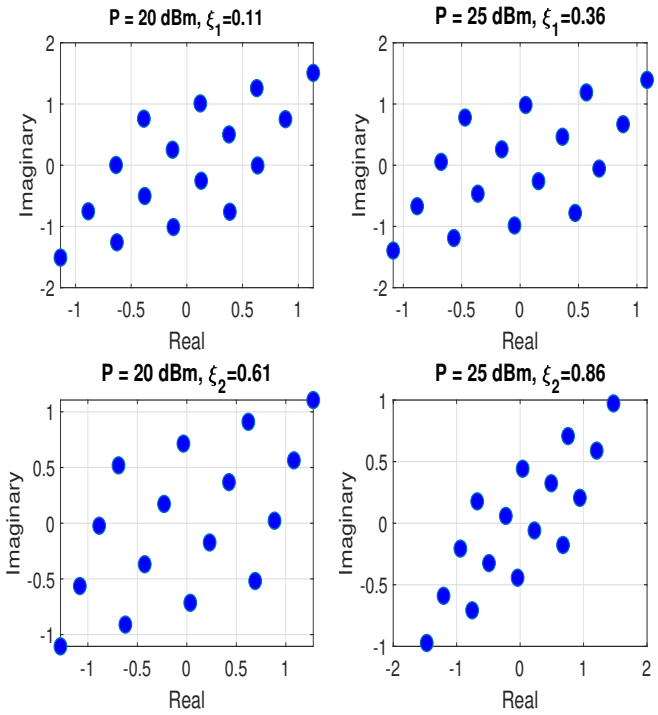

Figure 2.8: 16-QAM improper constellation diagrams with different ξ_x.

resource allocation in wireless systems including but not limited to power control, sub-channel allocation, and user rate allocation etc. In this thesis, we focus on power allocation and channel circularity coefficient optimization to maximize the system spectral efficiency and energy efficiency in NOMA and RSMA networks. Power allocation is performed at the the BS, where scheduler needs to allocate different powers to different users to guarantee certain performance gains. The optimization of channel circularity coefficient is done at the BS where the degree of signaling impropriety is decided to achieve better user throughput performance.

One of the most widely used and effective mathematical method to solve the resource allocation problems in wireless communication networks is the convex optimization tool. For an optimization problem to be convex, the objective function to be minimized or maximized and the inequality constraint functions should be convex or concave, and the equality constraint functions should be affine. However, not all resource allocation problem are convex. Many of them are non-convex and NP hard. Generally, it is easy

to find global optimal solution for the convex optimization problems but it is rather difficult to find the optimal solution for non-convex optimization problems. In non-convex problems, there may exist many locally optimal solutions which are not globally optimal. In addition, it is analytically and practically difficult to check whether a given local optimal solution is globally optimal, and this prevents the development of efficient solution methods for such problems [152].

2.8.1 The Lagrangian Method

The Lagrangian multiplier method is a mechanism for obtaining the local maxima and minima of a multi-variable function $f(x, y)$, subject to some equality constraints $g(x, y) = c$, where c is a constant. The general Lagrangian optimization can be written as

$$\underset{x,y}{\text{optimize}}\ f(x, y) \tag{2.18a}$$

$$\text{s.t. } g(x, y) = c. \tag{2.18b}$$

The Lagrangian method can be briefly introduced as follows. To find the maximum or minimum of a function $f(x, y)$ subjected to the equality constraint $g(x, y) = 0$,

- Introduce a new factor λ, and define a new function $\mathcal{L}$ as follows: $\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda(g(x, y) - c)$, where the function $\mathcal{L}$ is called the Lagrangian function and λ is the Lagrangian multiplier.

- To find the possible solutions and compute the stationary points of the Lagrange function, the gradient of $\mathcal{L}$ is set to equal to zero, $\nabla\mathcal{L}(x, y, \lambda) = 0$.

- All valid solutions are substituted in the objective function and the one that results in the greatest (or smallest) value is the local maxima (or minima) we are looking for. Note that the Lagrangian method does not guarantee global optimal solution.

2.8.2 KKT Conditions

The KKT conditions are called first-order necessary conditions and defined as first derivative tests for a possible solution in non-linear programming to be optimal, subject to some regularity conditions are satisfied. KKT approach is considered a generalization for the Lagrangian method in solving non-linear programming since it allows inequality constraints as opposed to Lagrangian methods that allows only equality constraints.

$$\underset{x,y}{\text{optimize}}\ f(x,y) \tag{2.19a}$$

$$\text{s.t. } g_i(x,y) = 0, i = 1, ..., m \tag{2.19b}$$

$$h_j(x,y) \leq 0, j = 1, ..., l \tag{2.19c}$$

That said, the constrained minimization (maximization) problem is re-formulated as a Lagrange function whose optimal point is a saddle point that represents the optimal maxima or minima over the the domain of the choice variables.

If x^*, y^* are local optimum points and the optimization problem satisfies some regularity conditions, then there exist constants, λ_i and μ_j, called KKT multipliers, such that the following necessary conditions hold.

- Stationarity

$$\text{Minimization}: \nabla f(x^*,y^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*,y^*) + \sum_{j=1}^{l} \mu_j \nabla h_j(x^*,y^*) = 0, \tag{2.20}$$

$$\text{Maximization}: -\nabla f(x^*,y^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*,y^*) + \sum_{j=1}^{l} \mu_j \nabla h_j(x^*,y^*) = 0, \tag{2.21}$$

- Primal feasibility

$$g_i(x^*,y^*) = 0, \forall i \tag{2.22}$$

$$h_j(x^*,y^*) \leq 0, \forall j \tag{2.23}$$

- Dual feasibility

$$\mu_j \geq 0, \forall j \tag{2.24}$$

- Complementary slackness

$$\sum_{j=1}^{l} \mu_j h_j(x^*, y^*) = 0. \tag{2.25}$$

The number of equations and inequalities corresponding to the KKT conditions is usually not easy to solve simultaneously, except in a few special cases where a closed-form solution can be analytically found. In general, many iterative algorithms can be developed as methods for numerically solving the KKT equations and inequalities.

2.8.3 Dinkelbach Algorithm

Dinkelbach algorithm was introduced in [181]. It is classified as a one category of parametric algorithms, whose basic principle is to handle a concave-convex fraction problems (CCFP) by solving a sequence of less complex problems. To explain its idea, let's consider the following CCFP:

$$\max_x \frac{f(x)}{g(x)} \tag{2.26a}$$

$$\text{s.t. } c_i(x) \leq 0, \forall i, \tag{2.26b}$$

where f is concave, differential, and non-negative function, g is convex, differential, and positive, and c_i are convex for all i. The fundamental result upon which Dinkelbach's algorithm is based is the relation between the CCFP (2.26) and the parametric function of a real variable as follows

$$F(\lambda) = \max_{x \in c_i} \{f(x) - \lambda g(x)\}. \tag{2.27}$$

Dinkelbach Algorithm

1: $\epsilon > 0, n = 0, \lambda_n = 0$;
2: **while** $F(\lambda_n) > \epsilon$ **do**
3: $x_n^* = \arg \max_{x \in c_i} \{f(x) - \lambda_n g(x)\}$;
4: $F(\lambda_n) = f(x_n^*) - \lambda_n g(x_n^*)$;
5: $\lambda_{n+1} = \frac{f(x_n^*)}{g(x_n^*)}$;
6: $n = n + 1$;
7: **end while**

Proposition: Consider $x^* \in c_i$ and $\lambda^* = \frac{f(x^*)}{g(x^*)}$, then x^* is a solution of (2.26) if and only if [181]

$$x^* = \arg \max_{x \in c_i} \{f(x) - \lambda^* g(x)\} \tag{2.28}$$

Based on the this above-mentioned proposition, solving a fractional problem is equivalent to finding the unique zero of the auxiliary function $F(.)$. A pseudo code of how the Dinkelbach algorithm is accomplished is presented as shown in the top of the page.

2.9 Chapter Summary

In this chapter, the background and recent literature of power-domain NOMA systems were reviewed with a focus on the following aspects: the basic principles of NOMA, the resource allocation of NOMA, its emerging with other key 5G technologies, and the implementation challenges and standardization. RSMA concept and definition were also introduced. Furthermore, the advantages of RSMA over NOMA and OMA were listed. The rates of RSMA were also derived for simple SISO systems.

Preliminary introduction for IGS was also introduced. Background and literature review were also presented. Bearing in mind that NOMA superimposed users signals associated with different power levels with the aid of superposition coding techniques, the resource allocation, e.g., power sharing among the users should be carefully optimized for each practical scenario.

In addition, NOMA and RSMA systems should be tackled considering practical sce-

narios including imperfect SIC and imperfect self-interference cancellation etc. so that the potential capability of NOMA and RSMA are attained, which is still lack in the literature.

Chapter 3

Spectral and Energy Efficiency Maximization of NOMA Systems under Imperfect Successive Interference Cancellation

3.1 Introduction

PGS has been widely adopted in communication systems due to its attractive entropy-maximizing property, which is proved to achieve the maximum spectral efficiency in AWGN channels but not in the case of an interference channel [20–22]. Moreover, the majority of the wireless products assume a proper signal model for the underlying CSCG signals, which is in contrast with many of the real-world applications. The improper characteristics with non-identical variance on the real and imaginary components are known to exist in some well-known constellation diagrams such as binary phase-shift keying and continuous phase modulation [24].

The traditional complex-valued signal processing assumes a vanishing pseudo-covariance,

i.e., equals zero, which is not accurate for many real-world scenarios where the concerned signals are frequently improper [24]. Such simplified assumption not only results in misleading analysis and inaccurate results but also prohibits us from investigating the potential benefit of the additional degree of the freedom offered by improper signaling. By relaxing the constraints of PGS, i.e., equal power and uncorrelated real and imaginary components, we have what is called IGS [24], [25]. In practical systems, improper signaling techniques have already existed in linear receivers for GSM communication and 3GPP systems [23]. It has been shown that IGS attains higher degrees of freedom than PGS due to its ability to identify the interference signal dimension. In fact, it can be described as a kind of interference alignment technique since it mitigates interference by using its alignment in only one orthogonal signal space dimension, and by extracting the desired signal from the other orthogonal space [27]. Due to these impropriety characteristics, IGS has been identified as a potential candidate for improving the spectral efficiency in interference-limited systems [24], [25].

NOMA has been introduced as a potential candidate for enabling heavily-loaded systems using limited resources while improving the spectral efficiency [12]. In that sense, in the downlink of NOMA systems, multiple users can be served simultaneously on one resource block by implementing superposition coding at the BS side and SIC at the users' side [15]. This leads to spectral-efficient systems provided that perfect SIC is performed, which is not a realistic assumption. In practice, detectors suffer from imperfect SIC, which leads to an interference-limited NOMA system [31]. This encourages us to consider IGS as a promising solution to recompense for the SIC imperfections in such interference-limited scenarios.

3.2 Related Works

In OMA schemes, IGS has been investigated in interference Z-channel models [20–22], MIMO interference-limited systems [147], and broadcast channels [169], [170], where the

achievable throughout regions and degrees of freedom were derived. IGS was also considered in Z-channel CR networks, where IGS was proven to be beneficial in reducing the interference to primary users in various unlicensed spectrum-sharing models [173]. IGS was also evaluated in cooperative full-duplex relaying CR systems [174], where IGS was used to help the secondary user access the spectrum. Similar work in [175] was investigated in non-CR systems under Nakagami-m fading, where it was concluded that using IGS can eliminate the impact of residual self-interference by revising the signal impropriety.

Recent works considered IGS in NOMA interference-limited systems. The authors in [176] investigated the performance of a two-user downlink NOMA system using IGS, by deriving the outage probability and the ergodic capacity expressions. In [177], transmit precoding schemes were designed for a multi-cell network in order to maximize the users' minimum rates under various power budget constraints. In [178], IGS scheme was generated for signal beamforming with the goal of improving the spectral efficiency of a multi-cell network and protecting the users' secrecy.

3.3 Contributions

The main contributions of this chapter can be summarized as follows.

- This chapter studies a two-user downlink point-to-point NOMA system with both users employing IGS under imperfect SIC. We first derive the exact expressions for each user's rate when using IGS at both users.

- We formulate two optimization problems for spectral efficiency and energy efficiency. In each formulated problem, we jointly optimize the BS transmit power and the transmit signal's circularity coefficients given the QoS and power budget requirements.

- We also consider a special case of using IGS at the strong user only. In this case, a

sum-rate maximization problem is formulated where the IGS circularity coefficient is optimized. Hence, we derive closed-form expression for the optimized IGS circularity coefficient.

- We propose iterative algorithms to find sub-optimal solutions to the developed non-convex optimization problems based on the KKT conditions.

- Additionally, we show the efficacy of the optimized circularity coefficient on the transmit constellations at the BS and the system error performance, where improper constellation diagrams are designed using WLT based on the optimized IGS circularity coefficients.

- Simulation results demonstrate the vital impact of IGS over PGS in the context of NOMA systems under imperfect SIC. Results show that the performance of IGS-based NOMA system outperforms its counterpart PGS-based NOMA system under SIC imperfections.

3.4 Channel and System Models

3.4.1 Channel Model

A downlink two-user NOMA system model with a BS is considered in this work as shown in Fig. 3.1. The Rayleigh model is considered for the channels between the BS and users. We denote h_i, $\forall i = 1, 2$, as the channel variables with variance $\sigma^2_{h_i}$. Unlike the commonly used setups where users employ PGS, in this work we consider IGS at both users, i.e., x_i, $\forall i = 1, 2$, are improper signals.

As per power-domain NOMA basics, users experience different channel conditions with the assumption of $|h_1|^2 > |h_2|^2$ and $P_2 > P_1$. Consequently, user 1, which has a better channel condition, can decode its own signal after removing the signal of user 2, which has worse channel condition, through SIC processing [14]. Meanwhile, user 2 can decode

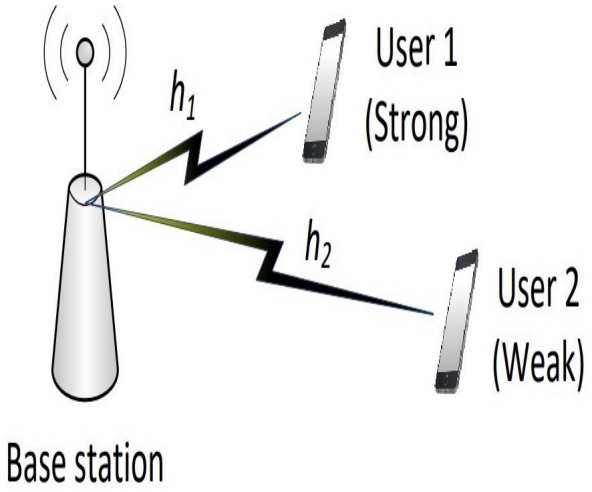

Figure 3.1: A NOMA system with IGS.

its own signal assuming the interference inflicted by user 1 is negligible and can thus be dealt with as noise.

Since user 1's receiver has imperfect SIC, there exists a residual interference as an outcome from this imperfection. Generally, the residual interference resulting from imperfect SIC is a complicated function of multiple factors, e.g., coding/modulation related parameters, channel related issues (fading and shadowing), device/hardware/battery related restrictions, etc. Furthermore, due to the characteristics of error propagation due to imperfect SIC, it is hard to model its impact. It is stated in [171] and [172] that a linear model can effectively demonstrate the relationship between the residual interference and the received signal power.

To introduce the impact of imperfect SIC, we adopt this linear model here. Under such a model of imperfect SIC, the received signal after SIC at user 1 is given in (3.1), where η denotes the coefficient of imperfect SIC at user 1. Note that $\eta = 0$, corresponds to perfect SIC, and $\eta = 1$, corresponds to no interference cancellation.

3.4.2 Spectral Efficiency Analysis

Since IGS is considered at both users, i.e., x_1 and x_2 are improper, and imperfect SIC is assumed, we first need to find the new expressions for the users' information rates. Hence, closed-form expressions are obtained for the spectral efficiency for each user in the considered system model.

To do so, we characterize the received signals at each user as

$$y_1 = \sqrt{P_1}h_1x_1 + \eta\sqrt{P_2}h_1x_2 + n_1, \tag{3.1}$$

$$y_2 = \sqrt{P_2}h_2x_2 + \sqrt{P_1}h_2x_1 + n_2, \tag{3.2}$$

where x_i, $\forall i = 1, 2$, is the signal for the ith user, $\mathcal{E}[|x_i|^2] = 1$, and n_i is additive white Gaussian noise (AWGN) at the ith user's receiver with variance σ_i^2. The improper circularity coefficient of x_i is denoted as $\kappa_{x_i} = \frac{|\hat{C}_i|}{C_i}$, where $\hat{C}_i = \mathcal{E}\{x_ix_i\} = \kappa_{x_i}\angle\hat{C}_i$ and $C_i = \mathcal{E}\{x_ix_i^*\}$. The component $\eta\sqrt{P_2}h_1x_2$ in (3.1) refers to the imperfect SIC at user 1. To continue, based on [154], the user's rate definition in the case of IGS is given as

$$\mathcal{R}_i = \frac{1}{2}\log_2\left(\frac{\Omega_{y_i}^2 - |\hat{\Omega}_{y_i}|^2}{\Omega_{z_i}^2 - |\hat{\Omega}_{z_i}|^2}\right), \tag{3.3}$$

where Ω_{y_i} and $\hat{\Omega}_{y_i}$ refer respectively to the covariance and pseudo-covariance components of the received signal and Ω_{z_i} and $\hat{\Omega}_{z_i}$ refer to covariance and pseudo covariance components of interference plus noise signal, respectively. The covariance and pseudo-covariance of y_i, $i = 1, 2$, can be obtained from (3.1) and (3.2) as

$$\Omega_{y_1} = \mathcal{E}(y_1y_1^*) = P_1|h_1|^2C_1 + \eta^2P_2|h_1|^2C_2 + \sigma_1^2, \tag{3.4}$$

$$\hat{\Omega}_{y_1} = \mathcal{E}(y_1y_1) = P_1h_1^2\hat{C}_1 + \eta^2P_2h_1^2\hat{C}_2, \tag{3.5}$$

$$\Omega_{y_2} = \mathcal{E}(y_2y_2^*) = P_2|h_2|^2C_2 + P_1|h_2|^2C_1 + \sigma_2^2, \tag{3.6}$$

$$\hat{\Omega}_{y_2} = \mathcal{E}(y_2y_2) = P_2h_2^2\hat{C}_2 + P_1h_2^2\hat{C}_1. \tag{3.7}$$

Define the noise and the interference-plus-noise terms in (3.1) and (3.2), as z_i, $i = 1, 2$, at each receiver, respectively, where $z_1 = \eta\sqrt{P_2}h_1x_2 + n_1$ and $z_2 = \sqrt{P_1}h_2x_1 + n_2$, we get

$$\Omega_{z_1} = \mathcal{E}(z_1 z_1^*) = \eta^2 P_2 |h_1|^2 C_2 + \sigma_1^2, \tag{3.8}$$

$$\hat{\Omega}_{z_1} = \mathcal{E}(z_1 z_1) = \eta^2 P_2 |h_1|^2 \hat{C}_2, \tag{3.9}$$

$$\Omega_{z_2} = \mathcal{E}(z_2 z_2^*) = P_1 |h_2|^2 C_1 + \sigma_2^2, \tag{3.10}$$

$$\hat{\Omega}_{z_2} = \mathcal{E}(z_2 z_2) = P_1 h_2^2 \hat{C}_1, \tag{3.11}$$

To continue, we compute the following components.

$$\begin{aligned} |\hat{\Omega}_{y_1}|^2 &= |P_1 h_1^2 \hat{C}_1 + \eta^2 P_2 h_1^2 \hat{C}_2|^2 \\ &= (P_1|h_1|^2\kappa_{x_1})^2 + (\eta^2 P_2|h_1|^2\kappa_{x_2})^2 + 2(P_1|h_1|^2\kappa_{x_1})(\eta^2 P_2 h_1^2 \kappa_{x_2}) \cos(\angle\hat{C}_1 - \angle\hat{C}_2), \end{aligned} \tag{3.12}$$

$$\begin{aligned} |\hat{\Omega}_{y_2}|^2 &= |P_2 h_2^2 \hat{C}_2 + P_1 h_2^2 \hat{C}_1|^2 \\ &= (P_2|h_2|^2\kappa_{x_2})^2 + (P_1|h_2|^2\kappa_{x_1})^2 + 2(P_2|h_2|^2\kappa_{x_2})(P_1|h_2|^2\kappa_{x_1}) \cos(\angle\hat{C}_1 - \angle\hat{C}_2), \end{aligned} \tag{3.13}$$

$$|\hat{\Omega}_{z_1}|^2 = (\eta^2 P_2|h_1|^2\kappa_{x_2})^2, \tag{3.14}$$

$$|\hat{\Omega}_{z_2}|^2 = (P_1|h_2|^2\kappa_{x_1})^2, \tag{3.15}$$

By substituting (3.4), (3.6), (3.8), (3.10), (3.12)-(3.15) into (3.3), and assuming, without loss of generality, $C_1 = C_2 = 1$, and $\sigma_1^2 = \sigma_2^2 = \sigma^2$, after a few mathematical manipulations

$$\mathcal{R}_1 = \frac{1}{2}\ \log_2\left(\frac{\left(P_1|h_1|^2 + \eta^2 P_2|h_1|^2 + \sigma^2\right)^2 - tz}{\left(\eta^2 P_2|h_1|^2 + \sigma^2\right)^2 - \left(\eta^2 P_2|h_1|^2\kappa_{x_2}\right)^2}\right), \tag{3.16}$$

where $tz = \Big((P_1|h_1|^2\kappa_{x_1})^2 + (\eta^2 P_2|h_1|^2\kappa_{x_2})^2 + 2(P_1|h_1|^2\kappa_{x_1})(\eta^2 P_2|h_1|^2\kappa_{x_2}) \cos(\angle\hat{C}_1 - \angle\hat{C}_2)\Big)$.

and

$$\mathcal{R}_2 = \frac{1}{2}\log_2\left(\frac{(P_2|h_2|^2 + P_1|h_2|^2 + \sigma^2)^2 - ty}{(P_1|h_2|^2 + \sigma^2)^2 - (P_1|h_2|^2\kappa_{x_1})^2}\right), \tag{3.17}$$

where $ty = \Big((P_2|h_2|^2\kappa_{x_2})^2 + (P_1|h_2|^2\kappa_{x_1})^2 + 2(P_2|h_2|^2\kappa_{x_2})(P_1|h_2|^2\kappa_{x_1})\cos(\angle\hat{C}_1 - \angle\hat{C}_2)\Big)$. Let's define $\Delta = \cos(\angle\hat{C}_1 - \angle\hat{C}_2)$. To find the optimal Δ that maximizes the sum rate $\mathcal{R}_1 + \mathcal{R}_2$, we derive each rate with respect to Δ as follows.

$$\frac{\partial\mathcal{R}_1}{\partial\Delta} = -\frac{1}{\ln 2}\frac{(P_1|h_1|^2\kappa_{x_1})(\eta^2 P_2|h_1|^2\kappa_{x_2})}{t_1}, \tag{3.18}$$

where $t_1 = (P_1|h_1|^2 + \eta^2 P_2|h_1|^2 + \sigma^2)^2 - tz =$
$(P_1|h_1|^2)^2(1-\kappa_{x_1}^2) + (\eta^2 P_2|h_1|^2)^2(1-\kappa_{x_2}^2) + \sigma^2 + 2(P_1|h_1|^2)(\eta^2 P_2|h_1|^2)(1-\kappa_{x_1}\kappa_{x_2}\cos(\angle\hat{C}_1 - \angle\hat{C}_2)) + 2\sigma^2(P_1|h_1|^2)(\eta^2 P_2|h_1|^2)$, and

$$\frac{\partial\mathcal{R}_2}{\partial\Delta} = -\frac{1}{\ln 2}\frac{(P_1|h_2|^2\kappa_{x_1})(P_2|h_2|^2\kappa_{x_2})}{t_2}, \tag{3.19}$$

where $t_2 = (P_2|h_2|^2 + P_1|h_2|^2 + \sigma^2)^2 - ty =$
$(\mathrm{P}_1|h_2|^2)^2(1-\kappa_{x_1}^2) + (P_2|h_2|^2)^2(1-\kappa_{x_2}^2) + \sigma^2 + 2(P_1|h_2|^2)(P_2|h_2|^2)(1-\kappa_{x_1}\kappa_{x_2}\cos(\angle\hat{C}_1 - \angle\hat{C}_2)) + 2\sigma^2(P_1|h_2|^2)(P_2|h_2|^2)$. Based on (3.18), (3.19), the derivation of the sum rate with respect to Δ results in

$$\frac{\partial(\mathcal{R}_1 + \mathcal{R}_2)}{\partial\Delta} = -\frac{1}{\ln 2}\frac{(P_1\kappa_{x_1}P_2\kappa_{x_2})}{t_1 t_2}\left[\eta^2|h_1|^4 t_2 + |h_2|^4 t_1\right], \tag{3.20}$$

Since $0 \le \kappa_{x_1}, \kappa_{x_2} \le 1$ and $-1 \le \cos(\angle\hat{C}_1 - \angle\hat{C}_2)) \le 1$, all the terms between brackets in (3.20) are positive values and the overall value of (3.20) is negative. Hence, the sum rate function is a decreasing function of Δ. Then, the optimal Δ to maximize the sum rate is to choose $\Delta = -1$, i.e., $\angle\hat{C}_1 - \angle\hat{C}_2 = \pi$.

By substituting the value of $\cos(\angle\hat{C}_1 - \angle\hat{C}_2) = -1$ into (3.16) and (3.17) and assuming the powers $P_1 = \alpha_1 P_T$ and $P_2 = \alpha_2 P_T$ are assigned to user 1 and user 2, respectively,

where $0 \leq \alpha_i \leq 1$, and $\alpha_1+\alpha_2 = 1$, and using the math identity $A^2+B^2-2AB = (A-B)^2$, the user rates $\mathcal{R}_1, \mathcal{R}_2$, are derived in closed-form as

$$\mathcal{R}_1 = \frac{1}{2}\log_2\left(\frac{\left(\alpha_1 P_T|h_1|^2 + \eta^2\alpha_2 P_T|h_1|^2 + \sigma^2\right)^2 - \left((\alpha_1 P_T|h_1|^2\kappa_{x_1}) - (\eta^2\alpha_2 P_T|h_1|^2\kappa_{x_2})\right)^2}{\left(\eta^2\alpha_2 P_T|h_1|^2 + \sigma^2\right)^2 - \left(\eta^2\alpha_2 P_T|h_1|^2\kappa_{x_2}\right)^2}\right), \tag{3.21}$$

and

$$\mathcal{R}_2 = \frac{1}{2}\ \log_2\left(\frac{\left(\alpha_2 P_T|h_2|^2 + \alpha_1 P_T|h_2|^2 + \sigma^2\right)^2 - \left((\alpha_2 P_T|h_2|^2\kappa_{x_2}) - (\alpha_1 P_T|h_2|^2\kappa_{x_1})\right)^2}{(\alpha_1 P_T|h_2|^2 + \sigma^2)^2 - (\alpha_1 P_T|h_2|^2\kappa_{x_1})^2}\right). \tag{3.22}$$

3.5 Optimization Problems

In this section, two optimization problems are formulated to maximize both the overall spectral efficiency and the energy efficiency under the constraints of each user's minimum rate requirements and power budget.

3.5.1 Spectral Efficiency Maximization

Maximizing the overall spectral efficiency is a common objective in wireless systems. However, to avoid having all resources taken by one user, the quality of service (QoS) constraints are often enforced when formulating the overall spectral efficiency maximization. A joint optimization problem is developed to optimize both the IGS circularity coefficients $(\kappa_{x_1}, \kappa_{x_2})$ and power allocation at the BS. The objective is to maximize the system spectral efficiency given that the QoS of each user (minimum user rate) is met and the power budget is not exceeded.

In this optimization problem, we assume that the BS uses the total available power. For simplicity, we denote $\alpha_1 = \alpha$ and $\alpha_2 = 1 - \alpha$. Based on the analysis of the rate expressions in Section 3.4.2, we can improve the overall sum rate in the case both users

are employing IGS. That mentioned, the joint optimization problem can be developed as

$$\mathcal{OP}1: \underset{\kappa_{x_1},\kappa_{x_2},\alpha}{\text{maximize}} \quad \mathcal{R}_1(\kappa_{x_1},\kappa_{x_2},\alpha)+\mathcal{R}_2(\kappa_{x_1},\kappa_{x_2},\alpha) \tag{3.23a}$$

$$\text{subject to} \quad \mathcal{C}1:\ \mathcal{R}_1(\kappa_{x_1},\kappa_{x_2},\alpha)\geq R_{m_1}, \tag{3.23b}$$

$$\mathcal{C}2:\ \mathcal{R}_2(\kappa_{x_1},\kappa_{x_2},\alpha)\geq R_{m_2}, \tag{3.23c}$$

$$\mathcal{C}3:\ 0\leq \kappa_{x_1},\kappa_{x_2}\leq 1, \tag{3.23d}$$

$$\mathcal{C}4:\ 0\leq \alpha\leq 1, \tag{3.23e}$$

where $\mathcal{R}_1(\kappa_{x_1},\kappa_{x_2},\alpha)$ and $\mathcal{R}_2(\kappa_{x_1},\kappa_{x_2},\alpha)$ are computed from (3.21) and (3.22), respectively. R_{m_1} and R_{m_2} are each user's minimum rate requirement. The conditions $\mathcal{C}1$ and $\mathcal{C}2$ stress that the achieved rate of user 1 and user 2 are greater than R_{m_1} and R_{m_2}, respectively, the condition $\mathcal{C}3$ enforced the range of IGS circularity coefficients between 0 and 1, and the condition $\mathcal{C}4$ represents the BS power budget.

The objective and rate constraints in the formulated optimization (3.23a)-(3.23e) are non-convex which lead to a non-convex problem. To tackle this issue, the formulated problem can be optimized by using the necessary but not sufficient Karush-Kuhn-Tucker (KKT) conditions to find sub-optimal solutions for the circularity coefficient $\kappa^*_{x_1},\kappa^*_{x_2}$, and power allocation parameter α^* at less computational complexity.

The Lagrangian function corresponding to (3.23a)-(3.23e) can be outlined as

$$\begin{aligned}\mathcal{L}(\kappa_{x_1},\kappa_{x_2},\alpha) = &-\left(\mathcal{R}_1(\kappa_{x_1},\kappa_{x_2},\alpha)+\mathcal{R}_2(\kappa_{x_1},\kappa_{x_2},\alpha)\right)\\ &+\lambda_1(R_{m_1}-\mathcal{R}_1(\kappa_{x_1},\kappa_{x_2},\alpha))+\lambda_2(R_{m_2}-\mathcal{R}_2(\kappa_{x_1},\kappa_{x_2},\alpha)),\end{aligned} \tag{3.24}$$

where $\lambda_1\geq 0$ and $\lambda_2\geq 0$ are the Lagrange multipliers associated with the minimum rates constraints of user 1 and user 2, respectively. Based on the above, the KKT conditions

can be described as follows [180]

$$\frac{\partial \mathcal{L}(\kappa_{x_1}, \kappa_{x_2}, \alpha)}{\partial \kappa_{x_1}} = 0, \tag{3.25a}$$

$$\frac{\partial \mathcal{L}(\kappa_{x_1}, \kappa_{x_2}, \alpha)}{\partial \kappa_{x_2}} = 0, \tag{3.25b}$$

$$\frac{\partial \mathcal{L}(\kappa_{x_1}, \kappa_{x_2}, \alpha)}{\partial \alpha} = 0, \tag{3.25c}$$

$$\lambda_1(R_{m_1} - \mathcal{R}_1(\kappa_{x_1}, \kappa_{x_2}, \alpha)) = 0, \tag{3.25d}$$

$$\lambda_2(R_{m_2} - \mathcal{R}_2(\kappa_{x_1}, \kappa_{x_2}, \alpha)) = 0, \tag{3.25e}$$

$$R_{m_1} - \mathcal{R}_1(\kappa_{x_1}, \kappa_{x_2}, \alpha) \leq 0, \tag{3.25f}$$

$$R_{m_2} - \mathcal{R}_2(\kappa_{x_1}, \kappa_{x_2}, \alpha) \leq 0, \tag{3.25g}$$

$$\lambda_1, \lambda_2 \geq 0. \tag{3.25h}$$

The results of (3.25a), (3.25b), and (3.25c) are found, respectively, as

$$\begin{aligned}
&-(1-\lambda_1)(\alpha P_T|h_1|^2)(\alpha P_T|h_1|^2\kappa_{x_1} - \eta^2(1-\alpha)P_T|h_1|^2\kappa_{x_2})u_2\, d_2 \\
&+(1-\lambda_2)[-(\alpha P_T|h_2|^2)(\alpha P_T|h_2|^2\kappa_{x_1} - (1-\alpha)P_T|h_2|^2\kappa_{x_2})d_2 + (\alpha P_T|h_2|^2)^2\kappa_{x_1}u_2]u_1 = 0,
\end{aligned} \tag{3.26}$$

$$\begin{aligned}
&(1-\lambda_1)[(\eta^2(1-\alpha)P_T|h_1|^2)(\alpha P_T|h_1|^2\kappa_{x_1} - \eta^2(1-\alpha)P_T|h_1|^2\kappa_{x_2})d_1 + (\eta^2(1-\alpha)P_T|h_1|^2)^2 \\
&\kappa_{x_2}u_1]u_2 + (1-\lambda_2)[((1-\alpha)P_T|h_2|^2)(\alpha P_T|h_2|^2\kappa_{x_1} - (1-\alpha)P_T|h_2|^2\kappa_{x_2})]u_1\, d_1 = 0,
\end{aligned} \tag{3.27}$$

$$(1-\lambda_1)u_3\ u_2\ d_2 + (1-\lambda_2)d_3\ u_1\ d_1 = 0, \tag{3.28}$$

where $u_1 = (\alpha P_T|h_1|^2 + \eta^2(1-\alpha)P_T|h_1|^2 + \sigma^2)^2 - ((\alpha P_T|h_1|^2\kappa_{x_1}) - (\eta^2(1-\alpha)P_T|h_1|^2\kappa_{x_2}))^2$,
$d_1 = (\eta^2(1-\alpha)P_T|h_1|^2 + \sigma^2)^2 - (\eta^2(1-\alpha)P_T|h_1|^2\kappa_{x_2})^2$,
$u_2 = ((1-\alpha)P_T|h_2|^2 + \alpha P_T|h_2|^2 + \sigma^2)^2 - (((1-\alpha)P_T|h_2|^2\kappa_{x_2}) - (\alpha P_T|h_2|^2\kappa_{x_1}))^2$,
$d_2 = (\alpha P_T|h_2|^2 + \sigma^2)^2 - (\alpha P_T|h_2|^2\kappa_{x_1})^2$,
$u_3 = 2\ d_1[(P_T|h_1|^2)^2\left(\alpha(1-\kappa_{x_1}^2) + \eta^2(1-2\,\alpha)(1+\kappa_{x_1}\kappa_{x_2}) - (\eta^4(1-\alpha)(1-\kappa_{x_2}^2))\right) + \sigma^2 P_T|h_1|^2$
$(1-\eta^2)] - 2\,u_1[-(P_T|h_1|^2)^2\eta^4(1-\alpha)(1-\kappa_{x_2}^2) - \sigma^2\eta^2 P_T|h_1|^2]$,

Algorithm I-a:Spectral Efficiency Maximization Algorithm for Two Users Using IGS:

1: **Initialization:** R_{m_1}, R_{m_2}, P_T, h_1, h_2, σ^2, and η;
2: **While:**$(\mathcal{R}_i - R_{m_i}) \neq 0, \forall i = 1, 2$, **do** {
3: assume that the optimal solution $\kappa_{x_1}^*, \kappa_{x_2}^*, \alpha^*$ belongs to *Case 1*, i.e. $\mathcal{R}_1 > R_{m_1}$ and $\mathcal{R}_2 > R_{m_2}$, then, find the sub-optimal solution $\kappa_{x_1}^*, \kappa_{x_2}^*, \alpha^*$ by solving (3.26)-(3.28) simultaneously when $\lambda_1 = \lambda_2 = 0$
4: **if** in Step 3, $\mathcal{R}_1 < R_{m_1}$ and $\mathcal{R}_2 \geq R_{m_2}$, **then** the sub-optimal solution $\kappa_{x_1}^*, \kappa_{x_2}^*, \alpha^*$ belongs to *Case 2*, i.e., find $\lambda_1 \geq 0$ by (3.29) such that $\mathcal{R}_1(\kappa_{x_1}, \kappa_{x_2}, \alpha) = R_{m_1}$, and re-calculate $\kappa_{s_1^*}, \kappa_{s_2^*}, \alpha^*$ from solving (3.26)-(3.28) simultaneously. Repeat until convergence.
5: **else if** in Step 3, $\mathcal{R}_1 \geq R_{m_1}$ and $\mathcal{R}_2 < R_{m_2}$, **then**, the sub-optimal solution $\kappa_{x_1}^*, \kappa_{x_2}^*, \alpha^*$ belongs to *Case 3*, i.e., find $\lambda_2 \geq 0$ by (3.29) such that $\mathcal{R}_2(\kappa_{x_1}, \kappa_{x_2}, \alpha) = R_{m_2}$ and re-calculate $\kappa_{s_1^*}, \kappa_{s_2^*}, \alpha^*$ from solving (3.26)-(3.28) simultaneously. Repeat until convergence.
6: **else if** in Step 3, $\mathcal{R}_1 < R_{m_1}$ and $\mathcal{R}_2 < R_{m_2}$, **then**, find $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ if feasible such that $\mathcal{R}_1(\kappa_{x_1}, \kappa_{x_2}, \alpha) = R_{m_1}$ and $\mathcal{R}_2(\kappa_{x_1}, \kappa_{x_2}, \alpha) = R_{m_2}$ and re-calculate $\kappa_{x_1^*}, \kappa_{x_2^*}, \alpha^*$ from solving (3.26)-(3.28) simultaneously. Repeat until convergence.
7: **Result:** $\kappa_{x_1^*}, \kappa_{x_2^*}, \alpha^*$ and substitute in (3.21), (3.22) to compute the maximum $\mathcal{R}_1(\kappa_{x_1}^*, \kappa_{x_2}^*, \alpha) + \mathcal{R}_2(\kappa_{x_1}, \kappa_{x_2}, \alpha^*)$

and

$d_3 = 2d_2[(P_T|h_2|^2)^2(-\alpha\kappa_{x_1}^2 + (1-2\,\alpha)(\kappa_{x_1}\kappa_{x_2}) + (1-\alpha)(1-\kappa_{x_2}^2))] - 2u_2[(P_T|h_2|^2)^2\alpha(1-\kappa_{x_1}^2) + \sigma^2 P_T|h_2|^2]$.

Then, (3.26), (3.27), and (3.28) can be solved simultaneously to compute optimal κ_{x_1}, κ_{x_2}, and the allocation power parameter α, at $\lambda_1 = \lambda_2 = 0$.

From the KKT conditions mentioned earlier, we need to check all alternatives of $\lambda_i, \forall i = 1, 2$. As can be seen from (4.24b) and (4.24c), it is either we have $\lambda_1 = 0$ or $\mathcal{R}_1(\kappa_{x_1}^*, \kappa_{x_2}^*, \alpha^*) = R_{m_1}$, or $\lambda_2 = 0$ or $\mathcal{R}_2(\kappa_{x_1}^*, \kappa_{x_2}^*, \alpha^*) = R_{m_2}$. This leads to four potential cases as below:

– *Case 1:* Inactive QoS constraints, when both $\lambda_1 = \lambda_2 = 0$.

– *Case 2:* The sub-optimal solutions of $\kappa_{x_1}, \kappa_{x_2}$, and α exist when $\lambda_1 \neq 0$ and $\lambda_2 = 0$, and $\mathcal{R}_1(\kappa_{x_1}^*, \kappa_{x_2}^*, \alpha^*) = R_{m_1}$.

– *Case 3:* The sub-optimal solutions of $\kappa_{x_1}, \kappa_{x_2}$, and α exist when $\lambda_1 = 0$ and $\lambda_2 \neq 0$, and $\mathcal{R}_2(\kappa_{x_1}^*, \kappa_{x_2}^*, \alpha^*) = R_{m_2}$.

– *Case 4:* The sub-optimal solutions of $\kappa_{x_1}, \kappa_{x_2}$, and α exist, if feasible, when $\lambda_1 \neq 0$

and $\lambda_2 \neq 0$ and both $\mathcal{R}_1(\kappa^*_{x_1}, \kappa^*_{x_2}, \alpha^*) = R_{m_1}$ and $\mathcal{R}_2(\kappa^*_{x_1}, \kappa^*_{x_2}, \alpha^*) = R_{m_2}$.

The Lagrangian values will be computed using the Subgradient method as

$$\lambda_i^{n+1} = [\lambda_i^n - \delta_i^n(\mathcal{R}_i - R_{m_i})]^+, \quad \forall i = 1, 2, \tag{3.29}$$

where δ_i^n is the a small increment at the nth iteration associated with the ith Lagrange multiplier.

Based on the above-mentioned cases, we introduce Algorithm I to solve the optimization problem in $\mathcal{OP}1$. The optimal solution belongs to one of the following four cases: 1) both minimum user rate constraints are inactive, 2) user 1 minimum rate constraint is active and user 2 minimum rate constraint is inactive, 3) user 1 minimum rate constraint is inactive and user 2 minimum rate constraint is active, and 4) both minimum user rate constraints are active. We explain the steps of the algorithm as follows.

- Step 3: the proposed algorithm starts by assuming that both the minimum user rate constraints are inactive. Then, we find the optimal solution based on this assumption. If the inactive constraints are satisfied, then the optimal solution is reached.

- Step 4: based on the assumption that solution belongs to *Case 1* (inactive constraints), the user 1 minimum rate constraint may be not inactive. This means that initial solution from step 3 is infeasible and the proposed algorithm finds the Lagrangian multipliers that enforce the solution to be in the feasible region. More specifically, the proposed algorithm finds the non-negative Lagrangian multiplier λ_1 that makes user 1 minimum rate constraint active (i.e., satisfied with equal sign)- *Case 2*.

- Step 5: based on the assumption that solution belongs to *Case 1* (inactive constraints), the user 2 minimum rate constraint may be not inactive. This means that initial solution from step 3 is infeasible and the proposed algorithm finds the

Lagrangian multipliers that enforce the solution to be in the feasible region. More specifically, the proposed algorithm finds the non-negative Lagrangian multiplier λ_2 that makes user 2 minimum rate constraint active (i.e., satisfied with equal sign)-*Case 3*.

- Step 6: based on the assumption that solution belongs to *Case 1* (inactive constraints), both minimum rate constraints may be not inactive. This means that initial solution from step 3 is infeasible, and the proposed algorithm finds the Lagrangian multipliers that enforce the solution to be in the feasible region. More specifically, the proposed algorithm finds the non-negative Lagrangian multiplier λ_1, λ_2 that makes both minimum rate constraints active (i.e., satisfied with equal sign)-*Case 4*.

- At the end, the optimal IGS circularity coefficients and power allocation parameter are obtained and the corresponding maximum sum-rate can be computed.

3.5.2 Special Case: IGS at Strong User Only

We focus on the case where we use IGS for strong user (i.e., x_1 is improper and $\kappa_{x_1} \neq 0$) and PGS for weak user (i.e., x_2 is proper and $\kappa_{x_2} = 0$). Also, we assume the powers P_1 and P_2 are already allocated to user 1 and 2, respectively (i.e., they are not optimization variables). That said, the optimization problem for maximizing the sum-rate under QoS constraints can be formulated as

$$\begin{aligned} \underset{\kappa_{s_1}}{\text{maximize}} \quad & \mathcal{R}_1(\kappa_{x_1}) + \mathcal{R}_2(\kappa_{x_1}) \\ \text{subject to} \quad & \mathcal{C}1: \ \mathcal{R}_1(\kappa_{x_1}) \geq R_{m_1}, \\ & \mathcal{C}2: \ \mathcal{R}_2(\kappa_{x_1}) \geq R_{m_2}, \\ & \mathcal{C}3: \ 0 \leq \kappa_{x_1} \leq 1, \end{aligned} \tag{3.30}$$

where $R_1(\kappa_{x_1})$ and $R_2(\kappa_{x_1})$ are obtained from (3.21) and (3.22), respectively, at $\kappa_{x_2} = 0$. R_{m_1} and R_{m_2} are the minimum rate requirements of the strong user and the weak user,

respectively. The constraint $\mathcal{C}3$ reflects that the circulatory coefficient is between 0 and 1.

The optimization problem in (3.30) can be solved by applying the KKT conditions; however, it is worthy to mention that the obtained circularity coefficient $\kappa_{s_1}^*$ will be sub-optimal as the the problem in (3.30) is non-convex. The Lagrangian function can be expressed as

$$L(\kappa_{x_1}, \lambda_1, \lambda_2) = -\left(\mathcal{R}_1(\kappa_{x_1}) + \mathcal{R}_2(\kappa_{x_1})\right) + \lambda_1(R_{m_1} - \mathcal{R}_1(\kappa_{x_1})) + \lambda_2(R_{m_2} - \mathcal{R}_2(\kappa_{x_1})), \tag{3.31}$$

where λ_1 and λ_2 are the non-negative Lagrange multipliers associated with the QoS constraints of user 1 and 2, respectively. The circularity coefficient constraint not considered in the Lagrangian function will be satisfied later. That said, the KKT conditions can be written as follows [180]

$$\begin{aligned}
\frac{\partial L(\kappa_{x_1}^*, \lambda_1, \lambda_2)}{\partial \kappa_{x_1}} &= 0, && (3.32)\\
\lambda_1(R_{m_1} - \mathcal{R}_1(\kappa_{x_1}^*)) &= 0, && (3.33)\\
\lambda_2(R_{m_2} - \mathcal{R}_2(\kappa_{x_1}^*)) &= 0, && (3.34)\\
R_{m_1} - \mathcal{R}_1(\kappa_{x_1}^*) &\leq 0, && (3.35)\\
R_{m_2} - \mathcal{R}_2(\kappa_{x_1^*}) &\leq 0, && (3.36)\\
\lambda_1, \lambda_2 &\geq 0. && (3.37)
\end{aligned}$$

From (3.32), we can obtain the circularity coefficient $\kappa_{x_1}^*$ as

$$\begin{aligned}
\kappa_{x_1}^2 &= 0.5\left((\Phi + \Psi) + (\Phi - \Psi)\left(\frac{1+\lambda_2}{1+\lambda_1}\right)\right)\\
&- 0.5\left(\left[(\Phi - \Psi)^2\left(1 + \left(\frac{1+\lambda_2}{1+\lambda_1}\right)^2\right) + (\Phi - \Psi)\left(\frac{1+\lambda_2}{1+\lambda_1}\right)(2(\Phi + \Psi) - 4\omega)\right]\right)^{\frac{1}{2}}.
\end{aligned} \tag{3.38}$$

Algorithm I-b:Spectral Efficiency Maximization Algorithm for IGS at Strong User Only :

1: **Input:** R_{m_1}, R_{m_2}, P_1, P_2, h_1, h_2, σ^2, $\kappa_{x_2} = 0$, and η.
2: Set $\lambda_1 = \lambda_2 = 0$. Calculate κ_{x_1} from (3.38). Calculate $\mathcal{R}_1$ and $\mathcal{R}_2$ from (3.21) and (3.22), respectively.
3: **if** $\mathcal{R}_1 \geq R_{m_1}$ and $\mathcal{R}_2 \geq R_{m_2}$, **then**, the sub-optimal solution $\kappa^*_{x_1}$ is reached.
4: **else if** $\mathcal{R}_1 < R_{m_1}$ and $\mathcal{R}_2 \geq R_{m_2}$, **then**, find non-negative λ_1 from (3.29) such that $\mathcal{R}_1(\kappa_{x_1}) = R_{m_1}$ and re-calculate $\kappa_{s_1^*}$ from (3.38). Repeat until convergence.
5: **else if** $\mathcal{R}_1 \geq R_{m_1}$ and $\mathcal{R}_2 < R_{m_2}$, **then**, find non-negative λ_2 from (3.29) such that $\mathcal{R}_2(\kappa_{x_1}) = R_{m_2}$ and re-calculate $\kappa_{s_1^*}$ from (3.38). Repeat until convergence.
6: **else** $\mathcal{R}_1 < R_{m_1}$ and $\mathcal{R}_2 < R_{m_2}$, **then**, find non-negative λ_1 and λ_2 from (3.29) if exists such that $\mathcal{R}_1(\kappa_{x_1}) = R_{m_1}$ and $\mathcal{R}_2(\kappa_{x_1}) = R_{m_2}$ and re-calculate $\kappa_{s_1^*}$ from (3.38). Repeat until convergence.
7: **Output:** $\kappa_{x_1^*}$.

where $\Phi = \left(1 + \frac{\sigma^2}{P_1|h_2|^2}\right)^2$, $\Psi = \left(1 + \frac{P_2}{P_1} + \frac{\sigma^2}{P_1|h_2|^2}\right)^2$, and $\omega = \left(1 + \eta^2\frac{P_2}{P_1} + \frac{\sigma^2}{P_1|h_1|^2}\right)^2$. To consider $\mathcal{C}3$, we need to guarantee that the term under the square root in (3.38) is positive and also the first term of (3.38) is greater than the second term of it. The values of λ_1 and λ_2 in (3.33) and (3.34) can be computed using the subgradient method [180] (3.29).

However, one can notice from (3.33) that either $\lambda_1 = 0$ or $\mathcal{R}_1(\kappa^*_{x_1}) = R_{m_1}$. Similarly, (3.34) implies that either $\lambda_2 = 0$ or $\mathcal{R}_2(\kappa^*_{x_1}) = R_{m_2}$. That said, four possible cases exist, as follows.

– *Case 1:* $\lambda_1 = 0$ and $\lambda_2 = 0$ means that both QoS constraints of user 1 and user 2 are inactive.

– *Case 2:* $\lambda_1 = 0$ and $\lambda_2 \neq 0$ implies that the sub-optimal circularity coefficient exists when $\mathcal{R}_2(\kappa^*_{x_1}) = R_{m_2}$.

– *Case 3:* $\lambda_1 \neq 0$ and $\lambda_2 = 0$ implies that the sub-optimal circularity coefficient exists when $\mathcal{R}_1(\kappa^*_{x_1}) = R_{m_1}$.

– *Case 4:* $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$ implies that if the problem is feasible, the sub-optimal circularity coefficient exists when both $\mathcal{R}_1(\kappa^*_{x_1}) = R_{m_1}$ and $\mathcal{R}_2(\kappa^*_{x_1}) = R_{m_2}$.

The proposed algorithm to solve the problem in (3.30) can be formally summarized as shown above. The algorithm iterates between the four cases until convergence.

3.5.3 Energy Efficiency Maximization

Since NOMA is considered as interference limited systems under the impact of imperfect SIC, needs additional transmit power plus extra circuit power consumption for detecting the weak user's signal due to the imperfection of SIC. In this subsection, we maximize the energy efficiency of the two-user system considering both QoS and BS power constraints. We jointly optimize the circularity coefficients and power allocation at the BS to maximize the energy efficiency. The optimization problem is formulated as follows

$$\mathcal{OP}2: \underset{\kappa_{x_i},\alpha_i}{\text{maximize}} \quad \zeta_{EE} = \frac{\mathcal{R}_1(\kappa_{x_i},\alpha_i)+\mathcal{R}_2(\kappa_{x_i},\alpha_i)}{P_c+(\alpha_1+\alpha_2)P_T} \tag{3.39a}$$

$$\text{subject to} \quad \mathcal{C}1:\ \mathcal{R}_1(\kappa_{x_i},\alpha_i) \geq R_{m_1}, \tag{3.39b}$$

$$\mathcal{C}2:\ \mathcal{R}_2(\kappa_{x_i},\alpha_i) \geq R_{m_2}, \tag{3.39c}$$

$$\mathcal{C}3:\ \alpha_1+\alpha_2 \leq 1, \tag{3.39d}$$

$$\mathcal{C}4:\ 0 \leq \alpha_i \leq 1, \forall i=1,2, \tag{3.39e}$$

$$\mathcal{C}5:\ 0 \leq \kappa_{x_i} \leq 1, \forall i=1,2, \tag{3.39f}$$

where P_c is the BS's circuitry power consumption. The optimization problem in (3.39a) is equivalent to the following minimization problem $\mathcal{OP}3$:

$$\mathcal{OP}3: \underset{\kappa_{x_i},\alpha_i}{\text{minimize}}\ \zeta_{EE}^{-1} \tag{3.40}$$

$$\text{subject to} \quad \mathcal{C}1-\mathcal{C}5.$$

The objective function of (3.40) and rate constraints are non-convex; hence, the overall problem is non-convex, and the global optimal solution cannot be ensured.

The fractional non-convex optimization problem in (3.40) can be converted to an equivalent parametric optimization problem using concepts from fractional programming, namely the Dinkelbach approach [181]. Using this conversion, a new objective function

can be found as

$$\Phi_{EE} = (P_c + (\alpha_1 + \alpha_2)P_T) - \mathcal{K}(\mathcal{R}_1(\kappa_{x_i}, \alpha_i) + \mathcal{R}_2(\kappa_{x_i}, \alpha_i)), \tag{3.41}$$

where $\mathcal{K}$ is a non-negative constant. Then, the new optimization problem $\mathcal{OP}4$ becomes

$$\mathcal{OP}4 : \underset{\kappa_{x_i}, \alpha_i}{\text{minimize}} \quad \Phi_{EE}, \text{ subject to} \quad \mathcal{C}1 - \mathcal{C}5. \tag{3.42}$$

It was proven in [181] that at a certain value of $\mathcal{K}$, which is defined as $\mathcal{K}^*$, an optimal solution to $\mathcal{OP}4$ is also an optimal solution to $\mathcal{OP}3$. Hence, obtaining the optimal values of κ_{x_i}, α_i for $\mathcal{OP}3$ can be reached by obtaining the optimal values of $(\kappa_{x_i}(\mathcal{K}), \alpha_i(\mathcal{K}))$ for $\mathcal{OP}4$. We can then update the value of $\mathcal{K}$ until it reaches $\mathcal{K}^*$, where $\mathcal{K}^*$ is obtained when $\Phi_{EE} = 0$ [181] at optimal $\kappa^*_{x_i}$ and α^*_i.

To find the sub-optimal solutions, we solve $\mathcal{OP}4$ using the KKT conditions. The Lagrangian function $\mathcal{L}_{EE}(\kappa_{x_i}, \alpha_i)$ based on $\mathcal{OP}4$ can be expressed as

$$\begin{aligned}\mathcal{L}_{EE}(\kappa_{x_i}, \alpha_i) = (P_c + (\alpha_1 + \alpha_2)P_T) - \mathcal{K}\left(\mathcal{R}_1(\kappa_{x_i}, \alpha_i) + \mathcal{R}_2(\kappa_{x_i}, \alpha_i)\right) \\ + \lambda_1(R_{m_1} - \mathcal{R}_1(\kappa_{x_i}, \alpha_i)) + \lambda_2(R_{m_2} - \mathcal{R}_2(\kappa_{x_i}, \alpha_i)) + \lambda_3(\alpha_1 + \alpha_2 - 1),\end{aligned} \tag{3.43}$$

where λ_1, λ_2 and λ_3 are the Lagrange multipliers connected with the QoS conditions of user 1, user 2, and power allocation at the BS, respectively. The impact of the constraints $\mathcal{C}4$ and $\mathcal{C}5$ determine the valid ranges of κ_{x_i}, α_i. The KKT conditions can consequently

be outlined as follows:

$$\frac{\partial \mathcal{L}_{EE}(\kappa_{x_i}^*, \alpha_i^*)}{\partial \kappa_{x_i}} = 0, \qquad \forall i = 1, 2, \tag{3.44a}$$

$$\frac{\partial \mathcal{L}_{EE}(\kappa_{x_i}^*, \alpha_i^*)}{\partial \alpha_i} = 0, \qquad \forall i = 1, 2, \tag{3.44b}$$

$$\lambda_1(R_{m_1} - \mathcal{R}_1(\kappa_{x_i}^*, \alpha_i^*)) = 0, \tag{3.44c}$$

$$\lambda_2(R_{m_2} - \mathcal{R}_2(\kappa_{x_i}^*, \alpha_i^*)) = 0, \tag{3.44d}$$

$$\lambda_3(\alpha_1 + \alpha_2 - 1) = 0, \tag{3.44e}$$

$$R_{m_1} - \mathcal{R}_1(\kappa_{x_i}^*, \alpha^*) \leq 0, \tag{3.44f}$$

$$R_{m_2} - \mathcal{R}_2(\kappa_{x_i}^*, \alpha_i^*) \leq 0, \tag{3.44g}$$

$$\alpha_1 + \alpha_2 - 1 \leq 0, \tag{3.44h}$$

$$\lambda_1, \lambda_2, \lambda_3 \geq 0. \tag{3.44i}$$

From (3.44a), we obtain

$$\frac{\partial \mathcal{L}_{EE}(\kappa_{x_i}^*, \alpha_i^*)}{\partial \kappa_{x_i}} = -(\lambda_1 + \mathcal{K})\frac{\partial \mathcal{R}_1(\kappa_{x_i}, \alpha_i)}{\partial \kappa_{x_i}} - (\lambda_2 + \mathcal{K})\frac{\partial \mathcal{R}_2(\kappa_{x_i}, \alpha_i)}{\partial \kappa_{x_i}} = 0. \tag{3.45}$$

and

$$\frac{\partial \mathcal{L}_{EE}(\kappa_{x_i}^*, \alpha_i^*)}{\partial \alpha_i} = P_T - (\lambda_1 + \mathcal{K})\frac{\partial \mathcal{R}_1(\kappa_{x_i}, \alpha_i)}{\partial \alpha_i} - (\lambda_2 + \mathcal{K})\frac{\partial \mathcal{R}_2(\kappa_{x_i}, \alpha_i)}{\partial \alpha_i} + \lambda_3 = 0. \tag{3.46}$$

We have same derivation for (3.45) as in (3.26) and (3.27), respectively. The results of (3.46) are obtained as

$$\frac{\partial \mathcal{R}_1(\kappa_{x_i}^*, \alpha_i^*)}{\partial \alpha_1} = \frac{1}{\ln 2}\frac{P_T|h_1|^2(\alpha_1 P_T|h_1|^2(1-\kappa_{x_1}^2) + \eta^2\alpha_2 P_T|h_1|^2(1+\kappa_{x_1}\kappa_{x_2}) + \sigma^2)}{(\alpha_1 P_T|h_1|^2 + \eta^2\alpha_2 P_T|h_1|^2 + \sigma^2) - (\alpha_1 P_T|h_1|^2\kappa_{x_1}) - (\eta^2\alpha_2 P_T|h_1|^2\kappa_{x_2})}, \tag{3.47}$$

and

$$\frac{\partial \mathcal{R}_1(\kappa_{x_i}^*, \alpha_i^*)}{\partial \alpha_2} = \frac{1}{\ln 2} \frac{zz(\eta^2 P_T |h_1|^2)(\alpha_1 P_T |h_1|^2 (1 + \kappa_{x_1}\kappa_{x_2}) + \eta^2 \alpha_2 P_T |h_1|^2 (1 - \kappa_{x_2}^2) + \sigma^2) - xx\, yy}{yy \times zz}, \tag{3.48}$$

where $xx = (\eta^2 P_T |h_1|^2)(\eta^2 \alpha_2 P_T |h_1|^2 (1 - \kappa_{x_2}^2) + \sigma^2)$,
$yy = (\alpha_1 P_T |h_1|^2 + \eta^2 \alpha_2 P_T |h_1|^2 + \sigma^2)^2 - ((\alpha_1 P_T |h_1|^2 \kappa_{x_1}) - (\eta^2 \alpha_2 P_T |h_1|^2 \kappa_{x_2}))^2$, and
$zz = (\eta^2 \alpha_2 P_T |h_1|^2 + \sigma^2)^2 - (\eta^2 \alpha_2 P_T |h_1|^2 \kappa_{x_2})^2$, and

$$\frac{\partial \mathcal{R}_2(\kappa_{x_i}^*, \alpha_i^*)}{\partial \alpha_1} = \frac{1}{\ln 2} \frac{zz_2(P_T |h_2|^2)(\alpha_2 P_T |h_2|^2 (1 + \kappa_{x_1}\kappa_{x_2}) + \alpha_1 P_T |h_2|^2 (1 - \kappa_{x_1}^2) + \sigma^2) - yy\ xx_2}{yy_2 \times zz_2}, \tag{3.49}$$

where $xx_2 = (P_T |h_2|^2)(\alpha_1 P_T |h_2|^2 (1 - \kappa_{x_1}^2) + \sigma^2)$,
$yy_2 = (\alpha_2 P_T |h_2|^2 + \alpha_1 P_T |h_2|^2 + \sigma^2)^2 - ((\alpha_2 P_T |h_2|^2 \kappa_{x_2}) - (\alpha_1 P_T |h_2|^2 \kappa_{x_1}))^2$
and $zz_2 = (\alpha_1 P_T |h_2|^2 + \sigma^2)^2 - (\alpha_1 P_T |h_2|^2 \kappa_{x_1})^2$.
and

$$\frac{\partial \mathcal{R}_2(\kappa_{x_i}^*, \alpha_i^*)}{\partial \alpha_2} = \frac{1}{\ln 2} \frac{(P_T |h_2|^2)(\alpha_2 P_T |h_2|^2 (1 - \kappa_{x_2}^2) + \alpha_1 P_T |h_2|^2 (1 + \kappa_{x_1}\kappa_{x_2}) + \sigma^2)}{(\alpha_2 P_T |h_2|^2 + \alpha_1 P_T |h_2|^2 + \sigma^2) - ((\alpha_2 P_T |h_2|^2 \kappa_{x_2}) - (\alpha_1 P_T |h_2|^2 \kappa_{x_1}))^2}, \tag{3.50}$$

From (3.44c)-(3.44e), we either have $\lambda_1 = 0$ or $\mathcal{R}_1(\kappa_{x_i}^*, \alpha_i^*) = R_{m_1}$, $\lambda_2 = 0$ or $\mathcal{R}_2(\kappa_{x_i}^*, \alpha_i^*) = R_{m_2}$, and $\lambda_3 = 0$ or $\alpha_1 + \alpha_2 = 1$. This leads to eight possible states through which we need to iterate to find the optimal solution.

The energy efficiency solution is found at $\mathcal{K} = \mathcal{K}^*$, where $\mathcal{K}^* = \frac{P_c + (\alpha_1^* + \alpha_2^*) P_T}{(\mathcal{R}_1(\kappa_i^*, \alpha_i^*) + \mathcal{R}_2(\kappa_i^*, \alpha_i^*))}$ and is computed by applying the Dinkelbach method [181]. Here, we develop Algorithm II, which follows an approach similar to Algorithm I with the aim of computing the optimal values of $\kappa_{x_i}^*, \alpha_i^*$ that satisfy $\Phi_{EE_{min}}(\kappa_i^*, \alpha_i^*) = 0$, where $\Phi_{EE_{min}}$ is the minimum of Φ_{EE}. Algorithm II starts with an initial value of $\mathcal{K}$, denoted as $\mathcal{K}_{\text{initial}}$, and employs an error

tolerance of δ. This energy efficiency Algorithm II is outlined at the top of the next page.

Description of Step 3 in the Algorithm II: We assume the optimal solution belongs to the case where the QoS and power allocation constraints are inactive. We initially find the sub-optimal solution when assuming all constraints are satisfied (inactive constraints), and then for those constraints that are not satisfied, we find their Lagrange multipliers to make them satisfied with equal sign. We initially set $\lambda_1 = \lambda_2 = \lambda_3 = 0$ and find initial $\kappa^*_{x_i}, \alpha^*_i$ by simultaneously solving (3.45), (3.46) using Newton's method.

- If $\mathcal{R}_1 \geq R_{m_1}$, $\mathcal{R}_2 \geq R_{m_2}$, and if the power constraint is true, then, the sub-optimal solution can be reached.

- If $\mathcal{R}_1 < R_{m_1}$, $\mathcal{R}_2 \geq R_{m_1}$, and the power constraint is true, i.e. this means that the initial solution of inactive constraints is infeasible, then we find non-negative λ_1 such that $\mathcal{R}_1(\kappa^*_{x_i}, \alpha^*_i) = R_{m_1}$ (that enforces the solution to be in the feasible region) and re-calculate $\kappa^*_{x_i}, \alpha^*_i$.

- If $\mathcal{R}_1 \geq R_{m_1}$, $\mathcal{R}_2 < R_{m_1}$, and the power constraint is true, then, we find non-negative λ_2 such that $\mathcal{R}_2(\kappa^*_{x_i}, \alpha^*_i) = R_{m_2}$ and re-calculate $\kappa^*_{x_i}, \alpha^*_i$

- If $\mathcal{R}_1 \geq R_{m_1}$, $\mathcal{R}_2 \geq R_{m_1}$, and the power constraint is not true, then, we find non-negative λ_3 such that $\alpha_1 + \alpha_2 = 1$ and re-calculate $\kappa^*_{x_i}, \alpha^*_i$

- If $\mathcal{R}_1 < R_{m_1}$, $\mathcal{R}_2 \geq R_{m_1}$, and the power constraint is not met, then, then we find non-negative λ_1, λ_3 such that $\mathcal{R}_1(\kappa^*_{x_i}, \alpha^*_i) = R_{m_1}$ and $\alpha_1 + \alpha_2 = 1$ and re-calculate $\kappa^*_{x_i}, \alpha^*_i$

- If $\mathcal{R}_1 \geq R_{m_1}$, $\mathcal{R}_2 < R_{m_1}$, and the power constraint is not met, then, we find non-negative λ_2, λ_3 such that $\mathcal{R}_2(\kappa^*_{x_i}, \alpha^*_i) = R_{m_2}$ and $\alpha_1 + \alpha_2 = 1$ and re-calculate ξ^*_i, ρ^*_i.

- If $\mathcal{R}_1 < R_{m_1}$, $\mathcal{R}_2 < R_{m_1}$, and the power constraint is true, then, we find non-negative λ_1, λ_2 such that $\mathcal{R}_1(\kappa^*_{x_i}, \alpha^*_i) = R_{m_1}$ and $\mathcal{R}_2(\kappa^*_{x_i}, \alpha^*_i) = R_{m_2}$ and re-calculate ξ^*_i, ρ^*_i.

Algorithm II: Energy Efficiency Maximization Algorithm:

1: **INPUT:** R_{m_1}, R_{m_2}, P, δ, $\mathcal{K}_{initial}$, and $\Phi_{EE_{min}} = -\infty$
2: **While:** $\Phi_{EE_{min}} < -\delta$, **do**
3: Find the values of κ_i^*, α_i^* as in **Algorithm I**
4: Update Φ_{EE} from (3.41) and calculate $\Phi_{EE_{min}}(\mathcal{K})$,
5: Calculate $\mathcal{K} = \frac{P_c + (\alpha_1^* + \alpha_2^*)P_T}{(\mathcal{R}_1(\kappa_i^*, \alpha_i^*) + \mathcal{R}_2(\kappa_i^*, \alpha_i^*))}$
6: **end While**
7: **OUTPUT:** $\mathcal{K}^*$, and $\kappa_{x_i}^*, \alpha_i^*$.

- If $\mathcal{R}_1 < R_{m_1}$, $\mathcal{R}_2 < R_{m_1}$, and the power constraint is not true, then, we find non-negative λ_1, λ_2 and λ_3 such that $\mathcal{R}_1(\kappa_{x_i}^*, \alpha_i^*) = R_{m_1}$, $\mathcal{R}_2(\kappa_{x_i}^*, \alpha_i^*) = R_{m_2}$, $\alpha_1 + \alpha_2 = 1$ are true, and re-calculate $\kappa_{x_i}^*, \alpha_i^*$.

3.5.4 Complexity Analysis of Algorithm II

The complexity analysis of Algorithm II can be described as follows. Step 3: the complexity order of this step is the complexity of Algorithm I. Let us assume that the maximum number of iterations needed for the subgradient method to converge is T, then the number of operations is of a complexity order equal to $\mathcal{O}(T)$, which is the complexity of Step 2 in Algorithm I. The computational requirement of Newton's method to solve a system of M equations in M unknowns is $\mathcal{O}(ML)$, where L is the number of required iterations [162], which is the complexity of Steps 3 to 6 of Algorithm I. Thus, the complexity order up to Step 4 (of Algorithm II) is $\mathcal{O}(TML)$. Accordingly, the complexity order of the proposed algorithm II is $\mathcal{O}(TMLN_k)$, where N_k is the number of executions of the while loop to update $\mathcal{K}$ in the Dinkelbach approach.

3.6 Discussion and Simulation Results

3.6.1 Simulation Setup

In this section, we evaluate the proposed solutions of the formulated optimization problems. A comparison in terms of the overall spectral efficiency and energy efficiency of

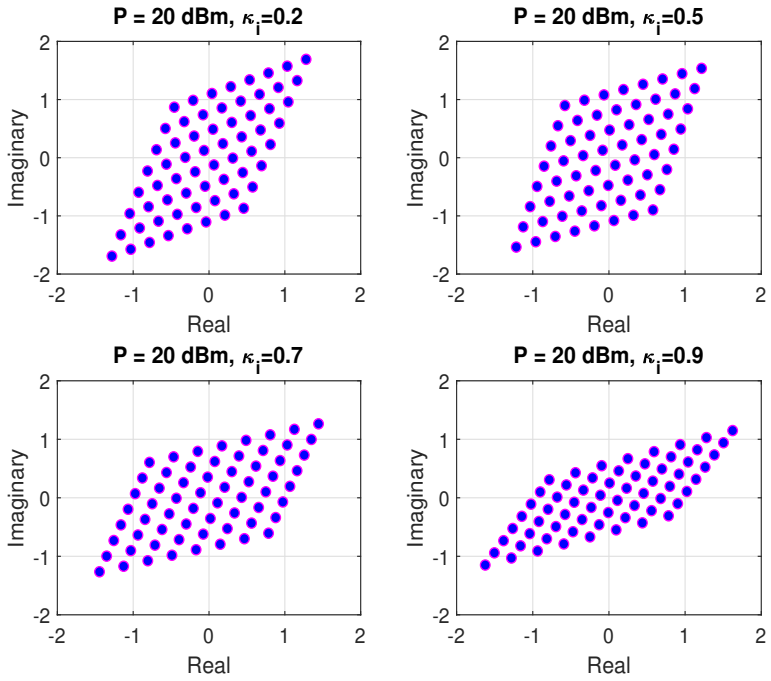

Figure 3.2: Improper constellation diagram with 64-QAM.

the system with its counterpart of a NOMA system employing traditional PGS at both users is performed. We also compare the proposed solution that considers IGS only at the strong user with PGS-based NOMA scheme. We consider the distance-dependent path-loss model as a form of large-scale fading, and the Rayleigh fading model as small-scale multi-path fading. The channel from the BS to user i, $\forall i = 1, 2$, at a distance of d_i meters is generated as $\sqrt{10^{-\frac{\sigma_{PL}}{10}}}h_i$, where h_i is a Rayleigh fading channel coefficient and $\sigma_{PL} = 38.46 + 10n\log_{10}(d_i)$ is the path-loss in dB.

In the definition of σ_{PL}, the loss factor 38.46 is the free space path loss at a reference distance of 1 meter and at carrier frequency of 2 GHz, and $n = 3$ is the path-loss exponent [182]. We set the noise power density $N_o = -174$ dBm/Hz with bandwidth $B = 20$ MHz. The distance d_1 between user 1 and the BS is set to 20 meters and the distance d_2 between user 2 and the BS is set to 100 meters. Unless otherwise stated, we assume that $\mathcal{E}[|h_1|^2] > \mathcal{E}[|h_2|^2]$, $P_c = 20$ dBm, and $\delta = 10^{-4}$. $R_{m_1} = R_{m_2} = 1.5$ bits/sec/Hz, and $\alpha = 0.4$. Unless otherwise mentioned, we assume $\eta = 0.1$.

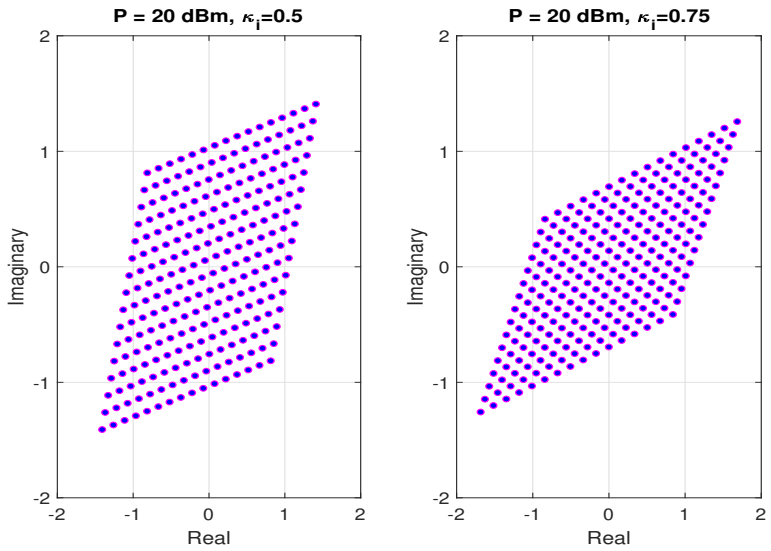

Figure 3.3: Improper constellation diagram with 256-QAM.

3.6.2 Improper Constellation Design

Figs. 3.2 and 3.3 show two examples of the improper constellation diagrams with 64-QAM and 256-QAM, respectively, which are designed based on WLT as described in Section II. We generate unit average energy standard 64-QAM and 256-QAM constellations. Next, we find the optimized $\kappa_{x_1}^*, \kappa_{x_2}^*$ at certain channel realizations. At these optimized values, we then generate the improper constellations using WLT based on these prescribed $\kappa_{x_1}^*, \kappa_{x_2}^*$. Considering this optimized improper constellation diagram, the minimum Euclidean distance between the constellations points is maximized due to WLT compared to the PGS standard constellation diagrams, which yields lower error probability and hence better spectral efficiency.

3.6.3 Spectral Efficiency Performance

In Fig. 3.4, the overall spectral efficiency of the system is depicted as a function of P_T, at different levels of the SIC imperfections η. The performance of the proposed algorithm I-a (jointly optimizing IGS circularity coefficients $\kappa_{x_1}^*, \kappa_{x_2}^*$ without power allocation, i.e., at fixed $\alpha = 0.4$) is compared with the following two cases: 1) strong user only employs

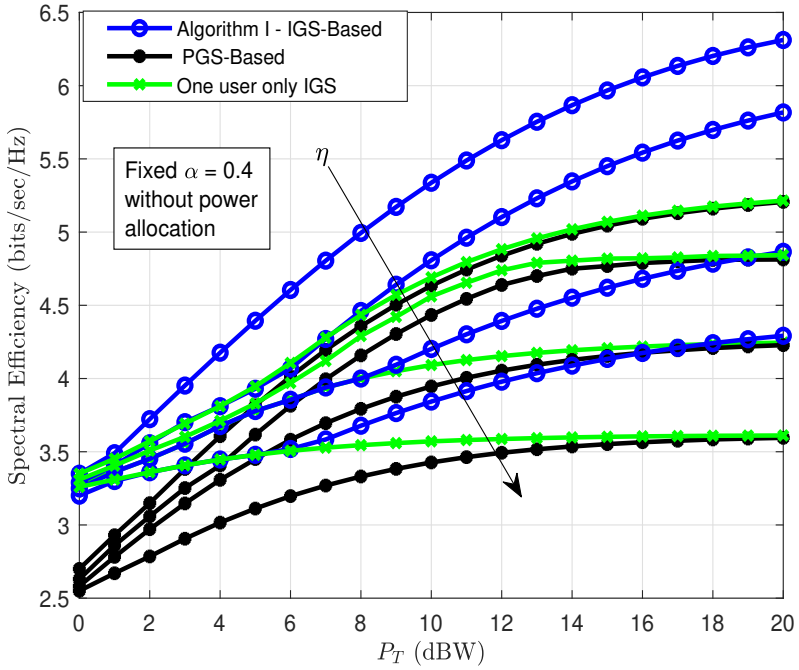

Figure 3.4: Spectral efficiency comparison between IGS-based and PGS-based for various $\eta = 0.2, 0.3, 0.4$, and fixed $\alpha = 0.4$.

IGS (optimized $\kappa^*_{x_1}$, algorithm I-b), and 2) both users employ PGS (traditional case).

As can be observed, the IGS-based scheme outperforms both one user only IGS-based and PGS-based schemes for all levels of η. Specifically, a considerable gain is attained at all power regimes compared with PGS-based. Moreover, the gain only appears at lower power regime in case of one-user only IGS-based scheme.

A spectral efficiency of about 4 bits/s/Hz can be attained by the modified proposed algorithm at 3 dBW power at $\eta = 0.1$ and $\alpha = 0.4$. To achieve the same spectral efficiency of 4 bits/s/Hz using only one user IGS-based and PGS-based schemes, nearly 2 to 2.5 dBW power are needed, respectively. It is also observed that as the SIC becomes worse, i.e., from $\eta = 0.1$ to $\eta = 0.4$, the performance of all schemes gets worse. However, the PGS-based scheme is the most impacted scheme.

While Fig. 3.4 assumes the fixed power allocation i.e. ($\alpha = 0.4$) scenario, Fig. 3.5 presents the optimized power allocation scenario. In particular, Fig. 3.5 shows the be-

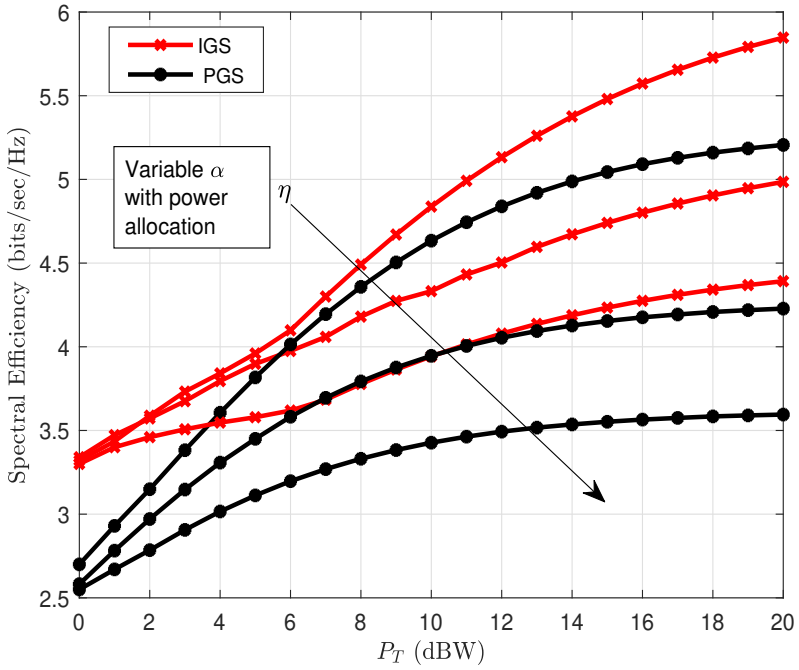

Figure 3.5: Spectral efficiency comparison between IGS-based and PGS-based for various values of $\eta = 0.1, 0.2, 0.3, 0.4$ and optimized α.

haviour of the proposed spectral efficiency algorithm (IGS-based with optimized power allocation) and its counterpart PGS-based NOMA system. It is clear that there is a further gain improvement for IGS-based scheme over the PGS-based one at different values of η. In overall, the results reveal the positive influence of jointly optimizing both IGS coefficients and power allocation parameter to enhance the NOMA system's spectral efficiency under imperfect SIC.

In Fig. 3.6, the convergence of the proposed spectral efficiency algorithm I-a is compared with the proposed algorithm I-b for the one user only IGS-based scheme at $\eta = 0.1$. The number of iterations required for IGS in the proposed algorithm is approximately double the number of iterations required for the algorithm I-b, but it is still relatively low.

In Fig. 3.7, the spectral efficiency is simulated versus P_T at different values of η. As can be seen, the IGS-based NOMA system outperforms PGS-based NOMA for all levels

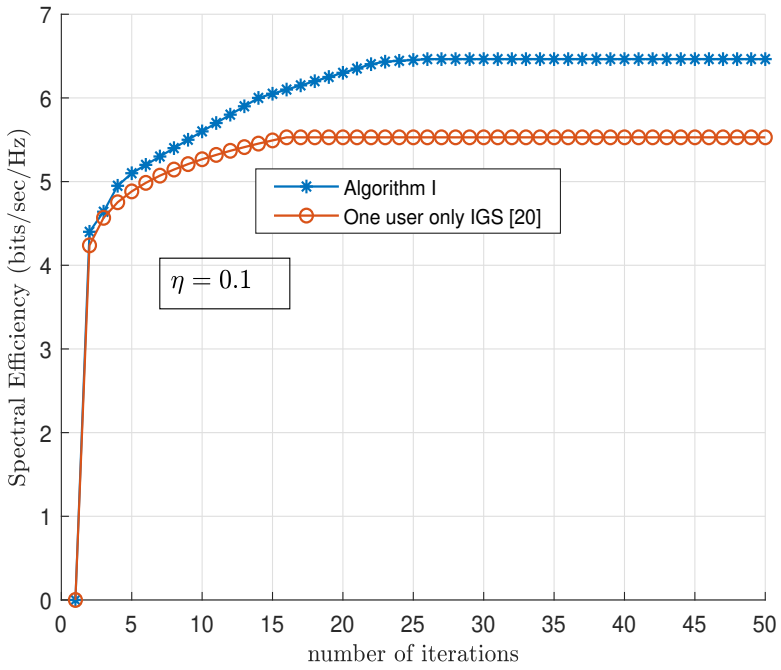

Figure 3.6: Convergence of the Algorithm I-a and the Algorithm I-b in terms of number of iterations at $\eta = 0.1$.

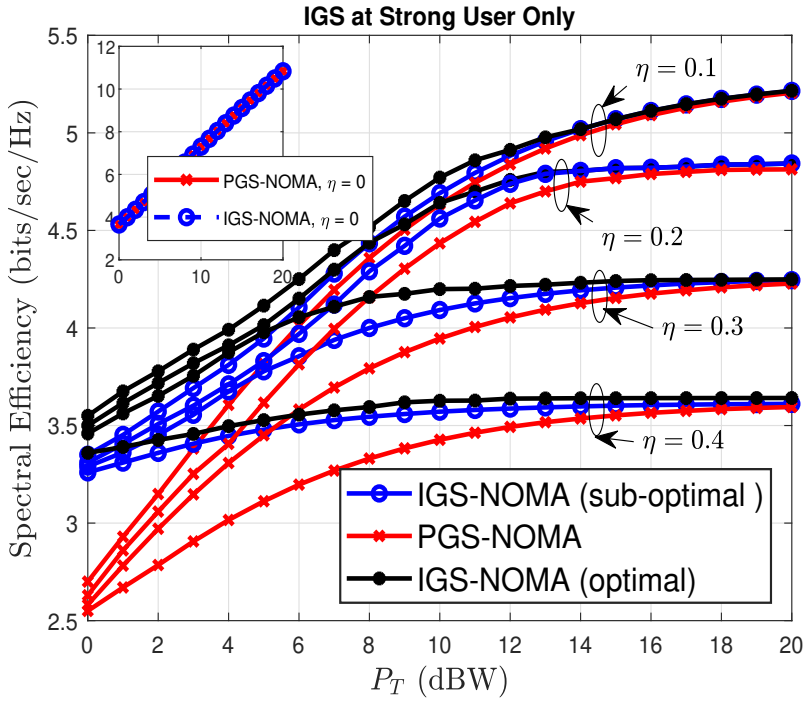

Figure 3.7: Spectral efficiency vs P_T for IGS-based and PGS-based NOMA systems for different η.

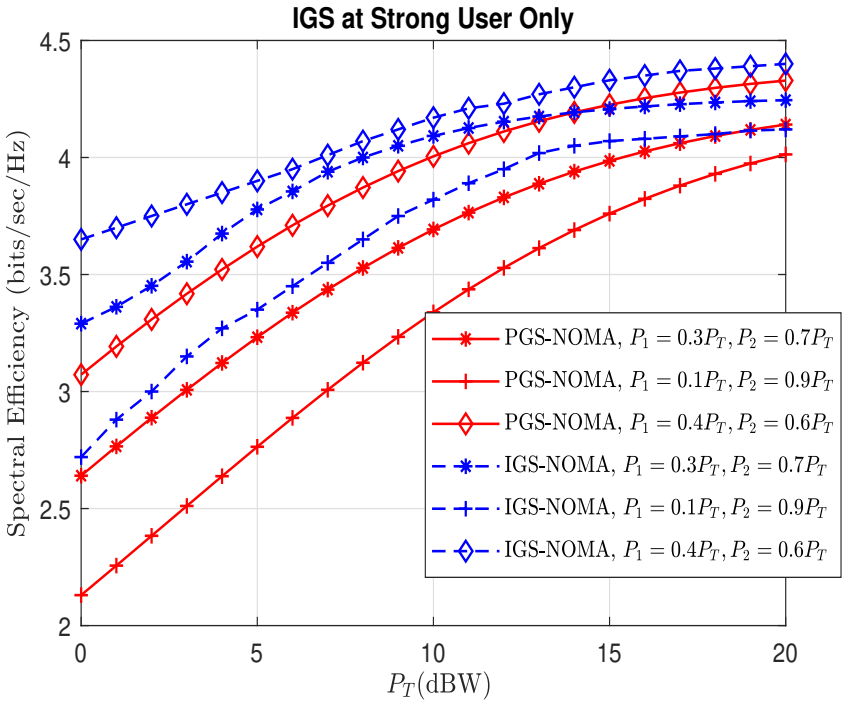

Figure 3.8: Spectral efficiency vs P_T for IGS-based and PGS-based NOMA systems for different P_1, P_2 values, with $\eta = 0.3$.

of imperfect SIC. In particular, as the SIC becomes worse, i.e., $\eta = 0.4$, the sum-rate gain of using IGS increases over PGS NOMA. IGS also offers a good gain in the low SNR region as the effect of the imperfect SIC is significant on the users' rate. At high SNR, the PGS-based NOMA system approaches the sum-rate performance of the IGS-based NOMA system.

In addition, we use exhaustive search method to find the optimal solution and compare it with the proposed KKT sub-optimal solution. The results show that there is a small performance gap between the optimal solution and proposed sup-optimal solution in terms of sum-rate at low SNR values and the gap tends to zero at high SNR values. It is worthy note that the proposed solution is far less complex than the optimal solution of the exhaustive search. The figure also shows that in case of perfect SIC, i.e., $\eta = 0$, both schemes perform similarly.

In Fig. 3.8, the spectral efficiency vs P_T for different values of P_1, P_2 at $\eta = 0.3$ is

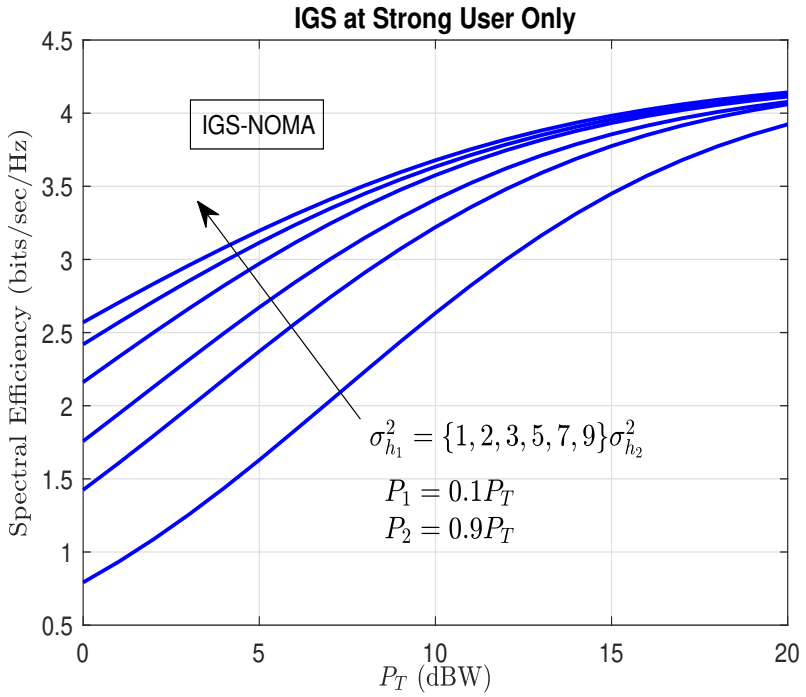

Figure 3.9: Spectral efficiency vs P_T for IGS-based NOMA system for different $\sigma^2_{h_1}$ to $\sigma^2_{h_2}$ ratios.

simulated for both IGS and PGS NOMA system. As the strong user gains more power, i.e., P_1 becomes larger, the sum-rate curves shift up and show higher spectral efficiency. One can notice from Fig. 3.8 that different power allocation ratios do not affect the gain of IGS over PGS based NOMA systems.

The effect of users' channel strength on the sum-rate for IGS-based NOMA performance is shown in Fig. 3.9. The sum-rate is simulated for the case of $\sigma^2_{h_1} = \{1, 2, 3, 5, 7, 9\}\sigma^2_{h_2}$ at $P_1 = 0.1P_T$, $P_2 = 0.9P_T$, and $\eta = 0.3$. It is clear that as $\sigma^2_{h_1}$ increases, the spectral efficiency enhances, i.e., as the channel of the first user becomes stronger, its rate becomes higher. Meanwhile, the rate of the user with weak channel is maximized by the proposed approach through the IGS-NOMA concept.

In Fig. 3.10, we show the convergence of the proposed algorithm at different values of η. On average, the algorithm needs small number of iterations to converge and the number increases as η increases.

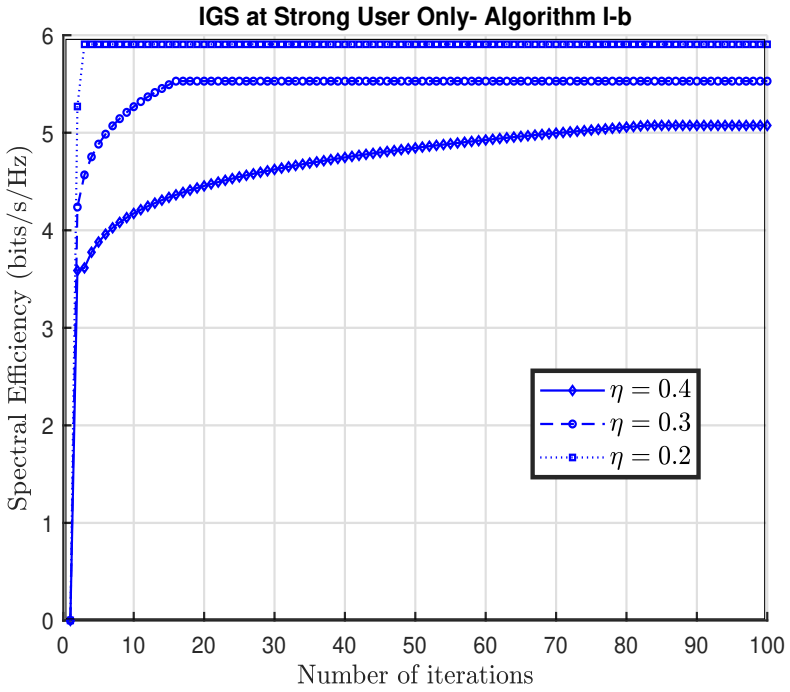

Figure 3.10: Spectral efficiency vs number of iterations for algorithm convergence with $\eta = 0.2, 0.3, 0.4$.

3.6.4 Energy Efficiency Performance

In Fig. 3.11, the performance of the the energy efficiency algorithm (in bits/joule) is plotted against P_T (dBm) at $P_c = 20$, 25 dBm. This figure shows that IGS enhances the energy efficiency performance in the proposed system by transmitting around 0.3 bits/joule more than PGS in all BS power regions. It is also observed that as the BS's power increases from low to medium, the energy efficiency performance improves. However, the energy efficiency saturates when the BS's transmit power increases which means that increasing the transmit power does not necessarily enhance the energy efficiency.

To demonstrate how close it is to the optimal solution, we compared the proposed sub-optimal solution with the exhaustive search solution, where the latter is performed through three nested loops with a step size of 0.05 (the three optimization variables are all bounded between 0 and 1). The result is a gap of around of 0.1 dB, which is acceptable performance loss given the high computational complexity of the exhaustive

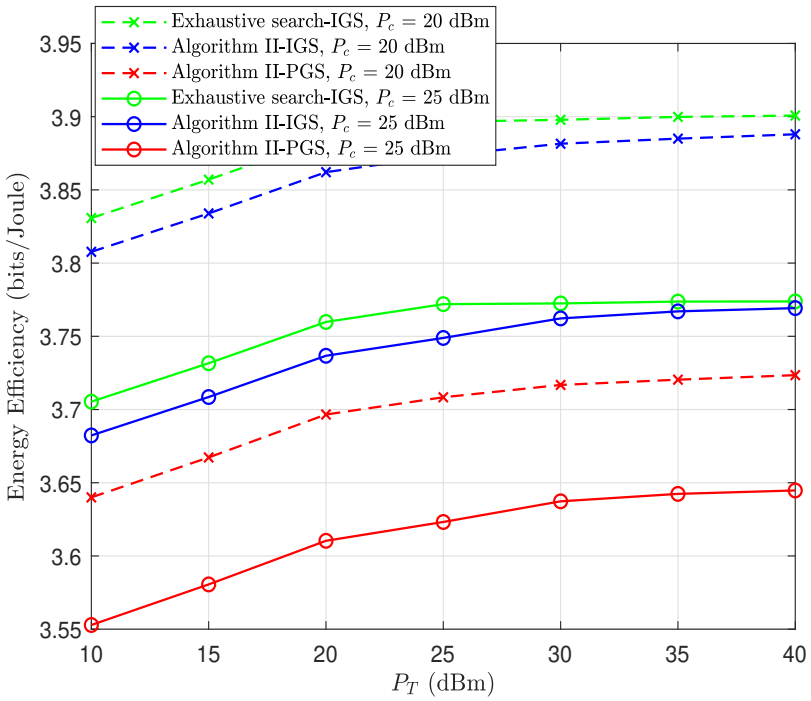

Figure 3.11: Energy efficiency vs P (dBm) at different values of P_c for proposed Algorithm II and exhaustive search.

search solution.

Figure 3.12 shows the energy efficiency performance against P_T dBm at different levels of SIC and at fixed $P_c = 20$ dBm. The figure shows the effectiveness of using IGS in the case of SIC imperfections when compared to the PGS case, especially at higher levels of η. As η increases, the gain due to using IGS becomes greater.

3.6.5 Error Performance

Fig. 3.13 shows the bit error rate (BER) through simulations versus P_T for both users in case of IGS and PGS at 16-QAM at $\eta = 0.1$. For each channel realization, we find the optimal improper coefficients, then we find the corresponding improper constellation diagram, simulate the error rate, and repeat.

For a given optimal IGS circularity coefficients, we study the BER performance of the new improper constellations, by means of 10^4 simulations, where each simulation

considering the decoding of 10^4 symbols. The optimal maximum likelihood detector is applied at the receiver side for this problem, which aims to finding the closest constellation point to a given noisy received signal. It should be noted that this approach may not result in the optimal BER performance since the decision variables κ_i and α_i are optimized to maximize the transmission rates (to approach Shannon capacity).

As can be seen in the figure, error performance in case of using improper constellation diagrams outperforms that of the proper constellation. The reason behind this is that since the improper constellation is designed based on WLT which relies on maximizing the minimum Euclidean distance, and hence achieves a better BER. Another observation is that as P_T increases, error performance improves as expected. However, error floor occurs at high P_T in case of user 1 due to residual interference resulting from the imperfect SIC.

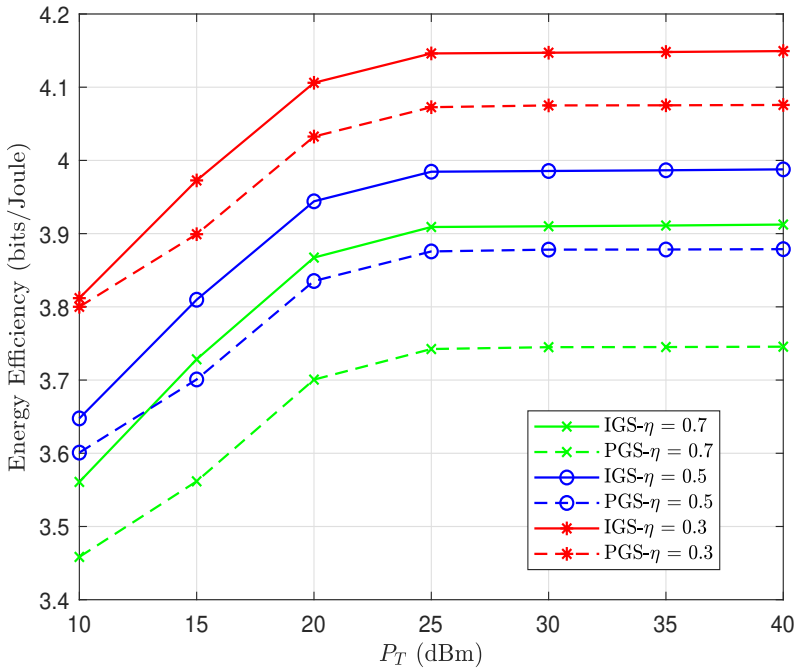

Figure 3.12: Energy efficiency vs P (dBm) at different values of η and fixed $P_c = 20$ dBm.

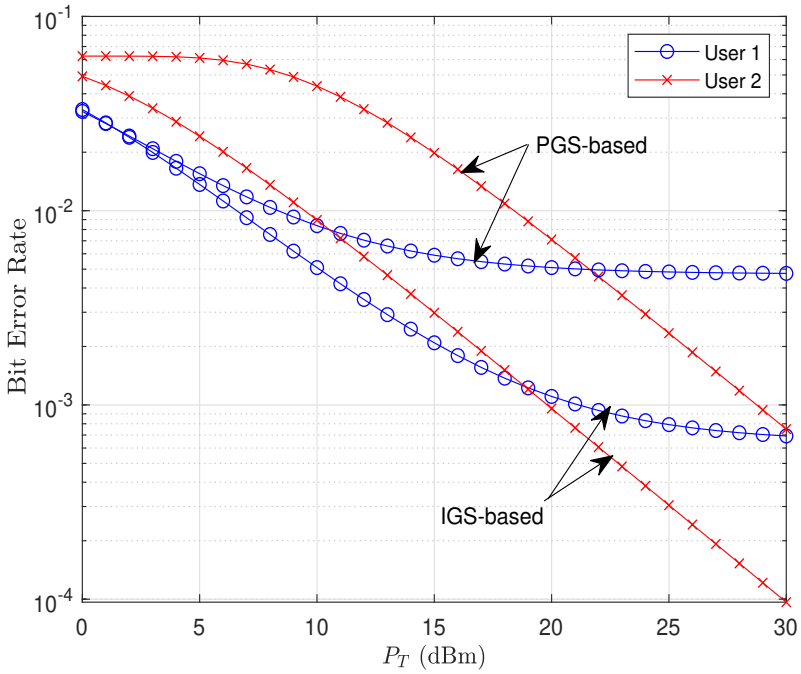

Figure 3.13: BER with improper constellation 16-QAM at $\eta = 0.1$.

3.7 Conclusion

In this chapter, the system spectral and energy efficiencies of a two-user NOMA system adopting IGS at both users are maximized such that the minimum rate requirements and power budget constraints are met under imperfect SIC. In addition, to study the system performance, improper constellations are designed using WLT based on predefined optimized IGS coefficients.

Results showed that system spectral and energy efficiencies of the IGS-based NOMA systems are further improved by jointly optimizing the circularity coefficients at both users compared to the case of optimizing the circularity coefficient of the single user IGS-based systems and PGS-based NOMA systems. A spectral efficiency of about 4 bits/s/Hz can be attained by the modified proposed algorithm at 3 dBW power at $\eta = 0.1$ and $\alpha = 0.4$. To achieve the same spectral efficiency of 4 bits/s/Hz using only one user IGS-based and PGS-based schemes, nearly 2 to 2.5 dBW power are needed, respectively. In addition, it

was observed that the gain from IGS increases when imperfect SIC gets higher. Moreover, spectral efficiency increases when channel gain ratio between users increases, but this improvement saturates at high SNR values.

In contrast to PGS-based NOMA systems, results revealed that IGS can save around 0.2 dBW of the transmit power, and hence, can be identified as an energy efficient signaling scheme. Furthermore, the results demonstrate that the error performance of the IGS-based system outperforms its counterpart PGS-based system due to impropriety characteristics of the constellation.

Chapter 4

Spectral and Energy Efficiency Maximization of Full-Duplex Relaying NOMA Systems with IGS Under Imperfect Self-Interference Cancellation

4.1 Introduction

Cooperative NOMA is introduced to enhance the overall system coverage and reliability in addition to spectral efficiency [71], [72], [73]. Cooperative NOMA can be categorized into two relaying strategies, namely dedicated relaying (employing an external entity as a relay) and user relaying (with one of the users acting as a relay). In the dedicated relaying NOMA systems, the BS transmits signals to the two users via an external relay or the two users send data to a common destination via an external relay [75]. In the user relaying NOMA systems, the BS transmits a superimposed mixture of two signals

to both the strong user U1 and the weak user U2, then U1 employs SIC to decode U2's signal and forwards this signal to U2. Depending on whether a direct link exists between the BS and U2, two scenarios of user relaying NOMA systems exist. The first scenario is when U2 receives its signal from both relaying phases and employs a combining method to improve its reception reliability, and the second scenario (adopted in this chapter) is when U2 receives its own message only through U1 [77].

To further improve the spectral efficiency of HD-NOMA systems that require two time slots for receiving and relaying data, full-duplex relaying has been recently investigated. In particular, the relay user U1 receives the superimposed signals from the BS and concurrently performs relay transmission in the same time slot. The authors in [163] studied the performance of the sum rate and outage probability in a cooperative FD-NOMA system, where the relay users receive and transmit signals simultaneously but on different frequency bands. The work in [164] computed the achievable rate region, assuming perfect SIC in a cooperative FD-NOMA network where relay users send and receive messages simultaneously on the same frequency band. Under imperfect SIC, authors in [165] analyzed outage performance in a cooperative FD-NOMA system. Furthermore, the authors in [166] investigated power minimization problems under outage constraints for a cooperative FD-NOMA system where outage probability and ergodic sum-rate expressions could be derived. With more realistic conditions, namely imperfect SIC in the FD-NOMA system, the authors in [167] were able to propose expressions for the achievable outage probability of both users and the ergodic sum capacity.

4.1.1 Related works

Recently, IGS was extended to full-duplex relaying in CR systems [174]. In the system model of [174], IGS was used to help the secondary user get the opportunity to access to the spectrum which was prevented from due to the self-interference introduced at the the primary network. Next, the authors in [175] employed IGS in full-duplex relaying with

non-negligible residual self-interference under Nakagami-m fading. In particular, a full-duplex decode-and-forward relay was utilized to assist in an end-to-end communication, where PGS was adopted at the source and IGS was used at the relay. The results in [175] showed that using IGS can mitigate the residual self-interference impact via tuning the signal impropriety and can make the system more energy efficient.

Similarly, a few works considered using IGS with NOMA in interference-limited networks. The authors in [176] analyzed the performance of a downlink NOMA system with IGS, deriving outage and ergodic rate expressions. In [177], transmit beamforming structures were developed for a multi-cell network in order to maximize the users' minimum throughput under various transmit power constraints. In [178], IGS was designed for information beamforming with the aim of improving the throughput of a multi-cell network and protecting user secrecy.

4.1.2 Contributions

Despite the advantages of using full-duplex relaying in NOMA systems, proper FD-NOMA systems still suffer from severe co-channel interference and are considered interference-limited [163]. To relieve the impact of the self-interference and enhance the system performance, IGS is adopted in this work. To the authors' best knowledge, IGS has not been considered in FD-NOMA relaying systems.

In this chapter, we investigate the potential performance merits of using IGS in the downlink of FD-NOMA relaying system where a BS serves two users on a near-far deployment. In particular, the user closer to the BS (U1) acts as a full-duplex relay (FD-relay) to the farther away user (U2) and forwards U2's signal beside extracting its own message. Unlike FD-OMA relaying system, the FD-relay in FD-NOMA system performs two roles; extracting both its own message and the other user's message, and forwarding the later to that user, whereas the FD-relay in the OMA system only acts as a relay and forwards the other user's message (there is no message for the relay itself).

Since the achievable rates of IGS are superior to those of PGS in interference-limited scenarios, we assume that both users employ improper signals. Due to superiority of IGS compared to PGS in terms of achievable rates and energy efficiency, we formulate three optimization problems for sum rate, energy efficiency, and max-min fairness where we jointly optimize the BS transmit power and the transmit signals circularity coefficients.

The main contributions of this work can be outlined as follows:

- We derive the exact expressions for the achievable user rates at each link in the cooperative FD-NOMA system when adopting IGS at both users.
- We formulate and solve three optimization problems to maximize the sum rate, maximize the energy efficiency, and maximize the minimum throughput of users. This is achieved by jointly optimizing the BS transmit power and the transmit signal circulatory coefficients.
- We propose iterative algorithms to find sub-optimal solutions of the formulated non-convex optimization problems based on the Karush-Kuhn-Tucker (KKT) conditions. Additionally, we show the effect of the optimized circulatory coefficient on the transmit constellation at the BS.
- Improper constellation diagram is designed based on the optimal circularity coefficient obtained from optimized sum rate algorithm. Error performance based on the improper constellations designed is also investigated.
- Simulation results demonstrate the effectiveness of IGS over PGS in the context of cooperative FD-NOMA systems.

4.2 System Model

We consider a cooperative FD-NOMA relaying system composed of a BS and two prepaired users (U1 and U2). In the setup of our system model, U2 has no-direct link with

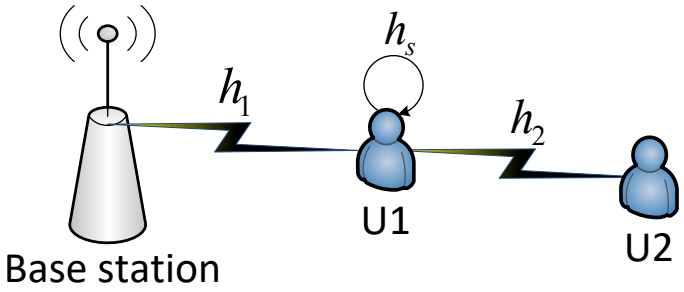

Figure 4.1: FD-NOMA system model with IGS.

the BS due to physical obstacles and/or heavy shadowing. As a result, U1 acts as a full-duplex relay by forwarding to U2 its message.

According to the theory behind NOMA, the BS transmits a superimposed message of both users' signals (x_1 and x_2) to U1. After receiving the signal, U1 extracts x_2 using SIC, and forwards it to U2. At the same time, U1 subtracts x_2 from the original signal and detects its own message x_1. It is worth-mentioning that in such a FD-NOMA relaying system model, an external relay is not needed to forward U2's signal since U1 receives its own message and acts as FD-relay at the same time/frequency.

It is assumed that channel coefficients experience Rayleigh fading and are modeled as $h_1 \sim \mathcal{CN}(0, \sigma_1^2)$, and $h_2 \sim \mathcal{CN}(0, \sigma_2^2)$, where h_1 denotes the channel coefficient between BS and U1 and h_2 denotes the channel coefficient between U1 and U2. Due to the FD-relaying strategy, residual interference occurs at the FD-relay. h_s denotes the self-interference coefficient and is modeled as $h_s \sim \mathcal{CN}(0, \sigma_s^2)$ [166].

We assume that both signals x_1 and x_2 are improper Gaussian with ξ_1 and ξ_2 as their IGS impropriety coefficients, respectively. Based on the NOMA concept, $|h_1|^2 > |h_2|^2$ and the BS transmits a superimposed downlink signal $x_1 + x_2$ to U1. Meanwhile, U1 transmits a delayed version of the decoded signal x_2 to U2. Thus, as a FD-relay, U1 receives $x_1 + x_2$ and transmits a processed x_2 version at the same time. This causes a self-interference due to the imperfections of the self-interference cancellation. We assume imperfect self-interference cancellation at U1, prompting to assign a factor $0 \leq \kappa \leq 1$ to

refer to the imperfection of self-interference cancellation technique. Based on this, the received signal at U1 in the first link is given as

$$y_1 = h_1 \left(\sqrt{\rho_1 P} x_1 + \sqrt{\rho_2 P} x_2 \right) + \sqrt{\kappa \hat{P}} h_s \hat{x}_2 + n_1, \tag{4.1}$$

where P is the total transmit power at the BS, ρ_1 and ρ_2, $(\rho_1 + \rho_2 \leq 1)$ are the power allocation parameters, $\hat{P}$ is the U1 transmit power, $\hat{x}_2$ is the processed version of x_2 where $\hat{x}_2 = x_2$ if U1 successfully decodes x_2, and n_1 is the additive white Gaussian noise (AWGN) at U1 and modeled as $n_1 \sim \mathcal{CN}(0, \sigma_{n1}^2)$. The received signal at U1 can be re-written as

$$y_1 = \underbrace{h_1 \sqrt{\rho_1 P} x_1}_{\text{desired signal}} + \underbrace{h_1 \sqrt{\rho_2 P} x_2 + \sqrt{\kappa \hat{P}} h_s \hat{x}_2 + n_1}_{\text{interference plus noise}}. \tag{4.2}$$

Considering perfect SIC, (4.1) is expressed as

$$y_1 = \underbrace{h_1 \sqrt{\rho_1 P} x_1}_{\text{desired signal}} + \underbrace{\sqrt{\kappa \hat{P}} h_s \hat{x}_2 + n_1}_{\text{interference plus noise}}. \tag{4.3}$$

Upon receiving y_1, U1 first decodes x_2 and then decodes its own message x_1 with the SIC technique. The received signal at U1 due to x_2 can be written as

$$y_{1,2} = \underbrace{\sqrt{\rho_2 P} h_1 x_2}_{\text{desired signal}} + \underbrace{\sqrt{\rho_1 P} h_1 x_1 + \sqrt{\kappa \hat{P}} h_s \hat{x}_2 + n_1}_{\text{interference plus noise}}. \tag{4.4}$$

The processed signal that U2 receives from U1 can be given as The processed signal that U2 received from U1 is given as

$$y_2 = \underbrace{\sqrt{\hat{P}} h_2 \hat{x}_2}_{\text{desired signal}} + \underbrace{n_2}_{\text{noise}}, \tag{4.5}$$

where n_2 is the AWGN at U2 and modeled as $n_2 \sim \mathcal{CN}(0, \sigma_{n2}^2)$.

4.3 Spectral Efficiency Analysis

In this section, user rates are derived assuming IGS-based signaling. We start by finding closed-form expressions. The achievable rate expression for a point-to-point IGS system is given as [21]

$$\mathcal{R}_i = \frac{1}{2}\log_2\left(\frac{\varphi_{y_i}^2 - |\hat{\varphi}_{y_i}|^2}{\varphi_{z_i}^2 - |\hat{\varphi}_{z_i}|^2}\right), \tag{4.6}$$

where φ_{y_i} and $\hat{\varphi}_{y_i}$ refer to the covariance and pseudo-covariance components of the received signal, while φ_{z_i} and $\hat{\varphi}_{z_i}$ refer to covariance and pseudo covariance components of interference plus noise signal, respectively. Define the circularity coefficient of x_i as $\xi_i = \frac{|\hat{C}_i|}{C_i}$, $\hat{C}_i = \mathcal{E}[x_i x_i] = \xi_i \angle \hat{C}_i$, and $C_i = \mathcal{E}[x_i x_i^*]$.

To find $\mathcal{R}_1$ for instance, we compute the covariance and pseudo-covariance components from (4.2)-(4.5) as follows.

$$\varphi_{y_1}^2 = (\mathcal{E}[y_1 y_1^*])^2 = (\rho_1 P|h_1|^2 C_1 + \kappa\hat{P}|h_s|^2 C_2 + \sigma_{n_1}^2)^2 \tag{4.7}$$

$$\hat{\varphi}_{y_1} = \mathcal{E}[y_1 y_1] = \rho_1 P h_1^2 \hat{C}_1 + \kappa\hat{P} h_s^2 \hat{C}_2, \tag{4.8}$$

$$\varphi_{z_1}^2 = (\mathcal{E}[z_1 z_1^*])^2 = (\kappa\hat{P}|h_s|^2 C_2 + \sigma_{n_1}^2)^2, \tag{4.9}$$

$$\hat{\varphi}_{z_1} = (\mathcal{E}[z_1 z_1]) = (\kappa\hat{P} h_s^2 \hat{C}_2). \tag{4.10}$$

and

$$|\hat{\varphi}_{y_1}|^2 = |\mathcal{E}[y_1 y_1]|^2 = (\rho_1 P|h_1|^2\xi_1)^2 + (\kappa\hat{P}|h_s|^2\xi_2)^2 + 2(\rho_1 P|h_1|^2\xi_1)(\kappa\hat{P}|h_s|^2\xi_2)$$
$$\cos(2\angle h_1 - 2\angle h_s + \angle\hat{C}_1 - \angle\hat{C}_2). \tag{4.11}$$

$$|\hat{\varphi}_{z_1}|^2 = |\mathcal{E}[z_1 z_1]|^2 = (\kappa\hat{P}|h_s|^2\xi_2)^2. \tag{4.12}$$

Assume the values of $\Gamma_1 = \frac{|h_1|^2}{\sigma_{n_1}^2}$, $\Gamma_2 = \frac{|h_2|^2}{\sigma_{n_2}^2}$, and $\Gamma_s = \frac{|h_s|^2}{\sigma_{n_1}^2}$ refer to the channel-to-noise ratio from the BS to U1, from U1 to U2, and the self-interference channel, respectively,

By substituting all these values in (4.6), and after some mathematical manipulations, the achievable rate at U1 is derived as

$$\mathcal{R}_1 = \frac{1}{2}\log_2\left[\frac{\Upsilon - |\rho_1 P\Gamma_1 \angle\Gamma_1 \hat{C}_1 + \kappa\hat{P}\Gamma_s\angle\Gamma_s\hat{C}_2|^2}{\left(\kappa\hat{P}\Gamma_s+1\right)^2 - \left(\kappa\hat{P}\Gamma_s\xi_2\right)^2}\right]. \tag{4.13}$$

where $\Upsilon = \left(\rho_1 P\Gamma_1 + \kappa\hat{P}\Gamma_s + 1\right)^2, \angle\Gamma_1 = 2\angle h_1, \angle\Gamma_s = 2\angle h_s$. Since we have an FD-NOMA system, $\mathcal{R}_2$ is expressed as [166]

$$\mathcal{R}_2 = \min\left[\mathcal{R}_{1,2}, \mathcal{R}_{2,2}\right], \tag{4.14}$$

where $\mathcal{R}_{1,2}$ is the rate at U1 when it detects U2's signal x_2, and $\mathcal{R}_{2,2}$ is the rate at U2 when it detects its signal x_2. Based on (4.6), $\mathcal{R}_{1,2}$ is derived as

$$\mathcal{R}_{1,2} = \frac{1}{2}\log_2\left[\frac{\Lambda - \left[|\rho_2 P\Gamma_1\angle\Gamma_1\hat{C}_2 + \kappa\hat{P}\Gamma_s\angle\Gamma_s\hat{C}_2 + \rho_1 P\Gamma_1\angle\Gamma_1\hat{C}_1|^2\right]}{\Upsilon - (|\kappa\hat{P}\Gamma_s\angle\Gamma_s\hat{C}_2 + \rho_1 P\Gamma_1\angle\Gamma_1\hat{C}_1|^2)}\right], \tag{4.15}$$

where $\Lambda = [(\rho_1+\rho_2)P\Gamma_1 + \kappa\hat{P}\Gamma_s + 1]^2$. To find $\mathcal{R}_2$ for instance, we compute the covariance and pseudo-covariance components from (4.2)-(4.5) as follows.

$$\varphi^2_{y_2} = (\mathcal{E}[y_2 y_2^*])^2 = \hat{P}|h_2|^2 C_2 + \sigma^2_{n_2}, \tag{4.16}$$

$$\hat{\varphi}_{y_2} = (\mathcal{E}[y_2 y_2]) = \hat{P}h_2^2 C_2, \tag{4.17}$$

$$|\hat{\varphi}_{y_2}|^2 = (\mathcal{E}[y_2 y_2]) = (\hat{P}|h_2|^2\xi_c)^2, \tag{4.18}$$

$$\varphi^2_{z_2} = (\mathcal{E}[z_2 z_2^*])^2 = \sigma^2_{n_2}, \tag{4.19}$$

$$\hat{\varphi}_{2_2} = 0. \tag{4.20}$$

In a similar fashion to (4.6), $\mathcal{R}_{2,2}$ is derived as

$$\mathcal{R}_{2,2} = \frac{1}{2}\log_2\left[\left(\hat{P}\Gamma_2\right)^2\left(1-\xi_2^2\right) + 2\left(\hat{P}\Gamma_2\right) + 1\right]. \tag{4.21}$$

4.4 Optimization Problems

In this section, optimization problems are formulated to achieve overall spectral efficiency maximization, max-min rate fairness, and energy efficiency maximization. We assume that QoS constraints are guaranteed for each user and that the power budget limit is satisfied.

4.4.1 Sum Rate Maximization

Maximizing the overall sum rate is a common objective in wireless networks. However, to avoid having all resources occupied by a single user, QoS or individual rate weights constraints are often enforced when determining the overall sum rate maximization. The weighted sum rate (WSR) maximization problem is advantageous in wireless applications where prioritization for certain users is desired. The weights can be selected by the scheduler based on the data traffic, throughput in the network, etc.

In this regard, we consider the weighted sum rate optimization problem to jointly optimize both circularity coefficients $\xi_i,\ \forall i = 1, 2$, and power allocation at the BS. We assume that the BS uses the total available power, i.e., $\rho_1 + \rho_2 = 1$. For simplicity, we denote $\rho_1 = \rho$ and $\rho_2 = 1 - \rho$. The following sum rate maximization problem considers the rates in (4.13) and (4.14), and can be written as

$$\mathcal{OP}1 : \underset{\xi_i,\rho}{\text{maximize}} \quad w_1\mathcal{R}_1(\xi_i,\rho) + w_2\mathcal{R}_2(\xi_i,\rho) \tag{4.22a}$$

$$\text{subject to} \quad \mathcal{C}1 :\ \mathcal{R}_1(\xi_i,\rho) \geq R_1^{\min}, \tag{4.22b}$$

$$\mathcal{C}2 :\ \mathcal{R}_2(\xi_i,\rho) \geq R_2^{\min}, \tag{4.22c}$$

$$\mathcal{C}3 :\ 0 \leq \xi_i \leq 1, \tag{4.22d}$$

$$\mathcal{C}4 :\ 0 \leq \rho \leq 1, \tag{4.22e}$$

where $R_i^{\min}$ is the QoS requirement at user i, w_1, w_2 are the weights of the users rates.

Keep in mind that the weights can be equal if both users have the same priority. The conditions $\mathcal{C}1$ and $\mathcal{C}2$ represent the fact that the rate of each user must meet $R_1^{\min}$ and $R_2^{\min}$, respectively, to guarantee QoS. The constraint $\mathcal{C}3$ defines the range of circularity coefficients between 0 and 1, and the constraint $\mathcal{C}4$ reflects the power allocation at the BS.

The problem formulated in (4.22) is non-convex due to non-linearity of the objective and QoS constraints, and thus, it is rather difficut to solve. However, (4.22) can be tackled by using KKT conditions [180], allowing us to obtain sub-optimal values for ξ_i, ρ. The Lagrangian function can be expressed as

$$\mathcal{L}(\xi_i, \rho) = -\left(w_1\mathcal{R}_1(\xi_i, \rho) + w_2\mathcal{R}_2(\xi_i, \rho)\right) + \lambda_1(R_1^{\min} - \mathcal{R}_1(\xi_i, \rho)) + \lambda_2(R_2^{\min} - \mathcal{R}_2(\xi_i, \rho)), \tag{4.23}$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are the Lagrange multipliers associated with the QoS requirements of U1 and U2, respectively. The effects of the constraints $\mathcal{C}3$ and $\mathcal{C}4$ will be considered later.

That said, the KKT conditions can be written as follows

$$\frac{\partial\mathcal{L}(\xi_i^*, \rho^*)}{\partial\xi_i, \partial\rho} = 0; \qquad \forall i = 1, 2, \tag{4.24a}$$

$$\lambda_1(R_1^{\min} - \mathcal{R}_1(\xi_i^*, \rho^*)) = 0, \tag{4.24b}$$

$$\lambda_2(R_2^{\min} - \mathcal{R}_2(\xi_i^*, \rho^*)) = 0, \tag{4.24c}$$

$$R_1^{\min} - \mathcal{R}_1(\xi_i^*, \rho^*) \leq 0, \tag{4.24d}$$

$$R_2^{\min} - \mathcal{R}_2(\xi_i^*, \rho^*) \leq 0, \tag{4.24e}$$

$$\lambda_1, \lambda_2 \geq 0. \tag{4.24f}$$

From (4.24a), we can conclude that

$$\frac{\partial\mathcal{L}(\xi_i^*, \rho^*)}{\partial\xi_i} = -(\lambda_1 + w_1)\frac{\partial\mathcal{R}_1(\xi_i, \rho)}{\partial\xi_i} - (\lambda_2 + w_2)\frac{\partial\mathcal{R}_2(\xi_i, \rho)}{\partial\xi_i} = 0. \tag{4.25}$$

To optimize the power at the BS, from (4.23), the following is determined:

$$\frac{\partial \mathcal{L}(\xi_i^*, \rho^*)}{\partial \rho} = -(\lambda_1 + w_1)\frac{\partial \mathcal{R}_1(\xi_i, \rho)}{\partial \rho} - (\lambda_2 + w_2)\frac{\partial \mathcal{R}_2(\xi_i, \rho)}{\partial \rho} = 0. \tag{4.26}$$

As previously mentioned, based on the FD-relaying concept, $\mathcal{R}_2$ is the minimum between $\mathcal{R}_{1,2}$ and $\mathcal{R}_{2,2}$. Hence, to be able to solve the optimization problem, we consider two scenarios: 1) *Case 1*: We assume $\mathcal{R}_2 = \mathcal{R}_{1,2}$, implying that we aim to maximize $\mathcal{R}_1 + \mathcal{R}_{1,2}$ and 2) *Case 2*: We assume $\mathcal{R}_2 = \mathcal{R}_{2,2}$, implying that we aim to maximize $\mathcal{R}_1 + \mathcal{R}_{2,2}$.

First, we find the range of ρ for each case in which either $\mathcal{R}_{1,2}$ or $\mathcal{R}_{2,2}$ is the minimum of $\mathcal{R}_2$. Second, within each region of ρ, we derive the sub-optimal solutions of ξ_1, ξ_2 and ρ by solving the corresponding equations in (4.25) and (4.26) for each case. Third, we verify the minimum between $\mathcal{R}_{1,2}$ and $\mathcal{R}_{2,2}$ and choose the appropriate case in order to find the sub-optimal solution that will be used to maximize the sum rate.

Case 1: $\mathcal{R}_2 = \mathcal{R}_{1,2}$

The value of ρ should be in the following range:

$$0 \leq \tilde{\rho} \leq \rho < 1. \tag{4.27}$$

The value of $\tilde{\rho}$ and explanatory proofs can be found in the *Appendix A.1.* When it comes to *Case 1*, we determine the candidates of sub-optimal solutions $\hat{\xi}_1, \hat{\xi}_2, \hat{\rho}$, by simultaneously solving (A.8)-(A.10), from *Appendix A.1* , which are the results of solving (4.25) and (4.26), respectively.

It can be seen that (A.8)-(A.10) are 3^{th} degree polynomials and are expected to have three roots each. Among all solutions obtained from (A.8)-(A.10), we accept the ones that satisfy the constraints $\mathcal{C}3$ and $\mathcal{C}4$ and result in the maximum sum rate. It should be noted that one of the possible solutions is $\hat{\xi}_1 = \hat{\xi}_2 = 0$ (i.e., the PGS case).

Case 2: $\mathcal{R}_2 = \mathcal{R}_{2,2}$

Here, ρ falls within the following range

$$0 \leq \rho \leq \tilde{\rho} < 1. \tag{4.28}$$

We can denote the sub-optimal solutions in *Case 2* as $\bar{\xi}_1, \bar{\xi}_2, \bar{\rho}$. As was noted in (A.7), $\mathcal{R}_{2,2}$ is only a function of ξ_2. On the other hand, $\mathcal{R}_1$ in (A.5) is an increasing function of ρ. As a result, in order to maximize (4.23) with respect to ρ, the upper bound of ρ (the highest value of ρ) in (4.27) (i.e. $\tilde{\rho}$) is the optimal solution of ρ in *Case 2*, ultimately implying that $\bar{\rho} = \tilde{\rho}$.

Next, to find the optimal value of ξ_1, in *Case 2*, we substitute $\mathcal{R}_2 = \mathcal{R}_{2,2}$ into (4.25), which leads to

$$-(\lambda_1 + w_1)(-2\xi_1(|h_1|^2\rho P)^2) = 0. \tag{4.29}$$

It is evident from (4.29) that the optimal value of ξ_1, in *case 2*, is zero. Thus, $\bar{\xi}_1 = 0$. Similar to what was seen in (4.25), the derivation with respect to ξ_2, leads to some candidate sub-optimal solutions for ξ_2 in *Case 2*, or $\bar{\xi}_2$ *(See* (A.11) *in Appendix A.2).* It should be mentioned that (A.11) is a 4^{th} degree polynomial and should have four roots, one of which is zero. The values that we consider to be valid are the ones that meet the constraints in $\mathcal{C}3$ and $\mathcal{C}4$.

We propose Algorithm I at the top of next page to solve the joint optimization problem in (4.22) and find the sub-optimal solutions for ξ_1^*, ξ_2^* and ρ^*. We consider both *Case 1* and *Case 2*, where the Lagrange multipliers λ_1 and λ_2 in ξ_1^*, ξ_2^*, and ρ^* are updated using the sub-gradient method [180]. After determining the solution in each case, we verify whether $\mathcal{R}_{1,2}$ or $\mathcal{R}_{2,2}$ was chosen as the minimum proceed to find the optimal solution for the sum rate with the corresponding case.

That said, following the KKT conditions above, we need to find the values of $\lambda_i, \forall i =$

Algorithm I: Weighted Sum Rate Maximization

INPUT: $R_1^{\min}, R_2^{\min}$, P, $\hat{P}$, κ, w_1, and w_2.

Solve the optimization problem (4.22) in two cases, i.e. when $\mathcal{R}_2 = \mathcal{R}_{1,2}$, and when $\mathcal{R}_2 = \mathcal{R}_{2,2}$,

Set $\lambda_1 = \lambda_2 = 0$ and find ξ_1^*, ξ_2^*, ρ^* in both cases.

- Case 1: $\mathcal{R}_2 = \mathcal{R}_{1,2}$, we maximize $\mathcal{R}_1 + \mathcal{R}_{1,2}$, and find $\hat{\xi}_1, \hat{\xi}_2, \hat{\rho}$ from jointly solving (A.8), (A.9), (A.10).
- Case 2: $\mathcal{R}_2 = \mathcal{R}_{2,2}$, we maximize $\mathcal{R}_1 + \mathcal{R}_{2,2}$, and find $\bar{\xi}_1, \bar{\xi}_2, \bar{\rho}$ from jointly solving (A.8), (A.11).
- only accepted candidates of $\hat{\xi}_1, \hat{\xi}_2, \hat{\rho}$ and $\bar{\xi}_1, \bar{\xi}_2, \bar{\rho}$ are the ones that satisfy $\mathcal{C}3$ and $\mathcal{C}4$ and choose the one that results in the maximum sum rate.

REPEAT: Until convergence, for each case, we continue to check:

if $\mathcal{R}_1 \geq R_1^{\min}$ and $\mathcal{R}_2 \geq R_2^{\min}$, **then**, the sub-optimal solution is reached.

else if $\mathcal{R}_1 < R_1^{\min}$ and $\mathcal{R}_2 \geq R_1^{\min}$, **then**, find non-negative λ_1 from (4.30) such that $\mathcal{R}_1(\xi_1^*, \xi_2^*, \rho^*) = R_1^{\min}$ and re-calculate ξ_1^*, ξ_2^*, ρ^* from the selected valid case. Repeat until convergence.

else if $\mathcal{R}_1 \geq R_1^{\min}$ and $\mathcal{R}_2 < R_1^{\min}$, **then**, find non-negative λ_2 from (4.30) such that $\mathcal{R}_2(\xi_1^*, \xi_2^*, \rho^*) = R_2^{\min}$ and re-calculate ξ_1^*, ξ_2^*, ρ^* from the selected valid case. Repeat until convergence.

else $\mathcal{R}_1 < R_1^{\min}$ and $\mathcal{R}_2 < R_1^{\min}$, **then**, find non-negative λ_1 and λ_2 from (4.30) if exists such that $\mathcal{R}_1(\xi_1^*, \xi_2^*, \rho^*) = R_1^{\min}$ and $\mathcal{R}_2(\xi_1^*, \xi_2^*, \rho^*) = R_2^{\min}$ and re-calculate ξ_1^*, ξ_2^*, ρ^* from the selected valid case. Repeat until convergence.

Check **if** $\mathcal{R}_{1,2}(\hat{\xi}_1, \hat{\xi}_2, \hat{\rho}) < \mathcal{R}_{2,2}(\hat{\xi}_1, \hat{\xi}_2, \hat{\rho})$, i.e. case 1 is valid and

Check **if** $\mathcal{R}_{2,2}(\bar{\xi}_1, \bar{\xi}_2, \bar{\rho}) < \mathcal{R}_{1,2}(\bar{\xi}_1, \bar{\xi}_2, \bar{\rho})$, i.e. case 2 is valid.

END

We have two options, check the region of ρ:

- Check **if** (4.27) is satisfied, *case 1* is valid and *case 2* is not valid, this leads to $\xi_1^*, \xi_2^*, \rho^* = \hat{\xi}_1, \hat{\xi}_2, \hat{\rho}$ is the optimal solution and then sum rate is equal $\mathcal{R}_1(\hat{\xi}_1, \hat{\xi}_2, \hat{\rho}) + \mathcal{R}_{1,2}(\hat{\xi}_1, \hat{\xi}_2, \hat{\rho})$.

- Check **if** (4.28) is satisfied, *case 2* is valid and *case 1*, this leads to $\xi_1^*, \xi_2^*, \rho^* = \bar{\xi}_1, \bar{\xi}_2, \bar{\rho}$ is the optimal solution and then sum rate is equal $\mathcal{R}_1(\bar{\xi}_1, \bar{\xi}_2, \bar{\rho}) + \mathcal{R}_{2,2}(\bar{\xi}_1, \bar{\xi}_2, \bar{\rho})$.

OUTPUT: ξ_1^*, ξ_2^*, ρ^*, and maximum sum-rate $\mathcal{R}_1(\xi_1^*, \xi_2^*, \rho^*) + \mathcal{R}_2(\xi_1^*, \xi_2^*, \rho^*)$.

$1, 2$. As can be seen from (4.24b) and (4.24c), we either have $\lambda_1 = 0$ or $\mathcal{R}_1(\xi_i^*, \rho^*) = R_1^{\min}$ and $\lambda_2 = 0$ or $\mathcal{R}_2(\xi_i^*, \rho^*) = R_2^{\min}$. This leads to four possible states:

– *State 1:* $\lambda_1 = \lambda_2 = 0$, a condition referring to inactive QoS constraints.

– *State 2:* $\lambda_1 = 0$ and $\lambda_2 \neq 0$, a condition that refers to the existence of the sub-optimal circularity coefficients and power parameter when $\mathcal{R}_2(\xi_i^*, \rho^*) = R_2^{\min}$.

– *State 3:* $\lambda_1 \neq 0$ and $\lambda_2 = 0$, a condition that refers to the existence of the sub-optimal circularity coefficients and power parameter when $\mathcal{R}_1(\xi_i^*, \rho^*) = R_1^{\min}$.

– *State 4:* $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$, a condition that, if feasible, refers to the existence of the sub-optimal circularity coefficients and power parameter when both $\mathcal{R}_1(\xi_i^*, \rho^*) = R_1^{\min}$ and $\mathcal{R}_2(\xi_i^*, \rho^*) = R_2^{\min}$.

The values of λ_1 and λ_2 in (4.23) can be computed using the subgradient method [180] as follows:

$$\lambda_i^{l+1} = \left[\lambda_i^l - s_i^l(\mathcal{R}_i - R_i^{\min})\right]^+, \quad \forall i = 1, 2, \tag{4.30}$$

where $[.]^+$ is defined as $\max(., 0)$ and s_i is a sufficiently small step size equal to $0.1/\sqrt{l}$ for l iterations [180].

4.4.2 Max-Min Rate Optimization

The max-min objective is a well-defined metric for system performance that guarantees fairness by providing effective rate balance across the different users. In contrast to a max-sum situation, a solution to the max-min problem does not accommodate multiple privileged users (i.e. users with stronger channels that can monopolize available resources).

To jointly find optimal power allocation and circularity coefficients that aim to maximize the minimum rate between $\{\mathcal{R}_1(\xi_i, \rho), \mathcal{R}_2(\xi_i, \rho)\}$ with constraints in ξ_i and ρ, we

formulate the following optimization problem:

$$\mathcal{OP}2: \max_{\xi_i,\rho} \min \quad \{\mathcal{R}_1(\xi_i,\rho), \mathcal{R}_2(\xi_i,\rho)\} \tag{4.31a}$$

$$\text{subject to} \quad \mathcal{C}1: \ 0 \leq \xi_i \leq 1, \forall i = 1,2 \tag{4.31b}$$

$$\mathcal{C}2: \ 0 \leq \rho \leq 1, \tag{4.31c}$$

where $\mathcal{R}_2(\xi_i,\rho) = \min\left[\mathcal{R}_{1,2}(\xi_i,\rho), \mathcal{R}_{2,2}(\xi_i,\rho)\right]$. Due to the non-concavity of the problem objective, the optimization problem (4.31) is non-convex. To solve this problem, we introduce a slack variable $\mathcal{X}$ and appropriately transform $\mathcal{OP}2$ into the following maximization problem:

$$\mathcal{OP}3: \max_{\xi_i,\rho,\mathcal{X}} \quad \mathcal{X} \tag{4.32a}$$

$$\text{subject to} \quad \mathcal{C}1: \ 0 \leq \xi_i \leq 1, \forall i = 1,2, \tag{4.32b}$$

$$\mathcal{C}2: \ 0 \leq \rho \leq 1, \tag{4.32c}$$

$$\mathcal{C}3: \ 0 \leq \mathcal{X} \leq \mathcal{R}_i(\xi_i,\rho). \tag{4.32d}$$

We can write the Lagrangian function of (4.32) as follows:

$$\mathcal{L}_{\mathcal{X}} = -\mathcal{X} + \lambda_1(\mathcal{X} - \mathcal{R}_1(\xi_i,\rho) + \lambda_2(\mathcal{X} - \mathcal{R}_2(\xi_i,\rho)). \tag{4.33}$$

Using the predefined KKT conditions, we can find sub-optimal solution that maximizes the minimum between the two rates. Following the same approach as in the sum-rate maximization problem, we can consider the two cases of $\mathcal{R}_2$ to analyze the max-min rate problem. This way, we will have two regions with two optimization problems.

To solve these two optimization problems, we first define the range of ρ for each case. Within each region of ρ, we derive the sub-optimal solutions of ξ_1, ξ_2, and ρ by deriving (4.33) for each case and simultaneously solving the corresponding equations. At the end of the algorithm, we have to check if the assumption we made (i.e. $\mathcal{R}_2 = \mathcal{R}_{1,2}$ or $\mathcal{R}_{2,2}$) is

still valid.

In the first region, from (4.33), the derivation with respect to ρ, ξ_i is given in *Appendix A.3*. We simultaneously solve (A.12)-(A.14) to find optimal values that satisfy constraints $\mathcal{C}1$ and $\mathcal{C}2$ in (4.32). For the second region, by deriving $\mathcal{L}_{\mathcal{X}}$ with respect to ξ_i, we can find that $\bar{\xi}_1 = 0$, and optimal $\bar{\xi}_2$ can be found from (A.15) outlined in *Appendix A.4* using sub-gradient method.

At the next page, we illustrate the proposed algorithm to solve the max-min rate problem that can determine the solutions for the above two cases and then verify the chosen minimum between $\mathcal{R}_{1,2}$ and $\mathcal{R}_{2,2}$.

4.4.3 Energy Efficiency

In this subsection, we maximize the energy efficiency of the cooperative FD-NOMA IGS-based system considering both QoS and BS power constraints.

$$\mathcal{OP}4 : \underset{\xi_i,\rho_i}{\text{maximize}} \quad \eta_{EE} = \frac{\mathcal{R}_1(\xi_i,\rho_i) + \mathcal{R}_2(\xi_i,\rho_i)}{P_c + (\rho_1+\rho_2)P + \hat{P}} \tag{4.34a}$$

$$\text{subject to} \quad \mathcal{C}1: \ \mathcal{R}_1(\xi_i,\rho_i) \geq R_1^{\min}, \tag{4.34b}$$

$$\mathcal{C}2: \ \mathcal{R}_2(\xi_i,\rho_i) \geq R_2^{\min}, \tag{4.34c}$$

$$\mathcal{C}3: \ \rho_1 + \rho_2 \leq 1, \tag{4.34d}$$

$$\mathcal{C}4: \ 0 \leq \rho_i \leq 1, \forall i = 1,2, \tag{4.34e}$$

$$\mathcal{C}5: \ 0 \leq \xi_i \leq 1, \forall i = 1,2, \tag{4.34f}$$

where P_c is the circuitry power consumption. The optimization problem in (4.34a) is equivalent to the following minimization problem $\mathcal{OP}5$:

$$\mathcal{OP}5 : \underset{\xi_i,\rho}{\text{minimize}} \ \eta_{EE}^{-1}, \quad \text{subject to} \quad \mathcal{C}1 - \mathcal{C}5. \tag{4.35}$$

Algorithm II: Max-Min Rate:

1: **INPUT:** P and $\hat{P}$.
2: We solve the max-min optimization problem (4.31) in two cases, i.e. when $\mathcal{R}_2 = \mathcal{R}_{1,2}$, and when $\mathcal{R}_2 = \mathcal{R}_{2,2}$,

- Case 1: $\mathcal{R}_2 = \mathcal{R}_{1,2}$, we maximize the minimum of $(\mathcal{R}_1, \mathcal{R}_{1,2})$, and find $\hat{\xi}_1, \hat{\xi}_2, \hat{\rho}$ from jointly solving (A.12), (A.13), (A.14).
- Case 2: $\mathcal{R}_2 = \mathcal{R}_{2,2}$, we maximize the minimum of $(\mathcal{R}_1, \mathcal{R}_{2,2})$, and find $\bar{\xi}_1, \bar{\xi}_2, \bar{\rho}$ from jointly solving (A.12), (A.15).
- only accepted candidates of $\hat{\xi}_1, \hat{\xi}_2, \hat{\rho}$ and $\bar{\xi}_1, \bar{\xi}_2, \bar{\rho}$ are the ones that satisfy $\mathcal{C}3$ and $\mathcal{C}4$ and choose the one that results in the maximum sum rate.

3: **REPEAT:**For each case, we continue to check **if** $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$, **then**, find non-negative λ_i from (4.30) such that $\mathcal{R}_1 = \mathcal{R}_2 = \mathcal{X}$ and the sub-optimal solution ξ_1^*, ξ_2^*, ρ^* is reached.
4: **else if** $\lambda_1 > 0$ and $\lambda_2 = 0$, **then**, find non-negative λ_1 from (4.30) such that $\mathcal{R}_1(\xi_1^*, \xi_2^*, \rho^*)$ is the minimum rate to maximized and re-calculate ξ_1^*, ξ_2^*, ρ^* from the selected valid case. Repeat until convergence.
5: **else if** $\lambda_1 = 0$ and $\lambda_2 > 0$, **then**, find non-negative λ_2 from (4.30) such that $\mathcal{R}_2(\xi_1^*, \xi_2^*, \rho^*)$ is the minimum rate to be maximized and re-calculate ξ_1^*, ξ_2^*, ρ^* from the selected valid case. Repeat until convergence.
6: **else** $\lambda_1 = 0$, and $\lambda_2 = 0$,, **then**, this case is not feasible, Repeat until convergence.
7: **END**
8: Check **if** $\mathcal{R}_{1,2}(\hat{\xi}_1, \hat{\xi}_2, \hat{\rho}) < \mathcal{R}_{2,2}(\hat{\xi}_1, \hat{\xi}_2, \hat{\rho})$, i.e. case 1 is valid and
9: Check **if** $\mathcal{R}_{2,2}(\bar{\xi}_1, \bar{\xi}_2, \bar{\rho}) < \mathcal{R}_{1,2}(\bar{\xi}_1, \bar{\xi}_2, \bar{\rho})$, i.e. case 2 is valid.
10: We have two options:
- Case 1 is valid and Case 2 is not valid, this leads to $\xi_1^*, \xi_2^*, \rho^* = \hat{\xi}_1, \hat{\xi}_2, \hat{\rho}$ is the optimal solution and then max-min throughput is equal max $\min(\mathcal{R}_1(\hat{\xi}_1, \hat{\xi}_2, \hat{\rho}), \mathcal{R}_{1,2}(\hat{\xi}_1, \hat{\xi}_2, \hat{\rho}))$.
- Case 2 is valid and Case 1 is not valid, this leads to $\xi_1^*, \xi_2^*, \rho^* = \bar{\xi}_1, \bar{\xi}_2, \bar{\rho}$ is the optimal solution and then max-min throughput is equal max $\min(\mathcal{R}_1(\bar{\xi}_1, \bar{\xi}_2, \bar{\rho}), \mathcal{R}_{2,2}(\bar{\xi}_1, \bar{\xi}_2, \bar{\rho}))$.

11: **OUTPUT:** ξ_1^*, ξ_2^*, ρ^*, and calculate the max-min throughput from max-$\min(\mathcal{R}_1(\xi_1^*, \xi_2^*, \rho^*), \mathcal{R}_2(\xi_1^*, \xi_2^*, \rho^*))$.

The objective function of (4.35) is non-convex; hence, the overall problem is non-convex, and the global optimal solution cannot be guaranteed.

The fractional non-convex optimization problem in (4.35) can be transformed to an equivalent parametric optimization problem using concepts from fractional programming, namely the Dinkelbach approach [181]. Using this transformation, a new objective function can be defined as

$$\mathcal{L}_{EE} = (P_c + (\rho_1 + \rho_2)P + \hat{P}) - q(\mathcal{R}_1(\xi_i, \rho_i) + \mathcal{R}_2(\xi_i, \rho_i)), \tag{4.36}$$

where q is a non-negative constant. Then, the new optimization problem $\mathcal{OP}6$ becomes

$$\mathcal{OP}6 : \underset{\xi_i,\rho_i}{\text{minimize}} \quad \mathcal{L}_{EE}, \quad \text{subject to} \quad \mathcal{C}1 - \mathcal{C}5. \tag{4.37}$$

It was shown in [181] that at a certain value of q, which is denoted by q^*, the optimal solution of $\mathcal{OP}6$ is also the optimal solution to $\mathcal{OP}5$. Hence, finding the optimal values of ξ_i, ρ_i for $\mathcal{OP}5$ can be reached by finding the optimal values of $(\xi_i(q), \rho_i(q))$ for $\mathcal{OP}6$. We can then update the value of q until it reaches q^*, where q^* is found by assigning $\mathcal{L}_{EE} = 0$ [181] at optimal ξ_i and ρ_i.

To find the sub-optimal solutions, we solve $\mathcal{OP}6$ using the KKT conditions. The Lagrangian function can be expressed as

$$\begin{aligned}\mathcal{L}_{EE}(\xi_i, \rho_i) = (P_c + (\rho_1+\rho_2)P + \hat{P}) - q\left(\mathcal{R}_1(\xi_i, \rho_i) + \mathcal{R}_2(\xi_i, \rho_i)\right) + \lambda_1(R_1^{\min} - \mathcal{R}_1(\xi_i, \rho_i)) \\ + \lambda_2(R_2^{\min} - \mathcal{R}_2(\xi_i, \rho_i)) + \lambda_3(\rho_1 + \rho_2 - 1),\end{aligned} \tag{4.38}$$

where λ_1, λ_2 and λ_3 are the Lagrange multipliers associated with the QoS conditions of U1 and U2, and power allocation at the BS, respectively. The impact of the constraints $\mathcal{C}4$ and $\mathcal{C}5$ will be considered later. The KKT conditions can consequently be written as

follows:

$$\frac{\partial \mathcal{L}_{EE}(\xi_i^*, \rho_i^*)}{\partial \xi_i, \partial \rho_i} = 0, \qquad \forall i = 1, 2, \tag{4.39a}$$

$$\lambda_1(R_1^{\min} - \mathcal{R}_1(\xi_i^*, \rho^*)) = 0, \tag{4.39b}$$

$$\lambda_2(R_2^{\min} - \mathcal{R}_2(\xi_i^*, \rho^*)) = 0, \tag{4.39c}$$

$$\lambda_3(\rho_1 + \rho_2 - 1) = 0, \tag{4.39d}$$

$$R_1^{\min} - \mathcal{R}_1(\xi_i^*, \rho^*) \leq 0, \tag{4.39e}$$

$$R_2^{\min} - \mathcal{R}_2(\xi_i^*, \rho^*) \leq 0, \tag{4.39f}$$

$$\rho_1 + \rho_2 - 1 \leq 0, \tag{4.39g}$$

$$\lambda_1, \lambda_2, \lambda_3 \geq 0. \tag{4.39h}$$

From (4.39a), we obtain

$$\frac{\partial \mathcal{L}_{EE}(\xi_i^*, \rho_i^*)}{\partial \xi_i} = -(\lambda_1 + q)\frac{\partial \mathcal{R}_1(\xi_i, \rho_i)}{\partial \xi_i} - (\lambda_2 + q)\frac{\partial \mathcal{R}_2(\xi_i, \rho_i)}{\partial \xi_i} = 0, \tag{4.40}$$

and

$$\frac{\partial \mathcal{L}_{EE}(\xi_i^*, \rho_i^*)}{\partial \rho_i} = P - (\lambda_1 + q)\frac{\partial \mathcal{R}_1(\xi_i, \rho_i)}{\partial \rho_i} - (\lambda_2 + q)\frac{\partial \mathcal{R}_2(\xi_i, \rho_i)}{\partial \rho_i} + \lambda_3 = 0. \tag{4.41}$$

Following the same steps seen in the sum rate maximization section, we simultaneously solve the results from (4.40)-(4.41) to find the optimal values of ξ_i, ρ_i.

Based on the FD-relaying concept, $\mathcal{R}_2$ is the minimum between $\mathcal{R}_{1,2}$ and $\mathcal{R}_{2,2}$. Hence, to be able to solve the optimization problem, we consider two scenarios: 1) *Case 1*: We assume $\mathcal{R}_2 = \mathcal{R}_{1,2}$, implying that we aim to maximize $\mathcal{R}_1 + \mathcal{R}_{1,2}$ and 2) *Case 2*: We assume $\mathcal{R}_2 = \mathcal{R}_{2,2}$, implying that we aim to maximize $\mathcal{R}_1 + \mathcal{R}_{2,2}$.

First, we find the range of ρ_i for each case in which either $\mathcal{R}_{1,2}$ or $\mathcal{R}_{2,2}$ is the minimum of $\mathcal{R}_2$. Second, within each region of ρ_i, we derive the sub-optimal solutions of ξ_1, ξ_2 and ρ_i by solving the corresponding equations in (4.40) and (4.41) for each case. Third, we

verify the minimum between $\mathcal{R}_{1,2}$ and $\mathcal{R}_{2,2}$ and choose the appropriate case in order to find the sub-optimal solution that will be used to maximize the energy efficiency. The region of ρ_i, for which either $\mathcal{R}_{1,2}$ or $\mathcal{R}_{2,2}$ is the minimum, is derived assuming $\mathcal{R}_{2,2} < \mathcal{R}_{1,2}$, after some mathematical manipulations, based on the following condition

$$\rho_2 > \frac{-((\rho_1 P\Gamma_1)(1-\xi_1\xi_2)+\kappa\hat{P}\Gamma_s+1)+\sqrt{Z}}{P\Gamma_1(1-\zeta_2^2)}, \tag{4.42}$$

where $Z = (\rho_1 P\Gamma_1 + \kappa\hat{P}\Gamma_s + 1)^2\left[(\hat{P}\Gamma_2+1)^2 - \zeta_2^2(\hat{P}\Gamma_2)^2(2-\zeta_2^2) - 2\zeta_2^2(\hat{P}\Gamma_2)\right] + (1-\zeta_2^2)((\hat{P}\Gamma_2+1)^2-\zeta_2^2(\hat{P}\Gamma_2)^2-1)((\kappa\hat{P}\Gamma_s\zeta_2)^2+(\rho_1 P\Gamma_1\zeta_1)^2)-2(\rho_1 P\Gamma_1+\kappa\hat{P}\Gamma_s+1)(\rho_1 P\Gamma_1\xi_1\xi_2)+(\rho_1 P\Gamma_1\xi_1\xi_2)^2$. If the condition in (4.42) is achieved, this leads to *Case 2*, i.e. $\mathcal{R}_2 = \mathcal{R}_{2,2}$, otherwise, $\mathcal{R}_2 = \mathcal{R}_{1,2}$.

We propose the following algorithm at the top of next page to solve the joint optimization problem in (4.34a) and find the sub-optimal solutions for ξ_1^*, ξ_2^* and ρ_i^*. Assuming that $\mathcal{C}1-\mathcal{C}3$ are satisfied, we start by simultaneously solving (4.40)-(4.41) numerically using the Gauss-Newton algorithm [180] to find the optimal values of ξ_i, ρ_i and ensure that $\mathcal{C}4-\mathcal{C}5$ are guaranteed. We consider both *Cases 1* and *2*, where the Lagrange multipliers $\lambda_i, \forall i = 1,2,3$ in ξ_1^*, ξ_2^*, and ρ^* are updated using the sub-gradient method [180]. After determining the solution in each case, we verify whether $\mathcal{R}_{1,2}$ or $\mathcal{R}_{2,2}$ was chosen as the minimum and proceed to find the optimal solution for the energy efficiency with the chosen case.

That said, following the KKT conditions above, we need to find the values of $\lambda_i, \forall i = 1,2,3$. As can be seen from (4.39b) and (4.39c), we either have $\lambda_1 = 0$ or $\mathcal{R}_1(\xi_i^*, \rho^*) = R_1^{\min}$, and $\lambda_2 = 0$ or $\mathcal{R}_2(\xi_i^*, \rho^*) = R_2^{\min}$, and $\lambda_3 = 0$ or $\rho_1 + \rho_2 = 1$. This leads to eight possible states which we need to iterate among them till reaching the optimal solution.

We set $\lambda_1 = \lambda_2 = \lambda_3 = 0$ and find ξ_i^*, ρ_i^* in both *Case 1* and *Case 2* from jointly solving (4.40), (4.41). If $\mathcal{R}_1 \geq R_1^{\min}$, $\mathcal{R}_2 \geq R_2^{\min}$, and power constraint is true, then, the sub-optimal solution is reached.

If $\mathcal{R}_1 < R_1^{\min}$, $\mathcal{R}_2 \geq R_1^{\min}$, and power constraint is true, then, find non-negative

Algorithm III: Energy Efficiency Maximization Algorithm:

1: **INPUT:** $R_1^{\min}, R_2^{\min}$, P, $\hat{P}$, δ, $q_{initial}$, and $\mathcal{L}_{EE_{min}} = -\infty$
2: **While:** $\mathcal{L}_{EE_{min}} < -\delta$, **do**
3: Find the values of ξ_i^*, ρ_i^* as in **Algorithm I**
4: Update $\mathcal{L}_{EE}$ from (4.36) and calculate $\mathcal{L}_{EE_{min}}(q)$,
5: Calculate $q = \frac{P_c+(\rho_1^*+\rho_2^*)P+\hat{P}}{(\mathcal{R}_1(\xi_i^*,\rho_i^*)+\mathcal{R}_2(\xi_i^*,\rho_i^*))}$
6: **end While**
7: **OUTPUT:** q^*, and ξ_i^*, ρ^*.

λ_1 such that $\mathcal{R}_1(\xi_i^*, \rho_i^*) = R_1^{\min}$ and re-calculate ξ_i^*, ρ_i^* from the selected valid case. If $\mathcal{R}_1 \geq R_1^{\min}$, $\mathcal{R}_2 \geq R_1^{\min}$, and power constraint is violated, then, find non-negative λ_3 such that $\rho_1 + \rho_2 = 1$ and re-calculate ξ_i^*, ρ_i^* from the selected valid case. Repeat until convergence.

If $\mathcal{R}_1 < R_1^{\min}$, $\mathcal{R}_2 \geq R_1^{\min}$, and power constraint is violated, then, find non-negative λ_1, λ_3 such that $\mathcal{R}_1(\xi_i^*, \rho_i^*) = R_1^{\min}$ and $\rho_1 + \rho_2 = 1$ and re-calculate ξ_i^*, ρ_i^* from the selected valid case. If $\mathcal{R}_1 \geq R_1^{\min}$, $\mathcal{R}_2 < R_1^{\min}$, and power constraint is violated, then, find non-negative λ_2, λ_3 such that $\mathcal{R}_2(\xi_i^*, \rho_i^*) = R_2^{\min}$ and $\rho_1 + \rho_2 = 1$ and re-calculate ξ_i^*, ρ_i^* from the selected valid case. If $\mathcal{R}_1 < R_1^{\min}$, $\mathcal{R}_2 < R_1^{\min}$, and power constraint is violated, then, find non-negative λ_1, λ_2 and λ_3 if exists such that $\mathcal{R}_1(\xi_i^*, \rho_i^*) = R_1^{\min}$, $\mathcal{R}_2(\xi_i^*, \rho_i^*) = R_2^{\min}$, $\rho_1 + \rho_2 = 1$ and re-calculate ξ_i^*, ρ_i^* from the selected valid case. We check if $\mathcal{R}_{1,2}(\hat{\xi}_i, \hat{\rho}_i)$ $< \mathcal{R}_{2,2}(\hat{\xi}_i, \hat{\rho}_i)$, and the region of ρ_i in (4.42) is not satisfied, i.e., *Case 1* is valid and if $\mathcal{R}_{2,2}(\bar{\xi}_i, \bar{\rho}_i) < \mathcal{R}_{1,2}(\bar{\xi}_i, \bar{\rho}_i)$ and the region of ρ_i in (4.42) is satisfied,, i.e., *Case 2* is valid.

The energy efficiency solution is found at $q = q^*$, where $q = \frac{P_c+(\rho_1^*+\rho_2^*)P+\hat{P}}{(\mathcal{R}_1(\xi_i^*,\rho_i^*)+\mathcal{R}_2(\xi_i^*,\rho_i^*))}$ and is computed by applying the Dinkelbach method [181]. Here, we develop Algorithm III, which follows an approach similar to Algorithm I with the aim of computing the optimal values of ξ_i^*, ρ_i^* that satisfy $\mathcal{L}_{EE_{min}}(\xi_i^*, \rho_i^*) = 0$, where $\mathcal{L}_{EE_{min}}$ is the minimum of $\mathcal{L}_{EE}$. Algorithm III starts with an initial value of q, denoted as q_{initial}, and employs an error tolerance of δ. This energy efficiency algorithm is outlined as follows.

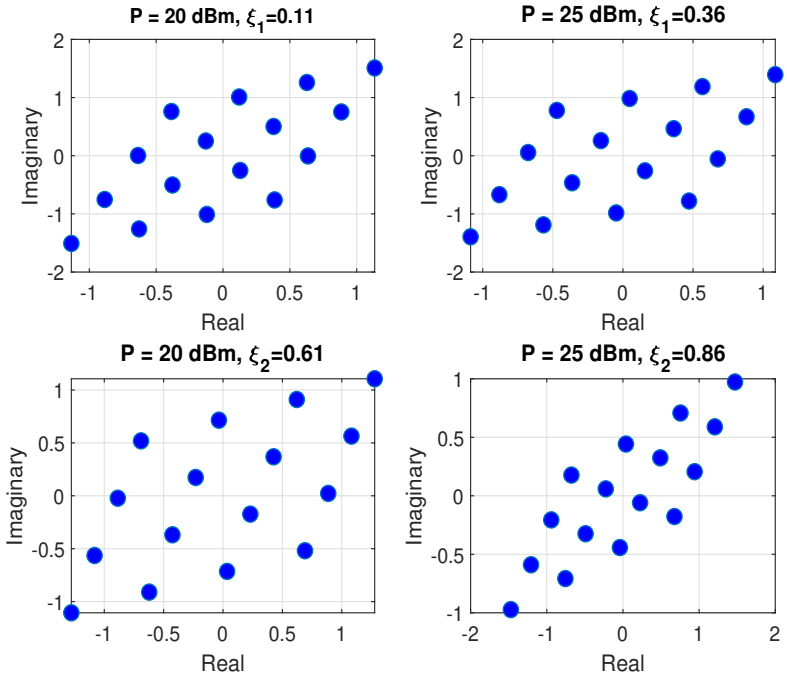

Figure 4.2: 16-QAM improper constellation diagrams with different ξ values at P =20 , 25 dBm, $\hat{P} = 0.3P$.

4.5 Simulation Results and Discussion

In this section, we present the simulation results obtained upon solving the joint optimization problems of the circularity coefficients and power allocation. Recall that the goals were to maximize the sum-rate, minimum rate (fairness) and the energy efficiency of a cooperative FD-NOMA IGS-based system. To demonstrate the effectiveness of the proposed schemes, we compare them with their cooperative FD-NOMA PGS-based counterparts.

The simulation results are obtained through Monte Carlo trials for 10^4 channel realizations. We consider the distance-dependent path-loss model as the large scale fading, where the path-loss exponent is 3, and Rayleigh fading as the small scale multi-path fading. Unless otherwise mentioned, we assume that $R_1^{\min} = R_2^{\min} = R^{\min} = 0.5$ bits/sec/Hz, $\hat{P} = 0.5P$, P ranges from 10-40 dBm, $\kappa = 1$, $\mathcal{E}[|h_1|^2] > \mathcal{E}[|h_2|^2]$, $P_c = 20$ dBm, and $\delta = 10^{-7}$.

Table 4.1: Comparison of sum rates using improper and proper constellations at different P values

Sum rate / Power	Proper bits/sec/Hz	Improper bits/sec/Hz	ξ_1^*	ξ_2^*
$P = 20$ dBm	2.54	3.1	0.11	0.61
$P = 25$ dBm	2.75	3.25	0.36	0.86

Fig. 4.2 shows the improper constellations with 16-QAM designed with prescribed procedures in Section II at a given instantaneous channel gain when $\Gamma_1 = 1$, $\Gamma_2 = 0.25, \Gamma_s = 0.35$, and $P = 20, 25$ dBm. The optimal ξ_1, ξ_2 are found by the proposed algorithm I. It can be seen that the minimum distance between constellations points decreases in the improper constellations compared to the proper constellations as ξ_i increases. Due to this impropriety, the throughput enhancement is achieved as proved in this work and also it has been shown that the error probability decreases as in [24], [146]. We further analyze the sum rate values achieved by the so-designed improper constellation diagrams as in Table. 4.1. As can be read, the improper constellation offers considerable sum rate gain compared to proper one. For instance, $P = 20$ dBm, the proper sum rate was equal to 2.54 bits/sec/Hz while the improper sum rate 3.1 bits/sec/Hz.

In Fig. 4.3, the sum rate is plotted against the BS transmit power P (dBm) at different values of $\hat{P}$, fixed $\kappa = 1$, and $R^{\min} = 0.5$ bits/sec/Hz. The sum rate is compared between the FD-NOMA IGS-based system considering the joint optimization of circularity coefficients and power allocation (Algorithm I) and its PGS-based counterpart. As shown in the figure, as P increases, sum rate improves for both systems as expected. The figure also shows the significant improvement of NOMA-IGS sum rate over that of NOMA-PGS from low to high values of P. This illustrates the superiority of IGS over PGS in the context of limited interference systems. We notice also that as $\hat{P}$ increases, the sum rate curves decreases, which can also be seen from (A.5) and (4.14).

Fig. 4.4 illustrates the impact of IGS when the self-interference cancellation is imperfect. The figure shows the sum rate performance at various values of κ at fixed $\hat{P}$. As

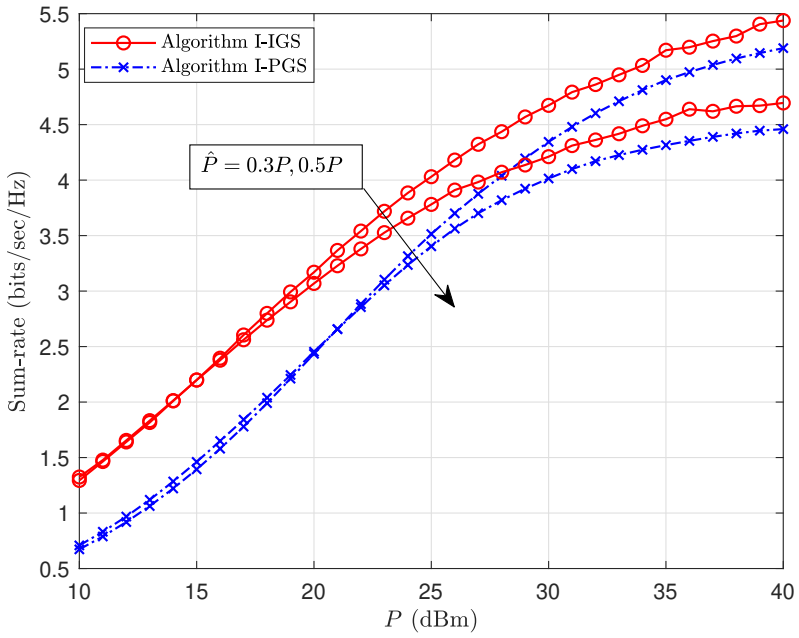

Figure 4.3: Sum rate vs P (dBm) for proposed NOMA-IGS compared with NOMA-PGS at different values of $\hat{P}$ and fixed $\kappa = 1$.

κ increases from 0.3 to 0.7, the improvement due to the IGS when compared with PGS increases where around one to two dBs gain is achieved. This justifies the importance of using IGS in limited-interference systems like cooperative FD-NOMA.

Fig. 4.5 presents the performance of Algorithm I for the weighted sum rate with QoS constraints imposed for both IGS-based and PGS-based systems. We assume $w_1 = 0.9, w_2 = 0.1$, which are the weights of users' rates. The result is a higher priority being given to U1 when compared to $w_1 = 0.5, w_2 = 0.5$, ultimately improving overall weighted sum rate. When compared to the PGS-based scheme, a considerable sum rate gain is achieved across the entire power region. In addition, Fig. 4.5 compares the proposed sub-optimal solutions of the weighted sum rate maximization with the exact optimal solution through an exhaustive search at $w_1 = 0.9, w_2 = 0.1$, where the latter is executed through three nested loops with a step size of 0.05 (the three optimization variables are all bounded between 0 and 1). This simulation shows a small gap of around 1 dB between the two solutions at low values of P.

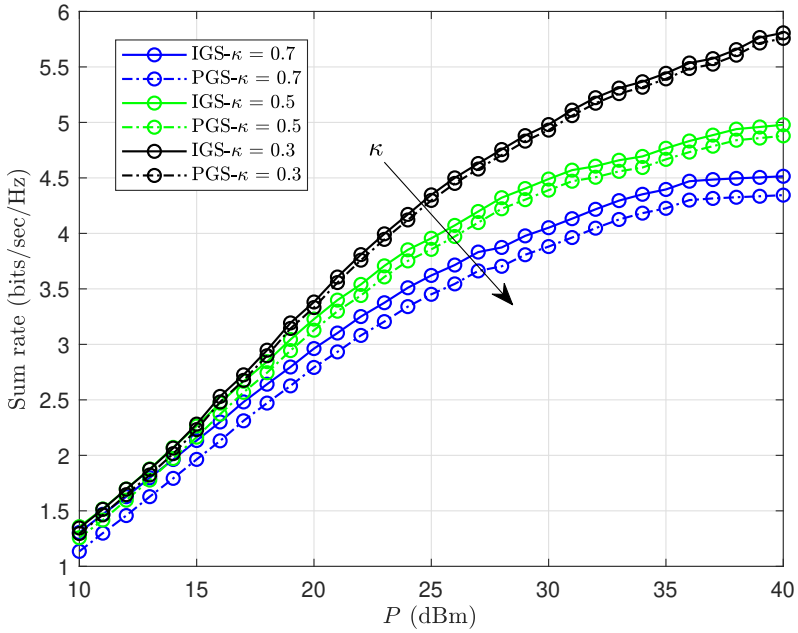

Figure 4.4: Sum rate vs P (dBm) for both IGS and PGS schemes at different values of κ and fixed value of $\hat{P} = 0.5P$.

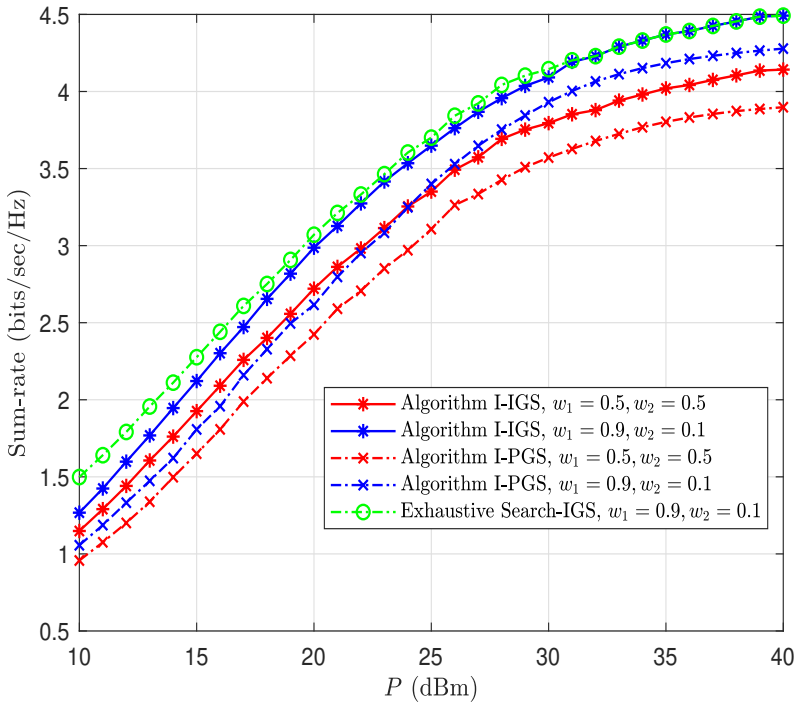

Figure 4.5: Comparison between proposed scheme and exhaustive search of weighted sum rate vs P dBm at different weights $w_1 = 0.9, w_2 = 0.1$ and $w_1 = 0.5, w_2 = 0.5$.

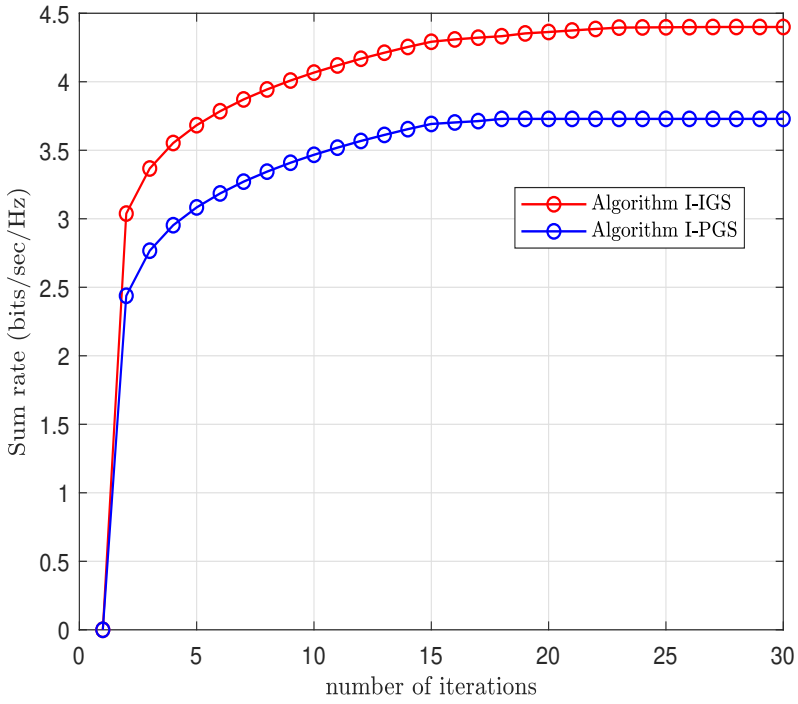

Figure 4.6: Convergence of Algorithm I for both IGS and PGS schemes.

For Algorithm I convergence, we show in Fig. 4.6 that the Algorithm I-IGS case needs around 20-25 iterations to converge while only 15 iterations are needed for PGS. This is due to fact that IGS needs to find the optimal values of ξ_i and ρ_i while PGS only needs to find the optimal value of ρ_i (since $\xi_i^* = 0$ when considering PGS).

Fig. 4.7 represents the results of the max-min optimization problem. We perform a comparison between IGS-based and PGS-based systems using Algorithm II. The improvement of the IGS-based max-min throughput outperforms its PGS-based counterpart. It is also observed from the figure that fairness is achieved between two users rates since we see that the gap between the rates of the two users shrinks using Algorithm II.

In Fig. 4.8, the overall sum-rate from Algorithm I, max-min sum rate from Algorithm II, $\mathcal{R}_1$ and $\mathcal{R}_2$ are simulated against P , where we assume $R_1^{\min} = R_2^{\min} = R_{min}$. The overall sum-rate of the system using Algorithm I is higher than overall max-min sum rate but at the expense of user fairness (which was achieved from Algorithm II). It is

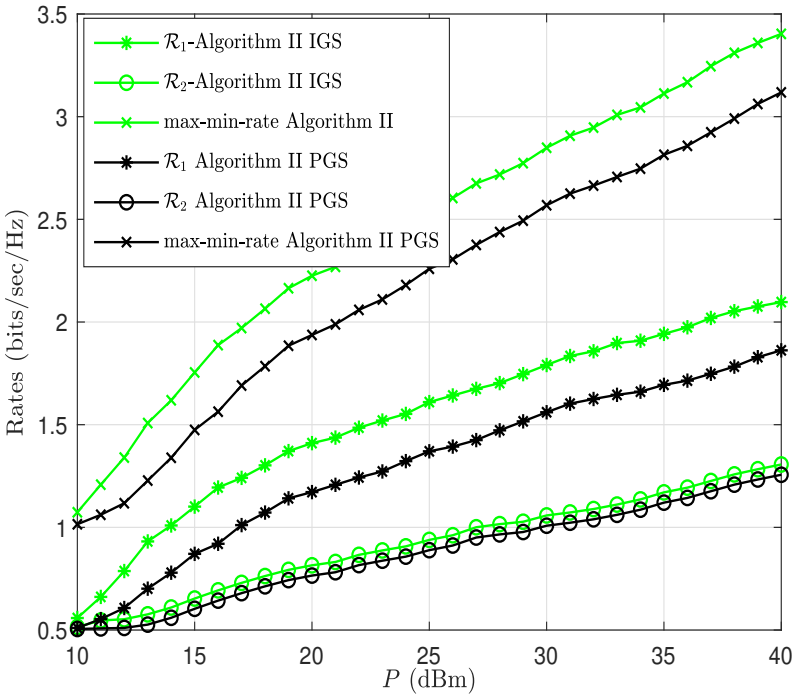

Figure 4.7: Comparison between max-min rates using Algorithm II in both IGS and PGS cases.

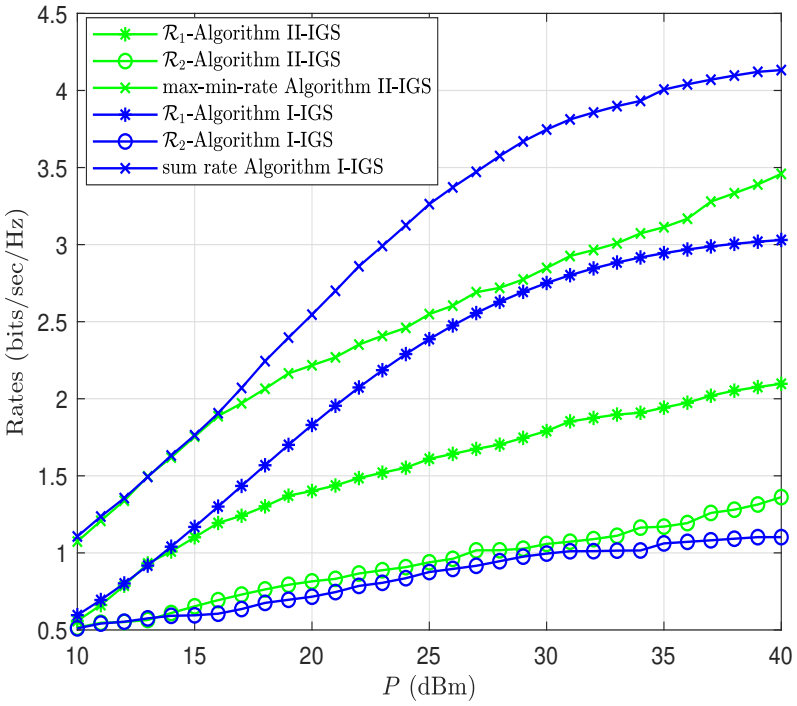

Figure 4.8: Comparison between users' rates and sum rates using Algorithms I and II.

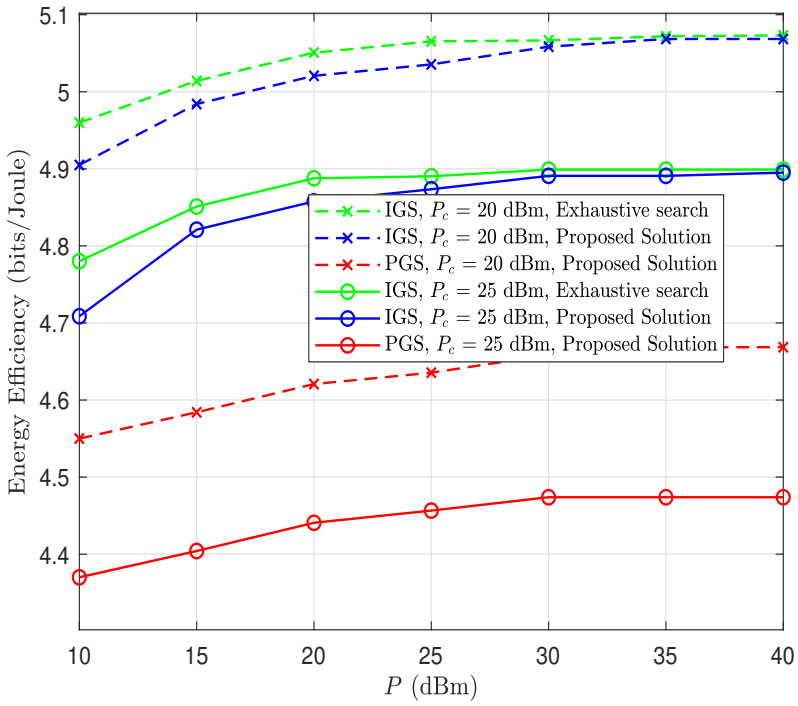

Figure 4.9: Energy efficiency vs P (dBm) at different values of P_c and fixed $\kappa = 1$ for proposed Algorithm III and exhaustive search.

readily noticed that the minimum rate is maximized in Algorithm II when compared to the minimum rate in Algorithm I, which was the main purpose behind implementing Algorithm II.

In Fig. 4.9, the energy efficiency (in bits/joule) is simulated against P (dBm) at $P_c =$ 20 , 25 dBm and $\kappa = 1$. It can be observed from this figure that IGS enhances the energy efficiency performance in the proposed system by transmitting around 0.5 bits/joule more than PGS in all BS power regions. It is also observed that as the BS's power increases from low to medium, the energy efficiency increases. However, energy efficiency decreases when the BS's transmit power consumption P_c increases because the increase in the transmit power is larger than the corresponding increase in the transmit rate. This means that increasing the transmit power does not necessarily enhance the energy efficiency.

To demonstrate closeness to the optimal solution, we compared the sub-optimal solu-

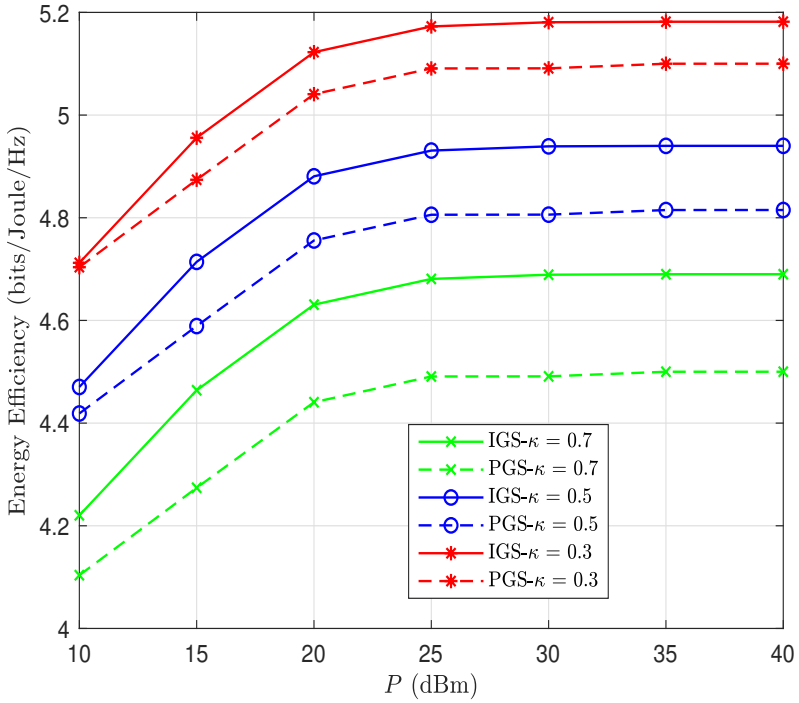

Figure 4.10: Energy efficiency vs P (dBm) at different values of κ and fixed $P_c = 20$ dBm.

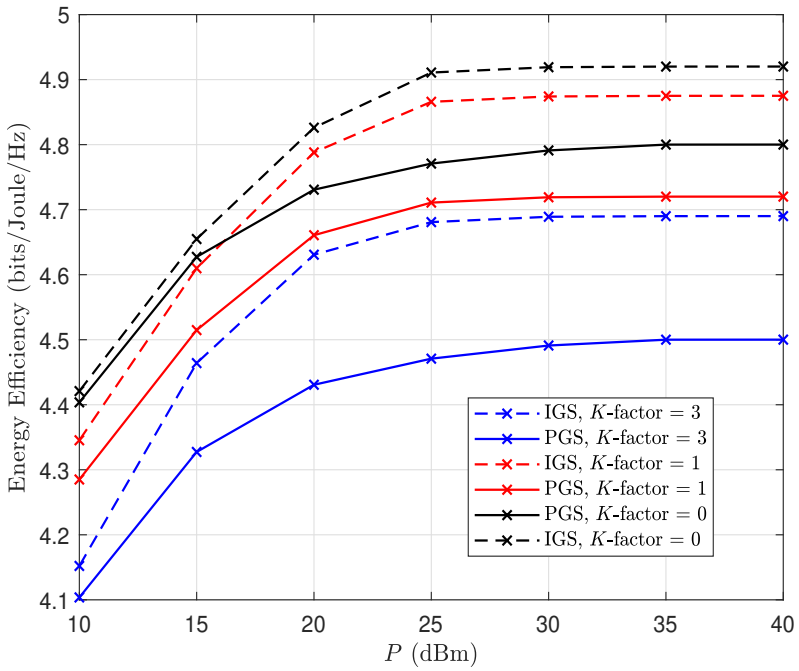

Figure 4.11: Energy efficiency vs P (dBm) at different values of K-Rician factor and fixed $\kappa = 0.5$.

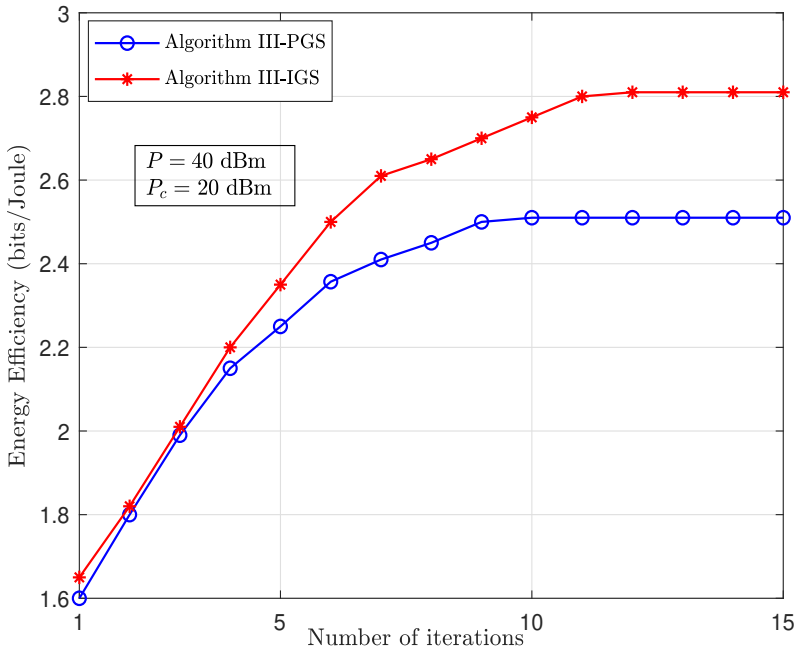

Figure 4.12: Convergence of Algorithm III at $P = 40$ dBm and P_c= 20 dBm for both IGS and PGS cases

tion with optimal solution through an exhaustive search. The result isa gap of around of 0.2-0.3 dB, which is acceptable performance loss given the impractical complexity of exhaustive search.

Figure 4.10 shows the energy efficiency performance against P dBm at different levels of κ and at fixed $P_c = 20$ dBm. The figure shows the effectiveness of using IGS in the case of self-interference existence when compared to the PGS, case especially at higher levels of κ. As κ increases, the gain due to using IGS becomes greater.

In Fig. 4.11, we consider a more realistic assumption where the self-interference channel is Rician fading and modeled as $h_s \sim \mathcal{CN}\left(\sqrt{\frac{\kappa K}{1+K}}, \frac{\kappa}{1+K}\right)$, where K is the Rician factor. We plot the energy efficiency versus P at different values of self-interference by changing the by changing the average mean of the interference channel. As K increases, i.e., level of interference is higher, the gain of IGS becomes better compared with PGS scheme, which proves that IGS is more effective in case of severe self-interference.

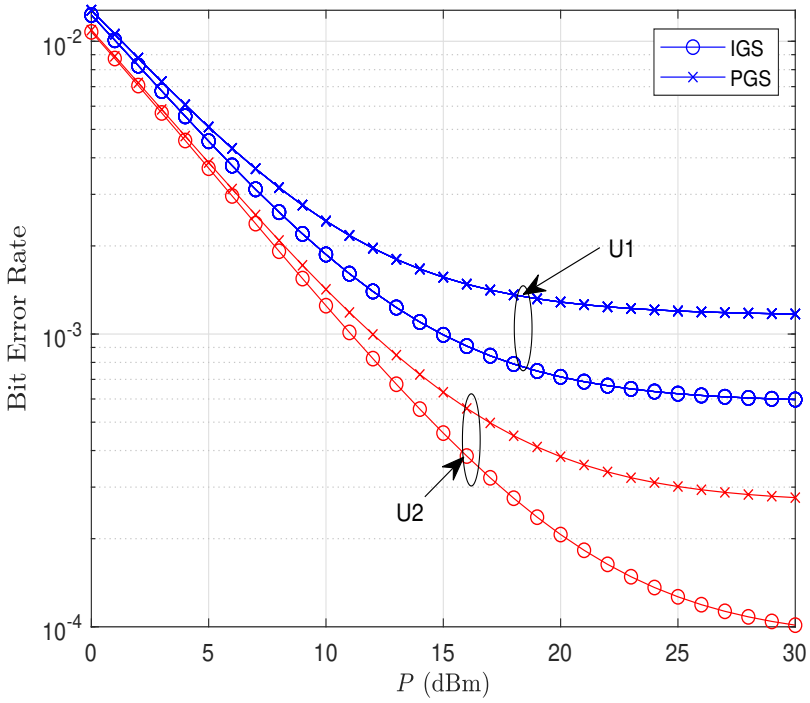

Figure 4.13: BER of U1 and U2 for both IGS and PGS cases.

Figure 4.12 presents the convergence of Algorithm III, where the energy efficiency against the number of iterations is plotted at $P_c = 20$ dBm and $P = 40$ dBm for both IGS and PGS cases. As observed in the figure, Algorithm III converges around 12 iterations in IGS case and after 8 iterations in PGS case.

Figure. 4.13 shows the BER performance through simulations versus P for both users in case of IGS and PGS at 16-QAM. For each channel realization, we find the optimal improper coefficients, then we find the corresponding improper constellation diagram, simulate the error rate, and repeat. For a given optimal IGS circularity coefficients, we study the BER performance of the new improper constellations, by means of 10^4 simulations, where each simulation considering the decoding of 10^4 symbols. The optimal maximum likelihood detector is applied at the receiver side for this problem, which aims to finding the closest constellation point to a given noisy received signal. As can be seen in the figure, error performance in case of using improper constellation diagrams outperforms

proper constellation diagram. The reason behind this is that since the improper constellation is designed based on WLT which relies on maximizing the minimum Euclidean distance, and hence, achieving better BER. Another observation is that as P increases, BER performance improves as expected. However, error floor occurs at high P due to residual interference resulted from the imperfect self-interference cancellation.

4.6 Conclusion

In this chapter, we investigated the potential merits (in terms of spectral and energy efficiencies and fairness among users) of using IGS in a cooperative FD-NOMA system. We derived exact expressions for the user throughput of each link in the cooperative system employing IGS. We formulated two joint optimization problems to maximize the overall sum rate and energy efficiency provided that the QoS of each user is satisfied. We also formulated a max-min throughput optimization problem to achieve fairness among users. Iterative algorithms, based on KKT conditions, were proposed to solve the formulated optimization problems. Simulation results showed the superiority of IGS over PGS in the context of cooperative FD-NOMA systems, where around 1.5 to 2 dB gain is achieved. IGS is also more energy-efficient than the PGS since it transmits around 0.5 bits/joule more at lower to medium BS's power values. In addition, the developed IGS max-min algorithm achieves a better trade-off when compared to its PGS counterpart.

Chapter 5

Rate-Splitting Multiple Access Using IGS under Imperfect SIC

5.1 Introduction

RSMA has been introduced as a novel technology for optimizing multiple access and interference management strategies for future wireless systems [9]. RSMA strategy relies on splitting of user messages as well as non-orthogonal transmission of common messages decoded by all users and private messages decoded by their corresponding users [19].

In RSMA, the signal transmitted to the users is divided into a common message and a private message. The common part is a message decoded by each user and the private part is a message that only intended to be received by specific user. Therefore, adapting the split of common and private signals can compromise the signal processing complexity and achieve higher data rates [19]. However, to attain such potential merits, perfect SIC and suitable splitting of common and private messages should be adopted. In practical scenarios, the assumption of perfect SIC at the receiver might not be possible. Hence, it is of great interest to compensate the impact of imperfect SIC for the RSMA systems.

IGS, as a generalized physical layer signaling scheme, has the capability to improve

the spectral efficiency of the interference-limited systems. By definition, IGS assumes correlated real and imaginary signal parts without necessarily equal powers as opposed to the traditional PGS scheme. Due to these extra degrees of signaling freedom, IGS is capable to compensate the impact of the residual interference due to the SIC imperfection [154]. This motivates us to consider IGS in RSMA under residual interference conditions.

5.1.1 Related Works

A number of existing works investigated interesting problems related to RSMA from many aspects including single and multiple antenna networks, interference and broadcasting channels etc. [19] and references therein. Appealing resource allocation methods using RSMA were proposed in the literature. In [125], the authors made use of RSMA to maximize the rate of all users in downlink multi-user MISO systems under imperfect CSI at the BS. In [127], the spectral and energy efficiencies were optimized using RSMA in non-orthogonal unicast and multicast transmissions. Recently, authors in [183] have considered the sum-rate maximization problem using RSMA subject to minimum rate demand. Rate allocation and power control iterative algorithms were proposed for SISO and MISO systems. However, all aforementioned works considered PGS and perfect SIC. Inspired by rate-splitting concept, work in [177] showed that the signal superposition which uses RSMA mechanism outperforms NOMA in a general multi-cell multiuser network. In this work, IGS was used for a two paired-users in the same cell to maximize the overall minimum throughput assuming perfect SIC.

5.1.2 Contributions

- In this work, we consider a downlink wireless network that serves multiple users using RSMA mechanism in the presence of imperfect SIC. To circumvent the impact of SIC imperfection, IGS is adopted at users' common messages which are decoded by all users.

- New expressions for the users' common rate and private rate are derived considering IGS and imperfect SIC.

- Using the derived expressions, an optimization problem is formulated to jointly optimize the circularity coefficient and power allocation to maximize the overall sum-rate subject to minimum rate requirements constraints.

- Simulation results show a significant sum-rate performance gain when using IGS-based RSMA systems compared with their counterpart PGS-based RSMA systems.

5.2 System Model and Rate Analysis

We consider a downlink wireless system that composes a BS that serves K users using RSMA mechanism. In RSMA, the transmitter splits the message W_k of the kth user into a common message $W_{c,k}$ and a private message $W_{p,k}$, $\forall k = 1, ..., K$. The common fractions $W_{c,1}, ..., W_{c,K}$ is combined into a common message W_c, which is encoded into the common stream s_c. This common stream is required to be decoded by all users. The private fraction $W_{p,k}$, containing the remaining portion of the message W_k, is encoded into the private stream s_k for user k.

PGS is adopted for the private stream s_k; however, due to the need of detecting the common messages at each user, and assuming imperfect SIC to cancel-out the common message when detecting the private messages, IGS is adopted for the common stream s_c to mitigate the impact of the residual interference at each user's receiver. The degree of impropriety of s_c is measured as discussed below.

We assume that the variance and pseudo-variance of s_c are denoted as $\omega_{s_c} = \mathcal{E}[s_c s_c^*]$ and $\hat{\omega}_{s_c} = \mathcal{E}[s_c s_c]$, respectively, where the superscript $(\cdot)^*$ refers to the complex conjugation operation and $\mathcal{E}[\cdot]$ refers to the expectation of a random variable. A signal is called proper if it has a zero pseudo-variance, i.e., $\hat{\omega}_{s_c} = 0$, otherwise it is called an improper signal. The impropriety degree (or the IGS circularity coefficient) of s_c is given as $\xi_c = |\hat{\omega}_{s_c}|/\omega_{s_c}$,

where $0 \leq \xi_c \leq 1$. If $\xi_c = 0$, then s_c is a proper signal, and if $\xi_c = 1$, we have a maximally improper signal.

The transmitted signal x of the BS is written as

$$x = \sqrt{p_c} s_c + \sum_{k=1}^{K} \sqrt{p_k} s_k, \tag{5.1}$$

where p_c is the transmit power of the common stream s_c, p_k is the transmit power of the private stream s_k transmitted to user $k, \forall k = 1, .., K$. We assume that the BS's budget power, P, is divided between the common stream and private stream as $p_c = \alpha_o P$, $p_k = \alpha_k P$, where $0 \leq \alpha_o, \alpha_k \leq 1$. Then, the transmitted signal can be re-written as

$$x = \sqrt{\alpha_o P} s_c + \sum_{k=1}^{K} \sqrt{\alpha_k P} s_k, \tag{5.2}$$

where α_o is the power allocation factor of the power allocated to the common stream, and α_k is the power allocation factor of the power allocated to the private stream of the user k. The total received signal at user k can be expressed as

$$y_k = h_k \sqrt{\alpha_o P} s_c + h_k \sum_{j=1}^{K} \sqrt{\alpha_j P} s_j + n_k, \tag{5.3}$$

where h_k refers to the channel coefficient between user k and the BS and n_k refers to the zero-mean additive white Gaussian noise with variance σ^2.

Without loss of generality, it is assumed that $|h_1|^2 \leq |h_2|^2 \leq ...|h_K|^2$, i.e. user with index 1 is with the weakest channel gain and the user with index K is with the strongest channel gain.

Based on the RSMA principle, each user detects the common message, and then, the user adopts SIC to cancel this common message and decode its own private message. To ensure that all users can successfully decode common stream s_c, the rate of common stream should be the common rate of the weakest user channel i.e. h_1 in our case [155].

To make sure that the common stream is detectable at each user receiver, the received signal to interference plus noise ratio (SINR) of the common stream at each user should be greater than its received SINR of the private stream. Next, to ensure that the common message can be detected through SIC operation at the receiver side, the transmit power of each user must satisfy the following constraint [155]

$$[(|h_1|^2P+\sigma^2)^2-(|h_1|^2\alpha_oP\xi_c)^2]^{1/2}-2(|h_1|^2P\sum_{j=1}^{K}\alpha_j+\sigma^2)\geq\epsilon. \tag{5.4}$$

where ϵ is the minimum difference between the common message power at kth user and the power of private messages of all users plus noise. The term $\alpha_oP|h_k|^2$ represents the received common message power at the kth user, while the term $\left(\sum_{j=1}^{K}\alpha_jP|h_k|^2+\sigma^2\right)$ represents the power of the non-decoded private message of all users plus noise. The minimum difference ϵ is needed to differentiate the common message to be decoded and the private messages of all users plus noise.

We assume that SIC process at users' receivers is imperfect and the residual interference component due to this imperfection is quantified by a factor β $(0\leq\beta\leq1)$, where $\beta=0$ refers to perfect SIC and $\beta=1$ refers to the fully imperfect SIC. Due to the SIC imperfection, the received signal at user k after decoding/canceling the common stream can be written as

$$\hat{y}_k=\beta h_k\sqrt{\alpha_oP}s_c+h_k\sum_{j=1}^{K}\sqrt{\alpha_jP}s_j+n_k, \tag{5.5}$$

where the first term in the right hand side refers to the residual interference.

In the following, we derive the common rate and the private rate expressions assuming that IGS is used for the common message while PGS is used for the private messages. , i.e., s_c is improper, and s_k is proper. Following [130], the achievable rate expression for a

SISO IGS-based system is given as

$$\mathcal{R}_i = \frac{1}{2}\log_2\left(\frac{\omega_{y_i}^2 - |\hat{\omega}_{y_i}|^2}{\omega_{z_i}^2 - |\hat{\omega}_{z_i}|^2}\right), \tag{5.6}$$

where ω_{y_i} and $\hat{\omega}_{y_i}$ refer respectively to the covariance and pseudo-covariance components of the received signal; while ω_{z_i} and $\hat{\omega}_{z_i}$ refer to covariance and pseudo covariance components of interference plus noise signal, respectively.

To derive the common rate, from the received signal in (5.3), the square of covariance and the square of the absolute value of pseudo covariance components are found as follows.

$$\omega_{y_k}^2 = (\mathcal{E}\{y_k y_k^*\})^2 = \left(|h_k|^2\alpha_o P + |h_k|^2\sum_{j=1}^{K}\alpha_j P + \sigma^2\right)^2, \tag{5.7}$$

$$|\hat{\omega}_{y_k}|^2 = |\mathcal{E}\{y_k y_k\}|^2 = (|h_k|^2\alpha_o P\xi_c)^2,, \tag{5.8}$$

where $\omega_{s_c} = \mathcal{E}[s_c s_c^*] = 1$. For the interference plus noise signal, $z_k = \sum_{j=1}^{K} h_k\sqrt{\alpha_j P}s_j + n_k$, the square of covariance and pseudo covariance components are found as follows

$$\omega_{z_k}^2 = (\mathcal{E}\{z_k z_k^*\})^2 = \left(|h_k|^2\sum_{j=1}^{K}\alpha_j P + \sigma^2\right)^2, \tag{5.9}$$

$$|\hat{\omega}_{z_k}|^2 = |\mathcal{E}\{z_k z_k\}|^2 = 0, \tag{5.10}$$

We first derive the common rate expression. By substituting (5.7)-(5.10) into (5.6), the achievable rate of user k decoding the common stream s_c can be expressed

$$R_c = \frac{1}{2}\log_2\left(\frac{(|h_k|^2 P + \sigma^2)^2 - (|h_k|^2\alpha_o P\xi_c)^2}{(|h_k|^2\sum_{j=1}^{K}\alpha_j P + \sigma^2)^2}\right), \tag{5.11}$$

To ensure that all users can successfully decode common stream s_c, the rate of common stream should be the common rate of the weakest user channel i.e. h_1 in our case [155].

Hence, based on (5.3) and (5.6), and after some mathematical manipulations, the common rate expression is derived as

$$R_c = \frac{1}{2}\log_2\left(\frac{(|h_1|^2P+\sigma^2)^2-(|h_1|^2\alpha_oP\xi_c)^2}{(|h_1|^2\sum_{j=1}^{K}\alpha_jP+\sigma^2)^2}\right), \tag{5.12}$$

After having decoded the common stream s_c, each user can decode its private stream. To derive the private rate, from (5.5), we have the desired signal $h_k\sqrt{\alpha_kP}s_k$ and the interference plus noise signal $\hat{z}_k = \beta h_k\sqrt{\alpha_oP}s_c + h_k\sum_{j=1,j\neq k}^{K}\sqrt{\alpha_jP}s_j + n_k$. Similarly, we find the square of the covariance and the square of the absolute value of the pseudo covariance components for the private messages and the interference plus noise signal $\hat{z}_k$ as follows

$$\omega_{\hat{y}_k}^2 = (\mathcal{E}\{\hat{y}_k\hat{y}_k^*\})^2 = \left(\beta^2|h_k|^2\alpha_oP + |h_k|^2\sum_{j=1}^{K}\alpha_jP+\sigma^2\right)^2, \tag{5.13}$$

$$|\hat{\omega}_{y_k}|^2 = |\mathcal{E}\{y_ky_k\}|^2 = (\beta^2|h_k|^2\alpha_oP\xi_c)^2, \tag{5.14}$$

$$\omega_{\hat{z}_k}^2 = (\mathcal{E}\{\hat{z}_k\hat{z}_k^*\})^2 = \left(\beta^2|h_k|^2\alpha_oP + |h_k|^2\sum_{j=1,j\neq k}^{K}\alpha_jP+\sigma^2\right)^2, \tag{5.15}$$

$$|\hat{\omega}_{\hat{z}_k}|^2 = |\mathcal{E}\{\hat{z}_k\hat{z}_k\}|^2 = (\beta^2|h_k|^2\alpha_oP\xi_c)^2, \tag{5.16}$$

By substituting (5.13)-(5.16) into (5.6) and after a few simplifications, the private rate for a user k is derived as

$$R_{p_k} = \frac{1}{2}\log_2\left(\frac{(\beta^2|h_k|^2\alpha_oP+|h_k|^2P\sum_{j=1}^{K}\alpha_j+\sigma^2)^2-(\beta^2|h_k|^2\alpha_oP\xi_c)^2}{(\beta^2|h_k|^2\alpha_oP+|h_k|^2P\sum_{j=1,j\neq k}^{K}\alpha_j+\sigma^2)^2-(\beta^2|h_k|^2\alpha_oP\xi_c)^2}\right). \tag{5.17}$$

Substituting $\xi_c = 0$ and $\beta = 0$ reduces to the rates of PGS case in perfect SIC.

Given the common rate R_c and the rate allocated at user k, a_k, we have $\sum_{k=1}^{K}a_k \leq R_c$, which indicates that the total data rates of all users receiving common stream should be less than or equal the common rate R_c because all users must decode the same shared message, which is broadcast only once and constrained by the weakest link. So we can

have the total rate of user k in RSMA scenario is given as

$$\begin{aligned} R_k &= a_k + R_{p_k} \\ &= a_k + \frac{1}{2}\log_2\left(\frac{(\beta^2|h_k|^2\alpha_o P + |h_k|^2 P\sum_{j=1}^{K}\alpha_j + \sigma^2)^2 - (\beta^2|h_k|^2\alpha_o P\xi_c)^2}{(\beta^2|h_k|^2\alpha_o P + |h_k|^2 P\sum_{j=1,j\neq k}^{K}\alpha_j + \sigma^2)^2 - (\beta^2|h_k|^2\alpha_o P\xi_c)^2}\right) \end{aligned} \tag{5.18}$$

5.3 Optimization Problem

In this section, an optimization problem is formulated to optimize the IGS circularity coefficient and power allocation in order to maximize the sum of the private rates of a K users RSMA system subject to minimum rate requirement of the common rate.

Based on RSMA concept, the common message has the higher power than the power of the private messages to ensure that common message can be detected at each user receiver, i.e., the private messages power is treated as a noise when detecting the common message. However, due to SIC imperfection, residual interference occurs when detecting the private messages. The goal is to benefit from IGS to mitigate the impact of the SIC imperfection in this interference-limited scenario to improve the sum of private rates. Hence, we employ the IGS at the common message and PGS at the private message.

That said, the joint optimization problem for maximizing the private sum rate under the minimum common rate demand and power budget can be formulated as

$$\underset{\xi_c,\alpha_c}{\text{maximize}} \quad \sum_{k=1}^{K} R_{p_k} \tag{5.19a}$$

$$\text{subject to} \quad \mathcal{C}1: \; R_c \geq R_{\text{th}}, \tag{5.19b}$$

$$\mathcal{C}2: (5.4) \tag{5.19c}$$

$$\mathcal{C}3: \; R_{p_k} \geq R_k^{\min}, \forall k, \tag{5.19d}$$

$$\mathcal{C}4: \; 0 \leq \xi_c \leq 1, \tag{5.19e}$$

$$\mathcal{C}5: \; p_c + \sum_{k=1}^{K} p_k = P, \tag{5.19f}$$

where R_c and R_{p_k} are obtained from (5.20) and (5.21), respectively. The constraint $\mathcal{C}1$ ensures the minimum rate requirement of the common rate where R_{th} is the minimum rate threshold. The constraint $\mathcal{C}2$ is to ensure that the common message is detectable at each user. The minimum private rate constraints for all users are ensured in $\mathcal{C}3$, where $R_k^{\min}$ is the minimum private rate requirement for each user k. The constraint $\mathcal{C}4$ reflects that the circularity coefficient is between 0 and 1. The constraint $\mathcal{C}5$ refers to the maximum power budget requirement.

Based on $\mathcal{C}5$ in (5.19f), $\sum_{j=1}^{K} \alpha_j = 1$, $\alpha_o + \sum_{k=1}^{K} \alpha_k = 1$. Hence, R_c and R_{p_k} are expressed as

$$R_c = \underbrace{\log_2\left(1 + \frac{\alpha_o P|h_1|^2}{(1-\alpha_o)P|h_1|^2 + \sigma^2}\right)}_{\text{proper}} + \underbrace{\frac{1}{2}\log_2\left(1 - \frac{(\alpha_o P|h_1|^2\xi_c)^2}{(P|h_1|^2 + \sigma^2)^2}\right)}_{\text{Improper}}, \tag{5.20}$$

and

$$R_{p_k} = \underbrace{\log_2\left(1+\frac{\alpha_k P|h_k|^2}{\beta^2\alpha_o P|h_k|^2+(1-\alpha_o-\alpha_k)P|h_k|^2+\sigma^2}\right)}_{\text{proper}} + \underbrace{\frac{1}{2}\log_2\left(1-\frac{(\beta^2\alpha_o P|h_k|^2\xi_c)^2}{(\beta^2\alpha_o P|h_k|^2+(1-\alpha_o)P|h_k|^2+\sigma^2)^2}\right)}_{\text{Improper}} - \underbrace{\frac{1}{2}\log_2\left(1-\frac{(\beta^2\alpha_o P|h_k|^2\xi_c)^2}{(\beta^2\alpha_o P|h_k|^2+(1-\alpha_o-\alpha_k)P|h_k|^2+\sigma^2)^2}\right)}_{\text{Improper}}. \tag{5.21}$$

5.3.1 Bounds of $\alpha_o, \xi_c, \alpha_k$

To find a sub-optimal solution of the optimization problem in (5.19), we need first to satisfy the constraint $\mathcal{C}1$. Based on the channel assumption $|h_1|^2 \leq |h_2|^2 \leq ...|h_K|^2$, the constraint $\mathcal{C}1$ in (5.4) can be simplified as

$$\alpha_o P|h_1|^2 - \left(\sum_{j=1}^{K}\alpha_j P|h_1|^2+\sigma^2\right) \geq \epsilon, \tag{5.22}$$

From (5.22) and given that $\sum_{j=1}^{K}\alpha_j = 1$, the power allocation α_o required to guarantee that the common message is detectable at each user is derived as

$$\alpha_o \geq \frac{\epsilon+\sigma^2}{2P|h_1|^2} + \sum_{j=1}^{K}\alpha_j. \tag{5.23}$$

Next, to satisfy the constraint $\mathcal{C}1$ in (5.19b), we solve for $R_c \geq R_{\text{th}}$. After mathematical simplifications, ξ_c is computed as

$$\xi_c \leq \sqrt{\frac{\phi-\theta}{(|h_1|^2\alpha_o P)^2}}, \tag{5.24}$$

where $\phi = (|h_1|^2 P+\sigma^2)^2$ and $\theta = 2^{2R_{\text{th}}}(|h_1|^2(1-\alpha_o)P+\sigma^2)^2$. To satisfy the minimum

private rate constraint $\mathcal{C}3$ in (5.19d), α_k is derived as follows

$$\alpha_k \geq \frac{\delta - \sqrt{\delta^2 + \Delta}}{1 - \alpha_o}, \tag{5.25}$$

where $\delta = \left(\beta^2\alpha_o + (1-\alpha_o) + \frac{\sigma^2}{P|h_k|^2}\right)$, $\Delta = (2^{-2R_k^{\min}} - 1)[\delta^2 - (\beta^2\alpha_o\xi_c)^2]$. This α_k is a non-negative value due to the constraint $\alpha_k \geq 0$. The proof is provided in Appendix B.2.

5.3.2 Proposed Solution

After finding the bounds of τ, ξ_c, α_k, the optimization problem in (5.19) can be re-written as

$$\underset{\xi_c,\alpha_k,\alpha_o}{\text{maximize}} \quad \sum_{k=1}^{K} R_{p_k}(\xi_c, \alpha_k, \alpha_o) \tag{5.26a}$$

$$\text{subject to} \quad \mathcal{C}1: \ \xi_c \leq \sqrt{\frac{\phi - \theta}{(|h_1|^2\alpha_o P)^2}}, \tag{5.26b}$$

$$\mathcal{C}2: \alpha_o \geq \frac{\epsilon + \sigma^2}{2P|h_1|^2} + \sum_{j=1}^{K} \alpha_j, \tag{5.26c}$$

$$\mathcal{C}3: \ \alpha_k \geq \frac{\delta - \sqrt{\delta^2 + \Delta}}{1 - \alpha_o} \, \forall k, \tag{5.26d}$$

$$\mathcal{C}4: \ 0 \leq \xi_c \leq 1. \tag{5.26e}$$

To solve this optimization problem, we optimize each variable separately as a sub-optimal solution. The objective of the optimization problem in (5.26) is to maximize the sum of private rates, however, $\sum_{k=1}^{K} R_{p_k}$ in (5.26a) is a decreasing function of α_o (the proof is in the Appendix B.3). Then, α_o with equality holds in (5.23) is the lower bound of the power allocation that is assigned to the common stream power to guarantee its detection, i.e., $\alpha_{oLP} = \frac{\epsilon+\sigma^2}{2P|h_1|^2} + \sum_{j=1}^{K} \alpha_j$, and the remaining amount of power $(1-\alpha_{oLP})$ is assigned to the private stream so that we achieve the maximum sum of the private rates, i.e. $\alpha_o^* = \alpha_{oLB}$.

Since $\sum_{k=1}^{K} R_{p_k}$ is an increasing function of ξ_c (the proof is provided in B.4), we can set

$\xi_c = \sqrt{\frac{\phi-\theta}{(|h_1|^2\alpha_o P)^2}}$ to satisfy the constraint $\mathcal{C}1$ in (5.26b) in order to achieve the maximum sum of private rates. By considering also the constraint $\mathcal{C}4$ in (5.26e), the sub-optimal solution of ξ_c is derived as follows

$$\xi_c^* = \begin{cases} \text{case } 1: 0, & \text{if } \phi \leq \theta, \\ \text{case } 2: \sqrt{\frac{\phi-\theta}{(|h_1|^2\alpha_o P)^2}}, & \text{if } \theta < \phi \leq \psi, \\ \text{case } 3: 1, & \text{if } \phi > \psi, \end{cases} \tag{5.27}$$

where $\psi = 2^{2R_{\text{th}}}(|h_1|^2(1-\alpha_o)P + \sigma^2)^2 + (|h_1|^2\alpha_o P)^2$. Next, we need to find the sub-optimal α_o^* in case 1 and 3. As for cases 1 and 3, we need to find the corresponding α_o to satisfy constraint $\mathcal{C}1$ in (5.26b). We find the following bounds α_{o1} and α_{o2} for both cases, respectively, as

$$\alpha_{o1} \geq \left(1 + \frac{\sigma^2}{|h_1|^2 P}\right)\left(1 - 2^{-R_{\text{th}}}\right), \tag{5.28}$$

and

$$\alpha_{o2} \geq \left(1 + \frac{\sigma^2}{|h_1|^2 P}\right)\left(\frac{2^{2R_{\text{th}}} - 1}{2^{2R_{\text{th}}} + 1}\right). \tag{5.29}$$

Based on the above analysis, by considering all constraints $\mathcal{C}1 - \mathcal{C}5$, the sub-optimal α_o^* is derived as follows

$$\alpha_o^* = \begin{cases} \text{case } 1: \max(\alpha_{o1}, \alpha_{oLP}), & \text{if } \phi \leq \theta, \\ \text{case } 2: \alpha_{oLP}, & \text{if } \theta < \phi \leq \psi, \\ \text{case } 3: \max(\alpha_{o2}, \alpha_{oLP}), & \text{if } \phi > \psi, \end{cases} \tag{5.30}$$

Next, after we get the sub-optimal solution for the IGS circularity coefficient and power allocation factor between the common stream and the private stream, i.e., (ξ_c^*, α_o^*), we need to find the sub-optimal solution of the power allocation factor α_k^* between the

private messages. Hence, our optimization problem can be re-written as

$$\underset{\alpha_k}{\text{maximize}} \quad \sum_{k=1}^{K} R_{p_k}(\xi_c^*, \alpha_k, \alpha_o^*) \tag{5.31a}$$

$$\text{subject to} \quad \mathcal{C}6: \sum_{k=1}^{K} \alpha_k = 1, \tag{5.31b}$$

$$\mathcal{C}7: \ \alpha_k \geq \alpha_k^{\min}, \forall k = 1, .., K. \tag{5.31c}$$

In (5.31c), $\alpha_k^{\min}$ is found in (5.25) with the equality sign holds is used to satisfy the minimum private rate constraint in (5.19d), and the problem in (5.31) is feasible, if and only if, $\sum_{k=1}^{K} \alpha_k^{\min} \leq 1$, which can be re-presented as follows

$$\sum_{k=1}^{K} \left(\delta - \sqrt{\delta^2 + \Delta}\right) \leq (1 - \alpha_o^*). \tag{5.32}$$

To be able to find the solution for problem (5.31), we need the following Lemma [183].

Lemma 1: For the sub-optimal α_k^* of problem (5.31), there exists one user k such that $\alpha_k^* = 1 - \sum_{j=1, j\neq k}^{K} \alpha_j^*$ and $\alpha_j^* = \alpha_j^{\min}$, $\forall j \in K, j \neq k$.

Proof. See Appendix B.5. □

Though the objective function in (5.31) is convex function of α_k (proof is in Appendix B.6), the maximization optimization problem is non-concave problem (maximization of convex function results in non-concave problem). Exploiting Lemma 1, we can find the sub-optimal solution for the problem in (5.31) in closed form as presented in the following Lemma.

Lemma 2:The sub-optimal power allocation of problem (5.31) is given as

$$\alpha_k^* = 1 - \sum_{j=1, j\neq k}^{K} \alpha_j^{\min}, \tag{5.33}$$

where

$$\alpha_j^{\min} = \sum_{j=1, j\neq k}^{K} \frac{\delta - \sqrt{\delta^2 + \Delta}}{(1-\alpha_o^*)}. \tag{5.34}$$

and

$$k = \arg\max_{m\in K} \left[\frac{(\beta^2\alpha_o P + P(1-\alpha_o)(1-\alpha_m^{\min}) + \frac{\sigma^2}{|h_m|^2})^2 - (\beta^2\alpha_o P\xi_c)^2}{(\beta^2\alpha_o P + P(1-\alpha_o)\sum_{j=1,j\neq m}^{K}\alpha_j^{\min} + \frac{\sigma^2}{|h_m|^2})^2 - (\beta^2\alpha_o P\xi_c)^2} \right]. \tag{5.35}$$

Proof. Based on the Lemma 1, we can substitute the sub-optimal power allocation $\alpha_k^* = 1 - \sum_{j=1,j\neq k}^{K}\alpha_j^{\min}$ into the the objective function of problem (5.31) and we get

$$\begin{aligned} &\sum_{j=1,j\neq k}^{K} \frac{1}{2}\log_2\left(\frac{(\beta^2|h_j|^2\alpha_o P + |h_j|^2 P(1-\alpha_o) + \sigma^2)^2 - (\beta^2|h_j|^2\alpha_o P\xi_c)^2}{(\beta^2|h_j|^2\alpha_o P + |h_s|^2 P(1-\alpha_o)(1-\alpha_j^{\min}) + \sigma^2)^2 - (\beta^2|h_j|^2\alpha_o P\xi_c)^2}\right) \\ &+ \frac{1}{2}\log_2\left(\frac{(\beta^2|h_k|^2\alpha_o P + |h_k|^2 P(1-\alpha_o) + \sigma^2)^2 - (\beta^2|h_k|^2\alpha_o P\xi_c)^2}{(\beta^2|h_t|^2\alpha_o P + |h_k|^2 P(1-\alpha_o)\sum_{j=1,j\neq k}^{K}\alpha_j^{\min} + \sigma^2)^2 - (\beta^2|h_k|^2\alpha_o P\xi_c)^2}\right). \end{aligned} \tag{5.36}$$

To maximize this sum private rate in (5.36), the optimal k should be selected as shown below

$$\begin{aligned} k = \arg\max_{m\in K} &\sum_{j=1,j\neq k}^{K} \frac{1}{2}\log_2\left(\frac{(\beta^2|h_j|^2\alpha_o P + |h_j|^2 P(1-\alpha_o) + \sigma^2)^2 - (\beta^2|h_j|^2\alpha_o P\xi_c)^2}{(\beta^2|h_j|^2\alpha_o P + |h_j|^2 P(1-\alpha_o)(1-\alpha_j^{\min}) + \sigma^2)^2 - (\beta^2|h_j|^2\alpha_o P\xi_c)^2}\right) \\ &- \frac{1}{2}\log_2\left(\frac{(\beta^2|h_m|^2\alpha_o P + |h_m|^2 P(1-\alpha_o) + \sigma^2)^2 - (\beta^2|h_m|^2\alpha_o P\xi_c)^2}{(\beta^2|h_m|^2\alpha_o P + |h_m|^2 P(1-\alpha_o)(1-\alpha_m^{\min}) + \sigma^2)^2 - (\beta^2|h_m|^2\alpha_o P\xi_c)^2}\right) \\ &+ \frac{1}{2}\log_2\left(\frac{(\beta^2|h_m|^2\alpha_o P + |h_m|^2 P(1-\alpha_o) + \sigma^2)^2 - (\beta^2|h_m|^2\alpha_o P\xi_c)^2}{(\beta^2|h_m|^2\alpha_o P + |h_m|^2 P(1-\alpha_o)\sum_{j=1,j\neq m}^{K}\alpha_j^{\min} + \sigma^2)^2 - (\beta^2|h_m|^2\alpha_o P\xi_c)^2}\right). \end{aligned} \tag{5.37}$$

RSMA Proposed Algorithm:

1: **INPUT:** $R_{\text{th}}, R_k^{\min}, \epsilon, \beta, P$
2: Set $\alpha_o = \alpha_{oLP}$ from (5.23).
3: Check which case is valid and find the sub-optimal circularity coefficient ξ_c^* from (5.27).
4: For the corresponding case, find the sub-optimal α_o^* from (5.30).
5: Calculate $\alpha_j^{\min}, \forall j = 1, ..., K$ from (5.34).
6: Find the user k from (5.35).
7: For this kth user, find sub-optimal α_k^* from (5.33).

After further simplifications, (5.37) is presented as expressed as

$$\begin{aligned} k = \arg\max_{m \in K} &= \frac{1}{2}\log_2(\beta^2|h_m|^2\alpha_o P + |h_m|^2 P(1-\alpha_o)(1-\alpha_m^{\min}) + \sigma^2)^2 - (\beta^2|h_m|^2\alpha_o P\xi_c)^2 \\ &- \frac{1}{2}\log_2(\beta^2|h_m|^2\alpha_o P + |h_m|^2 P(1-\alpha_o)\sum_{j=1, j\neq m}^{K}\alpha_j^{\min} + \sigma^2)^2 - (\beta^2|h_m|^2\alpha_o P\xi_c)^2 \\ &= \arg\max_{m \in K} = \frac{(\beta^2|h_m|^2\alpha_o P + |h_m|^2 P(1-\alpha_o)(1-\alpha_m^{\min}) + \sigma^2)^2 - (\beta^2|h_m|^2\alpha_o P\xi_c)^2}{(\beta^2|h_m|^2\alpha_o P + |h_m|^2 P(1-\alpha_o)\sum_{j=1, j\neq m}^{K}\alpha_j^{\min} + \sigma^2)^2 - (\beta^2|h_m|^2\alpha_o P\xi_c)^2}. \end{aligned} \tag{5.38}$$

After a few algebraic simplifications, we get (5.35). This concludes the proof. □

Lemma 2 emphasizes that it is optimal for the BS to allocate more power to the user that can maximize the sum private rate while allocating the minimum transmit power that can satisfy the minimum rate requirement for all other users. The proposed algorithm is summarized as shown at the top of the page.

5.4 Simulation Results

In this section, we simulate a downlink wireless system that consists of a BS serving K users using RSMA mechanism and employing IGS at the common message. We compare its achieved sum private rate to its counterpart of PGS-based RSMA systems.

We consider the distance-dependent path-loss model as a form of large-scale fading, and the Rayleigh fading model as small-scale multi-path fading. The channel from the BS to user k, $\forall k = 1, ..., K$, at a distance of d_k meters is generated as $\sqrt{10^{-\frac{\sigma_{PL}}{10}}}h_k$, where

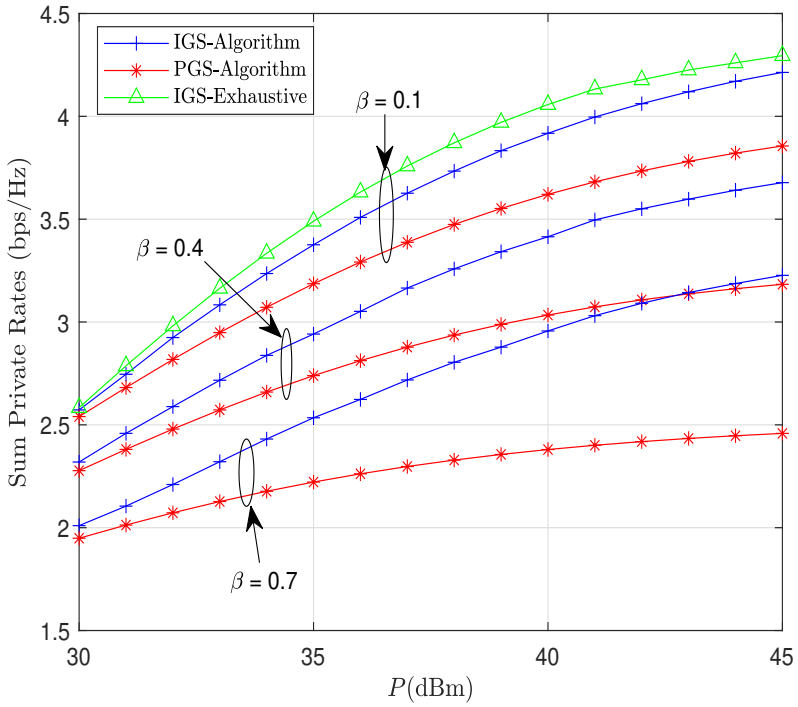

Figure 5.1: Sum private rate vs P (dBm) for IGS-based and PGS-based RSMA systems for different β and $K = 3$.

h_k is a Rayleigh fading channel coefficient and $\sigma_{PL} = 38.46 + 10n\log_{10}(d_k)$ is the path-loss in dB, where $n = 3$ is the path-loss exponent. We assume users are uniformly distributed in a circle area of diameter 300 meter. Unless otherwise mentioned, we assume that the maximum transmit power is $P = 42$ dBm, the noise power is σ^2 = -104 dBm/Hz, $R_k^{\min}$= 0.5 bps, and $R_{\text{th}} = 0.5$ bps/Hz. The SIC detection threshold is set as ϵ = -94 dBm (the value of ϵ is selected to ensure that the error rate of decoding the common message is below certain threshold) [183]. All simulation results are averaged over 10^4 Monte Carlo channel realizations.

In Fig. 5.1, the sum private rate is simulated versus P (dBm) where optimal ξ_c^*, α_o^*, and α_k^* are used at different values of $\beta = 0.1, 0.4, 0.7$, and $K = 3$. It is clearly seen that IGS-based RSMA outperforms PGS-based at all level of SIC imperfections. In particular, as the SIC becomes worse, the sum private rate gain of using IGS-based RSMA increases over PGS-based RSMA. This is due to the fact that IGS circumvents the severe impact

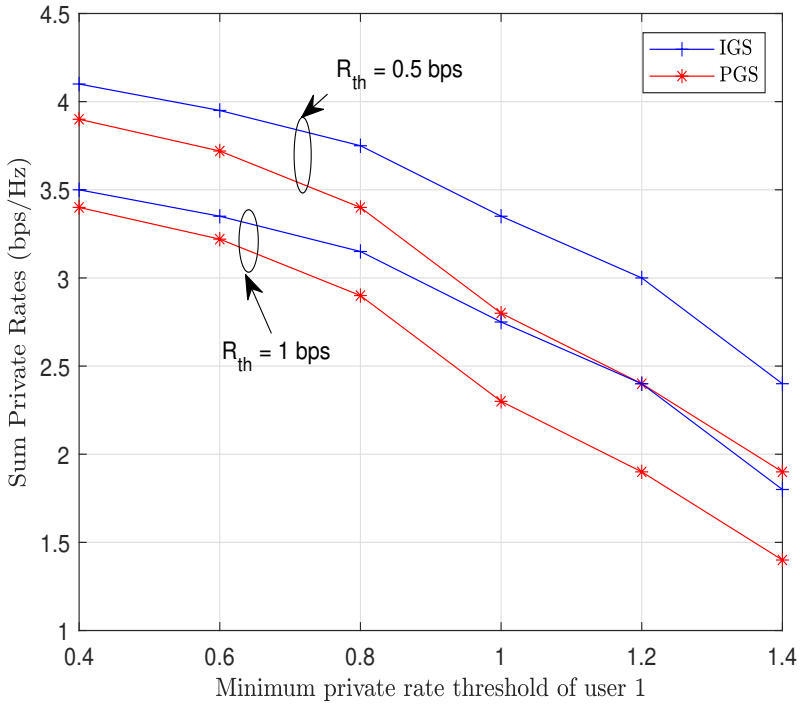

Figure 5.2: Comparison of the sum private rate performance vs. minimum private rate demand of user 1 under different minimum common rate demand, with $\beta = 0.1$ and $K = 3$.

of the residual interference occurring due to SIC imperfections. In addition, we compared the proposed IGS-based sub-optimal solution to the exhaustive search solution at $\beta = 0.1$. It can be observed that a small performance gap exists between the sub-optimal and the exhaustive search solutions.

Fig. 5.2 shows the sum private rates versus minimum private rate threshold for each user at different common rate thresholds $R_{\text{th}} = 0.5, 1$ bps/Hz. From this figure, RSMA IGS-based scheme always achieves a better performance than PGS-based scheme under imperfect SIC. It can be observed that the sum private rate degrades softly when the minimum private rate threshold is a bit low. However, when the minimum private rate threshold is high, the sum private rate degrades sharply. This is due to the reason that a high minimum private rate threshold requests the BS to allocate more power to the users with worse channel condition, which consequently degrades the private sum rate performance. It is also observed that as R_{th} increases, the private sum rate performance

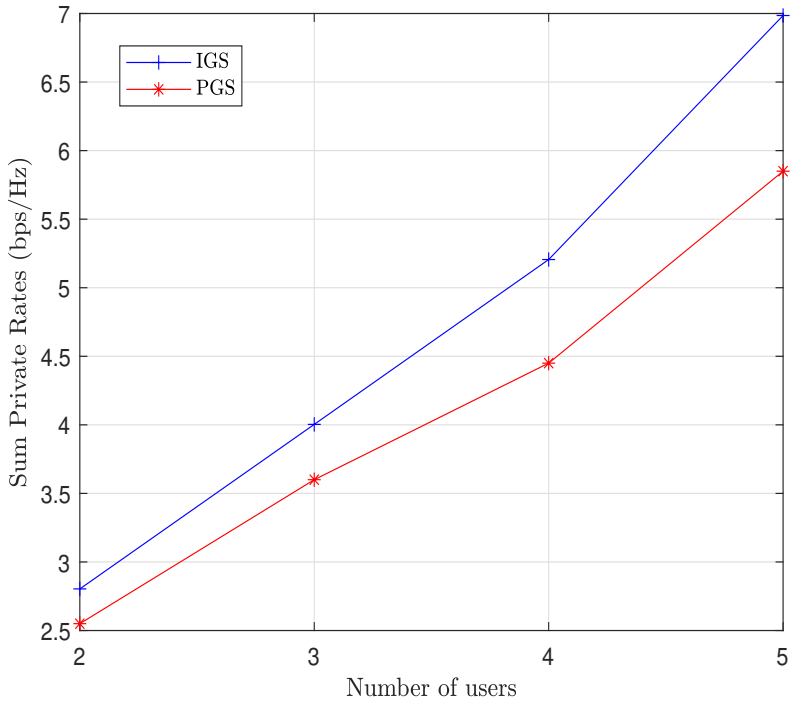

Figure 5.3: Sum private rate vs number of users for both IGS-based and PGS-based RSMA systems with $\beta = 0.1$.

in both schemes degrades because the BS requires to allocate more power for the common message to be decoded by all users.

In Fig. 5.3, the sum private rate is simulated vs the number of users at $\beta = 0.1$ and $P = 42$ dBm. As the number of users increases, the total sum private rate increases for both IGS and PGS RSMA schemes due to the multi-user gain that RSMA achieves. However, the sum private rate gain of IGS-bsed RSMA scheme outperforms PGS-based scheme due to the impact of extra degrees of freedom that IGS signaling offers.

5.5 Conclusion

In this chapter, we maximized the sum private rate of a multi-user RSMA system subject to minimum rate and power budget requirements under imperfect SIC. The optimal IGS circularity coefficient and power allocation solutions are obtained. Simulation results

showed that IGS-based RSMA systems outperform their counterpart PGS-based RSMA systems in terms of private sum rate. A gain of 0.5 bps/Hz is attained at medium SNR region, In addition, it was observed that the gain from the IGS increases when the imperfect SIC increases. Results also revealed that the optimal power allocation scheme enhances the gain of both IGS-based RSMA and PGS-based RSMA systems over the fixed power allocation scheme.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

NOMA and RSMA are considered potential solutions to suffice the the new requirements for B5G and 6G wireless communication systems. They have the capability to support low-loaded and heavily-loaded systems with much higher data rate. In this research work, we focused on employing IGS as a generalized signaling scheme to improve the spectral and energy efficiency of NOMA and RSMA technologies by exploiting the extra degrees of freedom that IGS can offer.

In this thesis, we first investigated the the performance of using IGS in direct point-to-point NOMA systems under imperfect SIC assumption. Achievable user rates expressions were derived when users adopt IGS. The spectral and energy efficiency maximization problems of a two-user NOMA system were formulated under minimum user-rate requirements and total power constraints. Two iterative algorithms were proposed to jointly optimize the IGS circularity coefficients and power allocation for the formulated problems using sub-optimal solutions of KKT and Dinkelbach approaches. Furthermore, improper constellation diagrams are designed using WLT with predefined optimized IGS coefficients to study the impact of IGS on throughput and error performance.

Results indicated that the spectral and energy efficiencies of IGS-based NOMA systems were significantly enhanced by jointly optimizing the circularity coefficients for both users compared to PGS-based NOMA systems and scenarios where only the circularity coefficient of a single user was optimized in IGS-based systems. A spectral efficiency of about 4 bits/s/Hz can be attained by the modified proposed algorithm at 3 dBW power at $\eta = 0.1$ and $\alpha = 0.4$. To achieve the same spectral efficiency of 4 bits/s/Hz using only one user IGS-based and PGS-based schemes, nearly 2 to 2.5 dBW power are needed, respectively. Moreover, it was observed that the benefits of using IGS became more pronounced as the level of SIC imperfection increased. Additionally, the findings revealed that IGS could offset the extra power required for processing at the strong user in a NOMA system under imperfect SIC. Furthermore, the results demonstrated that the error performance of the IGS-based system surpassed that of its PGS-based counterpart, owing to the unique impropriety characteristics of the constellation.

Second, we analyzed a cooperative FD-NOMA system employing IGS under the influence of imperfect self-interference cancellation. In this context, three optimization problems were formulated to maximize spectral efficiency, ensure max-min rate fairness, and optimize energy efficiency. The IGS circularity coefficients and power allocation at the BS were jointly optimized, subject to each user's rate constraints and the system's power budget. New expressions for user rates were derived under the assumption that both users adopt IGS. To solve the formulated problems, we proposed iterative algorithms based on solving the KKT conditions, yielding sub-optimal solutions. Furthermore, the effect of the IGS circularity coefficient on each user's constellation diagram was analyzed. We also designed improper constellations using the optimized IGS coefficients to evaluate the impact of IGS on error performance.

Simulation results demonstrated the superiority of IGS-based NOMA systems over their PGS-based counterparts in the context of cooperative FD-NOMA systems, achieving a performance gain of approximately 1.5 to 2 dB. Additionally, IGS was shown to be more energy-efficient than PGS, as it requires less power in the low to medium power

regions. IGS is considered more energy-efficient than the PGS since it transmits around 0.5 bits/joule more at lower to medium BS's power values. Furthermore, the developed IGS max-min algorithm provided a better trade-off compared to its PGS counterpart, highlighting its effectiveness in achieving improved system performance.

Third, RSMA combined with IGS was investigated in a multi-user environment under the influence of imperfect SIC. New expressions were derived for the achievable user rates, including the common and private rates, assuming IGS is employed for common message signaling. Using these expressions, a spectral efficiency maximization problem was formulated for the users' private rates, subject to a constraint ensuring that the minimum users' common rate remains above a specified threshold. Optimal solutions were obtained for the IGS circularity coefficients and power allocation at the BS. The results highlighted the effectiveness of IGS in enhancing the spectral efficiency of the RSMA system compared to its PGS-based counterpart.

6.2 Future Work

6.2.1 Direct Extensions

The use of IGS with NOMA and RSMA is still in its infancy in academic research, leaving numerous open challenges to be addressed. In the context of point-to-point NOMA systems, this thesis considered a two-user SISO system model. Future work could extend this research to MISO or MIMO NOMA systems with multiple users. A key challenge in such an extension would be designing practical and efficient solutions for these more complex systems, particularly when IGS is adopted for a group of users or all users in the system. This would involve addressing issues related to user grouping, interference alignment, resource allocation, and computational complexity, while ensuring the benefits of IGS are fully realized in multi-user and multi-antenna scenarios.

Extensions of IGS in cooperative relaying systems include exploring multi-hop relay-

ing to manage accumulated interference, cluster-based relaying with dynamic cluster formation, and hybrid decode-and-forward (DF) and amplify-and-forward (AF) schemes for improved spectral and energy efficiency. Relay selection algorithms optimized for IGS can enhance throughput, while alternating relaying under imperfect self-interference cancellation can further mitigate interference [134]. Other extensions include improving physical layer security in secure relaying systems, leveraging energy-harvesting relays to optimize power usage, and integrating IGS with cooperative systems for sensing and communication (ISAC) applications. These extensions focus on joint optimization of beamforming, power allocation, and circularity coefficients to maximize system performance.

In this research, we focused on an FD-NOMA relaying system with only two users to simplify the analysis and reduce the complexity of the proposed solutions. A more generalized FD-NOMA system model could involve multiple groups of users, where each group includes two paired NOMA users, and conventional OMA techniques are employed between these groups in a multi-user system. Furthermore, alternative cooperative schemes, such as alternating relaying NOMA systems under imperfect SIC, present promising opportunities for employing IGS. Incorporating IGS in these setups could potentially enhance both spectral and energy efficiency, offering valuable advancements in the performance of future communication systems.

Another extension could be the multi-relay FD-NOMA with IGS. We can extend the FD-NOMA relaying system to a multi-relay setup, where multiple relays simultaneously assist user transmissions. In this extension, we may investigate how IGS can be leveraged to manage the interference caused by multiple relays and optimize relay selection, circularity coefficients, and power allocation. The study can be extended to scenarios involving joint beamforming design and IGS optimization in FD-NOMA systems equipped with multiple antennas at the base station. In such systems, beamforming vectors, power allocation, and circularity coefficients can be jointly optimized to enhance both spectral and energy efficiency, leveraging the benefits of IGS in mitigating interference and improving overall system performance.

Since the RSMA is still an open research area, there is significant potential for further development in this area as an extension of our work, particularly by incorporating IGS into MIMO systems. Additionally, other performance metrics such as energy efficiency and max-min fairness could be explored through the use of IGS in RSMA. Furthermore, cooperative relaying networks could be examined using RSMA technology, considering practical challenges such as SIC imperfections, self-interference cancellation issues, and phase and quadrature imbalance.

One possible extension could involve investigating the incorporation of advanced modulation schemes, such as OFDM and QAM, within RSMA and IGS systems to enhance spectral efficiency and robustness against interference by enabling efficient signal transmission and better adaptability to channel conditions. Additionally, dynamic spectrum management strategies could be explored to optimize resource allocation in MIMO systems, particularly in time-varying and spatially heterogeneous wireless environments, further improving system performance and adaptability.

6.2.2 Open Research Directions

- MISO, MIMO, and Massive MIMO: While there has already been substantial work integrating NOMA and RSMA with MIMO systems, several intriguing research directions remain in this field [6]. One such direction involves utilizing IGS in multi-carrier settings, such as OFDM MIMO systems. The inclusion of sub-carrier assignments, alongside power allocation and user selection, introduces a significant level of complexity in terms of rate optimization problems in IGS. Additionally, uplink MIMO NOMA and RSMA with IGS are of great interest, particularly in massive machine-type connectivity scenarios within B5G systems, where the power budget of uplink transmitters (i.e., users) is a critical concern.

- mmW and Tera-Hertz (THz) Communications: A major challenge in high-frequency bands, such as mmW and Tera-Hertz (THz), is the significant path loss encountered

during signal propagation. The integration of promising techniques like IGS and multiple access schemes such as NOMA and RSMA offers a valuable complement to these systems, helping to efficiently serve a large set of users in such communication environments. There remains ample opportunity for innovative solutions that leverage the potential capabilities of IGS, mmW, and THz channels, benefiting NOMA and RSMA systems through both intelligent machine learning approaches and traditional optimization methods. UAVs Assisted Communications: Most existing rate optimization problems focus on a single unmanned aerial vehicle (UAV) acting as a flying BS to serve multiple users. However, incorporating UAV into NOMA and RSMA with IGS systems can offer a variety of solutions, each with its own unique sum-rate optimization challenges. For example, integrating NOMA, RSMA, and IGS in a multi-tier framework involving both terrestrial and aerial components, including ground BS, represents an exciting area for future research. The sum-rate optimization problem with IGS in such complex systems, incorporating NOMA and RSMA, remains a promising research avenue.

- Intelligent Reflecting Surfaces (IRS): While rate optimization problems have been studied in conventional NOMA and RSMA systems, these studies have generally not considered IGS. Additionally, existing models often focus on SISO configurations. Future research could explore the use of advanced multi-antenna setups, which could add significant complexity to rate optimization problems. The use of IGS, beamforming, and cluster formations introduces additional challenges. Moreover, managing inter-cluster interference between NOMA and RSMA clusters, which arises due to passive reflectors, presents an interesting design problem worth further exploration.

- Integrated Terrestrial-Satellite Networks: The integration of RSMA and NOMA in terrestrial-satellite communications holds great potential for B5G networks, as it offers extensive coverage and improved spectral efficiency. Future research could focus on optimizing rate and IGS coefficients for different user and system design

objectives in this context.

- Hybrid System Models: Terrestrial and Non-Terrestrial Networks: A hybrid system combining terrestrial and non-terrestrial networks (e.g., satellite, UAVs, and high-altitude platforms) could be a promising direction for the integration of NOMA, RSMA, and IGS. Future research could focus on optimizing resource allocation, beamforming, and user selection in such hybrid models, addressing the challenges of network synchronization, connectivity, and QoS management across different layers of communication.

- Mobile Edge Computing: A promising future research direction involves addressing the joint user clustering, rate allocation, and power allocation problems for uplink-downlink NOMA and RSMA systems with IGS in mobile edge computing environments, e.g. smart cities. Additionally, optimizing IGS coefficients to enhance system content delivery rates and improve user QoS is another area ripe for exploration.

- Integrated Machine-Type Communications (MTC) and NOMA/RSMA: The integration of NOMA, RSMA, and IGS in massive machine-type communication (mMTC) scenarios—such as smart cities, industrial IoT, and autonomous vehicles—could open new avenues for research. This could involve optimizing system throughput while maintaining reliability and energy efficiency for a large number of machine-type devices with varying channel conditions and traffic patterns.

- Multimodal Sensing and Communication in NOMA/RSMA/IGS Systems: Future networks might benefit from incorporating both sensing and communication capabilities in a seamless manner. Research could explore how NOMA, RSMA, and IGS can be integrated into systems that perform both communication and environmental sensing, such as for autonomous driving or smart environments. Optimizing the trade-offs between sensing quality and communication performance would be an interesting research challenge.

- AI-Driven Optimization/Innovations for NOMA, RSMA, and IGS: Artificial intelligence (AI) and machine learning (ML) techniques are gaining traction in communication systems. AI/ML could be used to optimize the combination of NOMA, RSMA, and IGS for network design, resource allocation, and interference management. Research could focus on developing deep learning models or reinforcement learning algorithms to dynamically adjust system parameters and improve performance in real-time, especially for network scenarios with high mobility or uncertain traffic patterns.

  AI offers transformative potential for NOMA and RSMA systems through various approaches. Reinforcement Learning, including multi-agent and deep RL, can dynamically optimize resource allocation, such as power control, user clustering, and subcarrier assignment. Deep learning models can approximate beamforming solutions, while Federated Learning enables decentralized optimization to preserve privacy and scalability. Graph Neural Networks enhance user clustering by modeling user interactions, and Generative Adversarial Networks improve CSI prediction for robust interference management. Transfer learning facilitates cross-domain adaptation for varying environments, and meta-learning ensures quick adaptation to dynamic conditions. Unsupervised learning extracts latent features for tasks like interference mitigation without requiring labeled data, while Explainable AI enhances transparency in resource allocation decisions. Together, these AI-driven techniques promise significant advancements in the efficiency, scalability, and adaptability of NOMA and RSMA systems for next-generation networks.

Bibliography

[1] O. Liberg et al., "Introducing 5G advanced," *IEEE Communications Standards Magazine,* vol. 8, no. 1, pp. 52-57, March 2024.

[2] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjoland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proceedings of the IEEE,* vol. 109, no. 7, pp. 1166– 1199, Jul. 2021.

[3] R. Steele and L. Hanzo, *Mobile Radio Communications: Second and Third Generation Cellular Systems,* 2nd. Hoboken, NJ, USA. Wiley, 1999.

[4] J. Li, X. Wu, and R. Laroia, *OFDMA Mobile Broadband Communications: A Systems Approach.* Cambridge, U.K. Cambridge Univ. Press, 2013.

[5] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Communications Magazine,* vol. 58, no. 3, pp. 55-61, March 2020.

[6] B. Clerckx et al., "Multiple access techniques for intelligent and multifunctional 6G: Tutorial, survey, and outlook," *Proceedings of the IEEE,* vol. 112, no. 7, pp. 832-879, July 2024.

[7] M. Na et al., "Operator's perspective on 6G: 6G services, vision, and spectrum," *IEEE Communications Magazine,* vol. 62, no. 8, pp. 178-184, August 2024.

[8] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of nonorthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.,* vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[9] Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski, and H. V. Poor, "Rate-splitting multiple access: fundamentals, survey, and future research trends", *IEEE Commun. Surveys Tuts.,* arXiv:2201.03192v1, Jan. 2022.

[10] J. G. Andrews et al., "What will 5G be?" *IEEE J. Sel. Areas Commun.,* vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[11] A. Ahmed et al., "Unveiling the potential of NOMA: A journey to next generation multiple access," *IEEE Communications Surveys Tutorials*, Early Access, 2025. doi: 10.1109/COMST.2024.3521647

[12] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. 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.

[13] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroglu and S. M. Sait, "A Survey of rate-optimal power domain NOMA With enabling technologies of future wireless networks," *IEEE Commun. Surveys Tuts.,* vol. 22, no. 4, pp. 2192-2235, Fourthquarter 2020.

[14] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "Capacity comparison between MIMO-NOMA and MIMO-OMA with multiple users in a cluster," *IEEE J. Sel. Areas in Commun.,* vol. 35, no. 10, pp. 2413–2424, Oct. 2017.

[15] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of nonorthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.,* vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[16] I. Aboumahmoud, E. Hossain and A. Mezghani, "Resource management in RIS-assisted rate splitting multiple access for next generation (xG) wireless communications: Models, state-of-the-art, and future directions," *IEEE Communications Surveys Tutorials,* 2024.

[17] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the gaussian multiple-access channel," *IEEE Trans. Inf. Theory,* vol. 42, no. 2, pp. 364–375, Mar. 1996.

[18] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory,* vol. 27, no. 1, pp. 49–60, Jan. 1981.

[19] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, "Rate splitting for MIMO wireless networks: A promising PHY-layer strategy for LTE evolution," *IEEE Commun. Mag.,* vol. 54, no. 5, pp. 98–105, May 2016.

[20] Y. Zeng, R. Zhang, E. Gunawan, and Y. L. Guan, "Optimized transmission with improper Gaussian signaling in the K-user MISO interference channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6303–6313, Dec. 2013.

[21] Y. Zeng, C. M. Yetis, E. Gunawan, Y. L. Guan, and R. Zhang, "Transmit optimization with improper Gaussian signaling for interference channels," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2899–2913, Jun. 2013.

[22] C. Lameiro, I. Santamaría, and P. J. Schreier, "Rate region boundary of the SISO Z-interference channel with improper signaling," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1022–1034, Mar. 2017.

[23] E. Kurniawan and S. Sun, "Improper Gaussian signaling scheme for the Z-interference channel," *IEEE Trans. Wireless Commun.,* vol. 14, no. 7, pp. 3912–3923, Jul. 2015.

[24] S. Javed, O. Amin, B. Shihada and M. -S. Alouini, "A journey from improper Gaussian signaling to asymmetric signaling," *IEEE Commun. Surveys & Tut.,* vol. 22, no. 3, pp. 1539–1591, Thirdquarter 2020.

[25] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals.* Cambridge, U.K.: Cambridge Univ. Press, 2010.

[26] "Ericsson mobility report June 2020," Ericsson, 2020.

[27] V. R. Cadambe, S. A. Jafar, and C. Wang, "Interference alignment with asymmetric complex signaling," *IEEE Trans. Inf. Theory,* vol. 56, no. 9, pp. 4552–4565, Sep. 2010.

[28] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, andT. Nakamura, "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access," *Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) 2013,* Naha, Japan, Nov. 2013.

[29] Y. Liu, Z. Qin, M. Elkashlan, A. Nallanathan, and J. A. McCann, "Non-orthogonal multiple access in large-scale heterogeneous networks," *IEEE Journal Sel. Areas in Commun.,* vol. 35, no. 12, pp. 2667–2680, Dec. 2017.

[30] L. P. Qian, Y. Wu, H. Zhou, and X. Shen, "Joint uplink base station association and power control for small-cell networks with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.,* vol. 16, no. 9, pp. 5567–5582, Sept. 2017.

[31] X. Yue, Z. Qin, Y. Liu, S. Kang, and Y. Chen, "A unified framework for non-orthogonal multiple access," to appear, *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5346–5359, Nov. 2018.

[32] E. Gelal, et al., "Topology control for effective interference cancellation in multiuser MIMO networks," *IEEE/ACM Trans. Netw.,* vol. 21, no. 2, pp. 455–468, Apr. 2013.

[33] C. Jiang, Y. Shi, Y. T. Hou, W. Lou, S. Kompella, and S. F. Midkiff, "Squeezing the most out of interference: An optimization framework for joint interference exploitation and avoidance," in *Proc. IEEE INFOCOM,* Mar. 2012, pp. 424–432.

[34] J. Lee, J. G. Andrews, and D. Hong, "Spectrum-sharing transmission capacity with interference cancellation," *IEEE Trans. Commun.,* vol. 61, no. 1, pp. 76–86, Jan. 2013.

[35] 3GPP, RP-170829, "New study item proposal: Study on non-orthogonal multiple access for NR," Mar. 2017.

[36] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. Wireless Commun.,* vol. 15, no. 6, pp. 4438–4454, Jun. 2016.

[37] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.,* vol. 15, no. 1, pp. 537–552, Jan. 2016.

[38] Z. Ding, P. Fan, and H. V. Poor, "Random beamforming in millimeter-wave NOMA networks," *IEEE Access,* vol. 5, pp. 7667–7681, 2017.

[39] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. 77th IEEE VTC-Spring,* Dresden, Germany, Jun. 2013, pp. 1–5.

[40] 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 Process. Lett.,* vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[41] 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.

[42] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE Annu. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC),* London, U.K., Sep. 2013, pp. 611–615.

[43] A. Benjebbour, A. Li, Y. Saito, Y. Kishiyama, A. Harada, and T. Nakamura, "System-level performance of downlink NOMA for future LTE enhancements," in

Proc. IEEE Global Commun. Conf. (GLOBECOM) Workshops, Dec. 2013, pp. 66–70.

[44] Z. Ding, "A study on the optimality of downlink hybrid NOMA," *IEEE Signal Processing Letters,* Early Access, 2025.

[45] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, "On the performance of non-orthogonal multiple access systems with partial channel information," *IEEE Trans. Commun.,* vol. 64, no. 2, pp. 654–667, Feb. 2016.

[46] S. Shi, L. Yang, and H. Zhu, "Outage balancing in downlink non-orthogonal multiple access with statistical channel state information," *IEEE Trans. Wireless Commun.,* vol. 15, no. 7, pp. 4718–4731, Jul. 2016.

[47] J. Cui, Z. Ding, and P. Fan, "A novel power allocation scheme under outage constraints in NOMA systems," *IEEE Signal Process. Lett.,* vol. 23, no. 9, pp. 1226–1230, Sep. 2016.

[48] P. Xu, Y. Yuan, Z. Ding, X. Dai, and R. Schober, "On the outage performance of non-orthogonal multiple access with one-bit feedback," *IEEE Trans. Wireless Commun.,* vol. 15, no. 10, pp. 6716–6730, Oct. 2016.

[49] Y. Zhang, H.-M. Wang, T.-X. Zheng, and Q. Yang, "Energy-efficient transmission design in non-orthogonal multiple access," *IEEE Trans. Veh. Technol.,* vol. 66, no. 3, pp. 2852–2857, Mar. 2017.

[50] F. Fang, H. Zhang, J. Cheng, and V. C. M. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. Commun.,* vol. 64, no. 9, pp. 3722–3732, May 2016.

[51] A. Powari, K. Z. Shen and D. K. C. So, "Sum rate maximization for NOMA with simultaneous cache-enabled D2D communications," *IEEE Wireless Communications Letters,* Early Access, 2024.

[52] C. L. Wang, J. Y. Chen, and Y. J. Chen, "Power allocation for a downlink non-orthogonal multiple access system," *IEEE Wireless Commun. Lett.,* vol. 5, no. 5, pp. 532–535, Oct. 2016.

[53] J. Cui, Z. Ding, and P. Fan, "A novel power allocation scheme under outage constraints in NOMA systems," *IEEE Signal Process. Lett.,* vol. 23, no. 9, pp. 1226–1230, Sep. 2016.

[54] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems," *IEEE Trans. Commun.,* vol. 65, no. 99, pp. 1077–1091, Mar. 2017.

[55] 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.

[56] F. Liu, P. Mahonen, and M. Petrova, "Proportional fairness-based power allocation and user set selection for downlink NOMA systems," in *Proc. IEEE Int. Commun. Conf. (ICC),* Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[57] J. Mei, L. Yao, H. Long, and K. Zheng, "Joint user pairing and power allocation for downlink non-orthogonal multiple access systems," in *Proc. IEEE Int. Commun. Conf. (ICC),* Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[58] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," *Proceedings of IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communica- tions (PIMRC)*, pp. 611-615, Sept. 2013.

[59] A. Benjebbour, A. Li, Y. Saito, Y. Kishiyama, A. Harada, and T. Nakamura, "Systemlevel performance of downlink NOMA for future LTE enhancements," in proc. *IEEE Globecom Workshops,* pp. 66-70, Dec. 2013.

[60]] S. Han, C.-L. I, Z. Xu, and Q. Sun, "Energy efficiency and spectrum efficiency co-design: From NOMA to network NOMA," *IEEE Commun. Soc. MMTC E-Lett.,* vol. 9, no. 5, pp. 21–24, Sep. 2014.

[61] B. Lyu, Z. Yang, G. Gui, and H. Sari, "Optimal time allocation in backscatter assisted wireless powered communication networks," *Sensors,* vol. 17, no. 6, p. 1258, Jun. 2017.

[62] G. Yang, X. Xu, and Y. Liang, "Resource allocation in NOMAenhanced backscatter communication networks for wireless powered IoT," *IEEE Wireless Communications Letters,* vol. 9, no. 1, pp. 117– 120, Jan. 2020.

[63] Y. Liao, G. Yang, and Y. Liang, "Resource allocation in NOMAenhanced full-duplex symbiotic radio networks," *IEEE Access,* vol. 8, pp. 22 709–22 720, Jan. 2020.

[64] Z. Lin, M. Lin, J. Wang, T. de Cola, and J. Wang, "Joint beamforming and power allocation for Satellite-Terrestrial integrated networks with non-orthogonal multiple

access," *IEEE Journal of Selected Topics in Signal Processing,* vol. 13, no. 3, pp. 657–670, Jun. 2019.

[65] X. Li, Q. Wang, H. Peng, H. Zhang, D. Do, K. M. Rabie, R. Kharel, and C. C. Cavalcante, "A unified framework for HS-UAV NOMA networks: Performance analysis and location optimization," *IEEE Access,* vol. 8, pp. 13 329–13 340, Jan. 2020.

[66] X. Wang, H. Chen and F. Tan, "Hybrid OMA/NOMA mode selection and resource allocation in space-air-ground integrated networks," *IEEE Transactions on Vehicular Technology,* Dec. 2024.

[67] Y. Huang, Y. Liu, and F. Chen, "NOMA-aided mobile edge computing via user cooperation," *IEEE Transactions on Communications,* vol. 68, no. 4, pp. 2221–2235, Jan. 2020.

[68] L. Xiang, D. W. K. Ng, X. Ge, Z. Ding, V. W. S. Wong, and R. Schober, "Cache-aided non-orthogonal multiple access: The two-user case," *IEEE Journal of Selected Topics in Signal Processing,* vol. 13, no. 3, pp. 436–451, Jun. 2019.

[69] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Exploiting intelligent reflecting surfaces in multi-antenna aided NOMA systems," *arXiv preprint arXiv:1910.13636,* Oct. 2019.

[70] G. Yang, X. Xu, and Y.-C. Liang, "Intelligent reflecting surface assisted non-orthogonal multiple access," *IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Korea (South),* Jul. 2020.

[71] M. Cheng, Y. Zhou, S. Wei and S. Qian, "Outage probability analysis of cooperative NOMA with successive refinement," *IEEE Transactions on Broadcasting,* Early Access, 2024.

[72] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.,* vol. 19, no. 8, pp. 1462–1465, Aug. 2015.

[73] Z. Ding, H. Dai, and H. V. Poor, "Relay selection for cooperative NOMA," *IEEE Wireless Commun. Lett.*, vol. 5, no. 4, pp. 416–419, Aug. 2016.

[74] S. Meng, X. Wang, H. Zhang, Z. Qian, Y. Zou and Y. Liu, "Capacity Enhancement for D2D-assisted cooperative NOMA systems," *IEEE Transactions on Communications*, Early Access, 2024.

[75] J. Kim and I. Lee, "Capacity analysis of cooperative relaying systems using non-orthogonal multiple access," *IEEE Communications Letters,* vol. 19, no. 11, pp. 1949–1952, Nov. 2015.

[76] P. Parida and S. Das, allocation in OFDM based NOMA system: A DC programming approach," in proc. *IEEE Globecom Workshops,* pp. 1026-1031, Dec. 2014.

[77] M. Zeng, W. Hao, O. A. Dobre and Z. Ding, "Cooperative NOMA: state of the art, key techniques, and open challenges," *IEEE Network,* vol. 34, no. 5, pp. 205–211, Sept. 2020.

[78] L. Dai et al., "Self-interference mitigation in FD-NOMA networks," *IEEE Communications Surveys and Tutorials,* vol. 20, no. 3, pp. 2110-2135, 2018.

[79] S. M. Islam et al., "Energy-efficient designs for FD-NOMA with SWIPT," *IEEE Transactions on Wireless Communications,* vol. 16, no. 4, pp. 2752-2765, 2017.

[80] Y. Liu et al., "Cooperative relaying techniques for FD-NOMA," *IEEE Transactions on Communications,* vol. 67, no. 4, pp. 2963-2977, 2019.

[81] X. Chen et al., "Physical layer security for FD-NOMA: Challenges and solutions," *IEEE Transactions on Information Forensics and Security,* vol. 15, pp. 2142-2156, 2020.

[82] X. Zhang and F. Wang, "Resource allocation for wireless power transmission over full-duplex OFDMA/NOMA mobile wireless networks," *IEEE Journal on Selected Areas in Communications,* vol. 37, no. 2, pp. 327–344, Feb. 2019.

[83] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems," *IEEE Transactions on Communications,* vol. 65, no. 3, pp. 1077–1091, Mar. 2017.

[84] G. Liu, Z. Wang, J. Hu, Z. Ding, and P. Fan, "Cooperative NOMA broadcasting/multicasting for low-latency and high-reliability 5G cellular V2X communications," *IEEE Internet of Things Journal,* vol. 6, no. 5, pp. 7828–7838, Oct. 2019.

[85] W. Wu, X. Yin, P. Deng, T. Guo, and B. Wang, "Transceiver design for downlink SWIPT NOMA systems with cooperative full-duplex relaying," *IEEE Access,* vol. 7, pp. 33 464–33 472, Mar. 2019.

[86] G. Liu, X. Chen, Z. Ding, Z. Ma, and F. R. Yu, "Hybrid halfduplex/ full-duplex cooperative non-orthogonal multiple access with transmit power adaptation," *IEEE Transactions on Wireless Communications,* vol. 17, no. 1, pp. 506–519, Jan. 2018.

[87] M. M. Salim, S. I. Al-Dharrab, D. Benevides da Costa and A. H. Muqaibel, "Rate-energy optimization for hybrid-powered full-duplex relays in cognitive C-NOMA with impairments," *IEEE Open Journal of the Communications Society,* vol. 5, pp. 7419-7433, 2024.

[88] A. Carleial, "Interference channels," *IEEE Trans. Inf. Theory,* vol. 24, no. 1, pp. 60–70, Jan. 1978.

[89] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory,* vol. 27, no. 1, pp. 49– 60, Jan. 1981.

[90] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inf. Theory,* vol. 42, no. 2, pp. 364–375, Mar. 1996.

[91] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory,* vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[92] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate splitting approach," *IEEE Trans. Commun.,* vol. 64, no. 11, pp. 4847– 4861, Nov. 2016.

[93] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, "Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," *IEEE Trans. Inf. Theory,* vol. 59, no. 1, pp. 315–328, Jan. 2013.

[94] H. Joudeh and B. Clerckx, "Robust transmission in downlink multiuser MISO systems: A rate-splitting approach," *IEEE Trans. Signal Process.,* vol. 64, no. 23, pp. 6227–6242, Dec. 2016.

[95] S. Yang and Z. Lit, "A constant-gap result on the multi-antenna broadcast channels with linearly precoded rate splitting," in Proc. *IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC),* 2018.

[96] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting multiple access for downlink communication systems: bridging, generalizing, and outperforming SDMA and NOMA," *EURASIP J. Wireless Commun. Netw.,* vol. 2018, no. 1, p. 133, May 2018.

[97] B. Clerckx, Y. Mao, R. Schober, and H. V. Poor, "Rate-splitting unifying SDMA, OMA, NOMA, and multicasting in MISO broadcast channel: A simple two-user rate analysis," *IEEE Wireless Commun. Lett.,* vol. 9, no. 3, pp. 349–353, Mar. 2020.

[98] O. Dizdar, Y. Mao, W. Han, and B. Clerckx, "Rate-splitting multiple access for downlink multi-antenna communications: Physical layer design and link-level simulations," in Proc. *IEEE Annu. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC),* 2020.

[99] B. Clerckx, Y. Mao, R. Schober, E. A. Jorswieck, D. J. Love, J. Yuan, L. Hanzo, G. Y. Li, E. G. Larsson, and G. Caire, "Is NOMA efficient in multi-antenna networks? A critical look at next generation multiple access techniques," *IEEE Open J. Commun. Soc.,* vol. 2, pp. 1310–1343, June 2021.

[100] O. Dizdar, Y. Mao, W. Han, and B. Clerckx, "Rate-splitting multiple access: A new frontier for the PHY layer of 6G," in *Proc. IEEE 92nd Veh. Technol. Conf. (VTC Fall),* 2020.

[101] O. Dizdar, Y. Mao, Y. Xu, P. Zhu, and B. Clerckx, "Rate-splitting multiple access for enhanced URLLC and eMBB in 6G," in Proc. *IEEE Int. Symp. Wireless Commun. Syst. (ISWCS),* 2021.

[102] O. Dizdar, Y. Mao, and B. Clerckx, "Rate-splitting multiple access to mitigate the curse of mobility in (massive) MIMO networks," *IEEE Trans. Commun.,* vol. 69, no. 10, pp. 6765–6780, Oct. 2021.

[103] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, "A rate splitting strategy for massive MIMO with imperfect CSIT," *IEEE Trans. Wireless Commun.,* vol. 15, no. 7, pp. 4611–4624, July 2016.

[104] M. Dai and B. Clerckx, "Multiuser millimeter wave beamforming strategies with quantized and statistical CSIT," *IEEE Trans. Wireless Commun.,* vol. 16, no. 11, pp. 7025–7038, Nov. 2017.

[105] A. Zappone, B. Matthiesen, and E. A. Jorswieck, "Energy efficiency in MIMO underlay and overlay device-to-device communications and cognitive radio systems," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 1026–1041, Feb. 2017.

[106] A. Papazafeiropoulos and T. Ratnarajah, "Rate-splitting robustness in multi-pair massive MIMO relay systems," *IEEE Trans. Wireless Commun.,* vol. 17, no. 8, pp. 5623–5636, Aug. 2018.

[107] Z. Yang, J. Shi, Z. Li, M. Chen, W. Xu, and M. Shikh-Bahaei, "Energy efficient rate splitting multiple access (RSMA) with reconfigurable intelligent surface," in Proc. *IEEE Int. Conf. Commun. (ICC) Workshop,* 2020.

[108] A. A. Ahmad, J. Kakar, R. Reifert, and A. Sezgin, "UAV-assisted CRAN with rate splitting under base station breakdown scenarios," in Proc. *IEEE Int. Conf. Commun. (ICC)* Workshop, 2019.

[109] A. Alameer Ahmad, H. Dahrouj, A. Chaaban, A. Sezgin, and M. Alouini, "Interference mitigation via rate-splitting and common message decoding in cloud radio access networks," *IEEE Access,* vol. 7, pp. 80 350–80 365, June 2019.

[110] Z. Yang, J. Shi, Z. Li, M. Chen, W. Xu, and M. Shikh-Bahaei, "Energy efficient rate splitting multiple access (RSMA) with reconfigurable intelligent surface," in Proc. *IEEE Int. Conf. Commun. (ICC) Workshop,* 2020.

[111] A. Bansal, K. Singh, and C.-P. Li, "Analysis of hierarchical rate splitting for intelligent reflecting surfaces-aided downlink multiuser MISO communications," *IEEE Open J. Commun. Soc,* vol. 2, pp. 785– 798, Apr. 2021.

[112] A. A. Ahmad, B. Matthiesen, A. Sezgin, and E. Jorswieck, "Energy efficiency in C-RAN using rate splitting and common message decoding," in Proc. *IEEE Int. Conf. Commun. (ICC) Workshop,* 2020.

[113] A. A. Ahmad, Y. Mao, A. Sezgin, and B. Clerckx, "Rate splitting multiple access in C-RAN," in Proc. *IEEE Annu. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC),* 2020.

[114] A. A. Ahmad, H. Dahrouj, A. Chaaban, A. Sezgin, T. Y. Al-Naffouri, and M.-S. Alouini, "Power minimization via rate splitting in downlink cloud-radio access networks," in Proc. *IEEE Int. Conf. Commun. (ICC) Workshop,* 2020.

[115] W. Jaafar, S. Naser, S. Muhaidat, P. C. Sofotasios, and H. Yanikomeroglu, "Multiple access in aerial networks: From orthogonal and non-orthogonal to rate-splitting," *IEEE Open J. of Veh. Technol.,* vol. 1, pp. 372–392, Oct. 2020.

[116] W. Jaafar, S. Naser, S. Muhaidat, P. C. Sofotasios, and H. Yanikomeroglu, "On the downlink performance of RSMA-based UAV communications," *IEEE Trans. Veh. Technol.,* vol. 69, no. 12, pp. 16 258–16 263, Dec. 2020.

[117] S. Zhang, Y. Mao, B. Clerckx and T. Q. S. Quek, "Interference management in space-Air-Ground Integrated networks with fully distributed rate-splitting multiple access," *IEEE Transactions on Wireless Communications,* Early Access, 2024.

[118] M. Z. Hassan, M. J. Hossain, J. Cheng, and V. C. M. Leung, "Device-clustering and rate-splitting enabled device-to-device cooperation framework in fog radio access

network," *IEEE Trans. Green Commun. Netw.,* vol. 5, no. 3, pp. 1482–1501, Sept. 2021.

[119] O. Tervo, L. Trant, S. Chatzinotas, B. Ottersten, and M. Juntti, "Multigroup multicast beamforming and antenna selection with ratesplitting in multicell systems," in Proc. *IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC),* 2018.

[120] H. Joudeh and B. Clerckx,, "Rate-splitting for max-min fair multigroup multicast beamforming in overloaded systems," *IEEE Trans. Wireless Commun.,* vol. 16, no. 11, pp. 7276–7289, Nov. 2017.

[121] A. Rahmati, Y. Yapici, N. Rupasinghe, I. Guvenc, H. Dai, and A. Bhuyan, "Energy efficiency of RSMA and NOMA in cellular connected mmWave UAV networks," in Proc. *IEEE Int. Conf. Commun. (ICC) Workshop,* 2019.

[122] S. Tao, H. Yu, Q. Li, Y. Tang, and D. Zhang, "One-layer rate-splitting multiple access with benefits over power-domain NOMA in indoor multi-cell visible light communication networks," in Proc. *IEEE Int. Conf. Commun. (ICC) Workshop,* 2020.

[123] S. Ma, H. Zhou, Y. Mao, X. Liu, Y. Wu, B. Clerckx, Y. Wang, and S. Li, "Robust beamforming design for rate splitting multiple access-aided MISO visible light communications," *arXiv preprint arXiv:2108.07014,* 2021.

[124] J. Cao and E. M. Yeh, "Asymptotically optimal multiple-access communication via distributed rate splitting," *IEEE Trans. Inf. Theory,* vol. 53, no. 1, pp. 304–319, Jan. 2007.

[125] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Trans. Commun.,* vol. 64, no. 11, pp. 4847–4861, Nov. 2016.

[126] Y. Mao, B. Clerckx, and V. O. K. Li, "Energy efficiency of rate-splitting multiple access, and performance benefits over SDMA and NOMA," in Proc. *IEEE Int. Symp. Wireless Commun. Syst. (ISWCS),* 2018.

[127] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: Spectral and energy efficiency analysis," *IEEE Trans. Commun.,* 2019.

[128] Z. Ho and E. Jorswieck, "Improper Gaussian signaling on the two-user SISO interference channel," *IEEE Trans. Wireless Commun.,* vol. 11, no. 9, pp. 3194–3203, Sep. 2012.

[129] Y. Zeng, R. Zhang, E. Gunawan, and Y. Guan, "Optimized transmission with improper Gaussian signaling in the K-user MISO interference channel," *IEEE Trans. Wireless Commun.,* vol. 12, no. 12, pp. 6303–6313, Dec. 2013.

[130] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory,* vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[131] E. Kurniawan and S. Sun, "Improper Gaussian signaling scheme for the Z-interference channel," *IEEE Trans. Wireless Commun.,* vol. 14, no. 7, pp. 3912–3923, Jul. 2015.

[132] S. Lagen, A. Agustin, and J. Vidal, "Improper Gaussian signaling for the Z-interference channel," in Proc. *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP),* Florence, Italy, May 2014, pp. 1140–1144.

[133] M. Gaafar, M. G. Khafagy, O. Amin, and M. S. Alouini, "Improper Gaussian signaling in full-duplex relay channels with residual selfinterference," in Proc. *IEEE Int. Conf. Commun. (ICC),* Kuala Lumpur, Malaysia, 2016, pp. 1–7.

[134] M. Gaafar, O. Amin, A. Ikhlef, A. Chaaban, and M.-S. Alouini, "On alternate relaying with improper Gaussian signaling," *IEEE Commun. Lett.,* vol. 20, no. 8, pp. 1683–1686, Aug. 2016.

[135] O. Amin, W. Abediseid, and M. Alouini, "Overlay spectrum sharing using improper Gaussian signaling," *IEEE Journal Sel. Areas in Commun.,* vol. 35, no. 1, pp. 50–62, Jan. 2017.

[136] S. Javed, O. Amin, B. Shihada, and M.-S. Alouini, "Improper Gaussian signaling for hardware impaired multihop full-duplex relaying systems," *IEEE Trans. Commun.,* vol. 67, no. 3, pp. 1858–1871, Mar. 2019.

[137] C. Kim, E.-R. Jeong, Y. Sung, and Y. H. Lee, "Asymmetric complex signaling for full-duplex decode-and-forward relay channels," in Proc. *Int. Conf. ICT Converg. (ICTC),* Jeju Island, South Korea, Oct. 2012, pp. 28–29.

[138] S. Javed, O. Amin, S. S. Ikki, and M.-S. Alouini, "Multiple antenna systems with hardware impairments: New performance limits," *IEEE Trans. Veh. Technol.,* vol. 68, no. 2, pp. 1593–1606, Feb. 2019.

[139] S. Javed, O. Amin, S. S. Ikki, and M.-S. Alouini, "Asymmetric hardware distortions in receive diversity systems: Outage performance analysis," *IEEE Access,* vol. 5, pp. 4492–4504, 2017.

[140] A. Kariminezhad, A. Chaaban, and A. Sezgin, "Improper signaling and symbol extensions: How far can we go with Gaussian P2P codebooks in the interfering MAC with TIN?" [Online]. Available:arXiv:1607.01995.2016.

[141] A. A. Nasir, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Improper Gaussian signaling for broadcast interference networks," *IEEE Signal Process. Lett.,* vol. 26, no. 6, pp. 808–812, Jun. 2019.

[142] A. Kariminezhad, A. Sezgin, and M. Pesavento, "Power efficiency of improper signaling in MIMO full-duplex relaying for K-user interference networks," in Proc. *IEEE Int. Conf. Commun. (ICC),* Paris, France, May 2017, pp. 1–6.

[143] J. Zhang and M. Haardt, "Widely linear signal processing for twoway relaying with MIMO amplify and forward relays," in Proc. *10th Int. Symp. Wireless Commun. Syst. (ISWCS),* Ilmenau, Germany, Aug. 2013, pp. 1–5.

[144] M. Soleymani, P. J. Schreier, and I. Santamaria, "Performance analysis of MIMO K-user interference channels with hardware impairments," 2020.

[145] I. Santamaria, P. M. Crespo, C. Lameiro, and P. J. Schreier, "Information-theoretic analysis of a family of improper discrete constellations," *Entropy,* vol. 20, no. 1, pp. 1–22, Jan. 2018.

[146] J. A. Lopez-Fernandez, R. G. Ayestaran, I. Santamaria, and C. Lameiro, "Design of asymptotically optimal improper constellations with hexagonal packing," *IEEE Trans. Commun.,* vol. 67, no. 8, pp. 5445–5457, Aug. 2019.

[147] H. T. Nguyen, H. D. Tuan, D. Niyato, D. I. Kim and H. V. Poor, "Improper Gaussian signaling for D2D communication coexisting MISO cellular networks," *IEEE Trans. Wireless Commun.,* vol. 20, no. 8, pp. 5186–5198, Aug. 2021.

[148] L. Hanzo, M. El-Hajjar, and O. Alamri, "Near-capacity wireless transceivers and cooperative communications in the MIMO era: Evolution of standards, waveform design, and future perspectives," *Proc. IEEE,* vol. 99, no. 8, pp. 1343–1385, Aug. 2011.

[149] Y. Cai, Z. Qin, F. Cui, G. Y. Li, andJ. A. McCann, "Modulation and multiple access for 5G networks," *IEEE Commun. Surveys Tuts.,* vol. 20, no. 1, pp. 629–646, First quarter 2018.

[150] Y. Zeng, C. M. Yetis, E. Gunawan, Y. L. Guan, and R. Zhang, "Transmit optimization with improper Gaussian signaling for interference channels," *IEEE Trans. Sig. Process.,* vol. 61, no. 11, pp. 2899–2913, Jun. 2013.

[151] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," *Lecture Notes of EE392o, Stanford University, Autumn Quarter,* 2003.

[152] S. Lim and J. Zhu, "Integrated data envelopment analysis: Global vs. local optimum," *European Journal of Operational Research,* vol. 229, no. 1, pp. 276–278, Aug. 2013.

[153] X. Zhang, Q. Gao, C. Gong, and Z. Xu, "User grouping and power allocation for NOMA visible light communication multi-cell networks," *IEEE Commun. Lett.,* vol. 21, no. 99, pp. 777–780, Mar. 2016.

[154] Y. Zeng, C. M. Yetis, E. Gunawan, Y. L. Guan, and R. Zhang, "Transmit optimization with improper Gaussian signaling for interference channels," *IEEE Trans. Sig. Process.,* vol. 61, no. 11, pp. 2899–2913, Jun. 2013.

[155] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting multiple access for downlink communication systems: Bridging, generalizing, and outperforming SDMA and NOMA," *EURASIP J. Wireless Commun. Netw.,* no. 1, pp. 1–54, May 2018.

[156] S. Javed, O. Amin, B. Shihada, and M. Alouini, "A Journey from improper Gaussian signaling to asymmetric signaling", *IEEE Commun. Surveys Tut.,* vol. 22, no. 3, pp. 1539–1591, third quarter, 2020.

[157] H. D. Tuan, A. A. Nasir, H. H. Nguyen, T. Q. Duong, and H. V. Poor, "Non-Orthogonal multiple access with improper Gaussian signaling," *IEEE Journal Sel. Topics in Sig. Process.,* vol. 13, no. 3, pp. 496–507, Jun. 2019.

[158] S. Hong and S. Bahk, "Performance analysis and fairness maximization in NOMA systems with improper Gaussian signaling under imperfect successive interference cancellation", *IEEE Access*, vol. 8, pp. 50439–50451, 2020.

[159] A. A. Nasir, H. D. Tuan, H. H. Nguyen, T. Q. Duong and H. V. Poor, "Signal superposition in NOMA with proper and improper Gaussian signaling," *IEEE Trans. on Commun.* vol. 68, no. 10, pp. 6537–6551, Oct. 2020

[160] I. Abu Mahady, E. Bedeer, S. Ikki, and H. Yanikomeroglu, "Sum-rate maximization of NOMA systems under imperfect successive interference cancellation," *IEEE Commun. Lett.,* vol. 23, no. 3, pp. 474–477, Mar. 2019.

[161] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," *Lecture Notes of EE392o, Stanford University, Autumn Quarter,* 2003.

[162] J. More and M. Cosnard, "Numerical solution of nonlinear equations," *ACM Trans. Mathematical Software,* vol. 5, no. 1, pp. 64–85, Mar. 1979.

[163] X. Liu and X. Wang, "Outage probability and capacity analysis of the collaborative NOMA assisted relaying system in 5G," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC),* Jul. 2016, pp. 1–5.

[164] J. So and Y. Sung, "Improving non-orthogonal multiple access by forming relaying broadcast channels," *IEEE Commun. Lett.,* vol. 20, no. 9, pp. 1816–1819, Sep. 2016.

[165] Z. Zhang, Z. Ma, M. Xiao, Z. Ding, and P. Fan, "Full-duplex device-to-device-aided cooperative nonorthogonal multiple access," *IEEE Trans. Veh. Technol.,* vol. 66, no. 5, pp. 4467–4471, May 2017.

[166] L. Zhang, J. Liu, M. Xiao, G. Wu, Y. Liang and S. Li, "Performance analysis and optimization in downlink NOMA systems with cooperative full-duplex relaying," *IEEE J. Sel. Areas Commun.,* vol. 35, no. 10, pp. 2398–2412, Oct. 2017.

[167] C. Zhong and Z. Zhang, "Non-orthogonal multiple access with cooperative full-duplex relaying," *IEEE Commun. Lett.,* vol. 20, no. 12, pp. 2478–2481, Dec. 2016.

[168] S. Lagen, A. Agustin, and J. Vidal, "On the superiority of improper Gaussian signaling in wireless interference MIMO scenarios," *IEEE Trans. Commun.,* vol. 64, no. 8, pp. 3350–3368, Aug. 2016.

[169] H.-Y. Shin, S.-H. Park, H. Park, and I. Lee, "A new approach of interference alignment through asymmetric complex signaling and multiuser diversity," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 880–884, Mar. 2012.

[170] A. A. Nasir, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Improper Gaussian signaling for broadcast interference networks," *IEEE Signal Process. Lett.,* vol. 26, no. 6, pp. 808–812, Jun. 2019.

[171] C. Lameiro, I. Santamaria, and P. J. Schreier, "Benefits of improper signaling for underlay cognitive radio," *IEEE Wireless Commun. Lett.,* vol. 4, no. 1, pp. 22–25, Feb. 2015.

[172] O. Amin, W. Abediseid, and M.-S. Alouini, "Underlay cognitive radio systems with improper Gaussian signaling: Outage performance analysis," *IEEE Trans. Wireless Commun.,* vol. 15, no. 7, pp. 4875–4887, Jul. 2016.

[173] C. Lameiro, I. Santamaria, and P. J. Schreier, "Improper Gaussian signaling for multiple-access channels in underlay cognitive radio," *IEEE Trans. Commun.,* vol. 67, no. 3, pp. 1817–1830, Mar. 2019.

[174] M. Gaafar, O. Amin, W. Abediseid, and M.-S. Alouini, "Underlay spectrum sharing techniques with in-band full-duplex systems using improper Gaussian signaling," *IEEE Trans. Wireless Commun.,* vol. 16, no. 1, pp. 235–249, Jan. 2017.

[175] M. Gaafar, M. G. Khafagy, O. Amin, R. F. Schaefer, and M.-S. Alouini, "Full-duplex relaying with improper Gaussian signaling over Nakagami-m fading channels," *IEEE Trans. Commun.,* vol. 66, no. 1, pp. 64–78, Jan. 2018.

[176] S. G. Hong and S. Bahk, "Performance analysis and fairness maximization in NOMA systems with improper Gaussian signaling under imperfect successive interference cancellation," *IEEE Access,* vol. 8, pp. 50439–50451, 2020.

[177] H. D. Tuan, A. A. Nasir, H. H. Nguyen, T. Q. Duong and H. V. Poor, "Non-orthogonal multiple access with improper Gaussian signaling," *IEEE J. Sel. Topics Signal Process.,* vol. 13, no. 3, pp. 496–507, Jun. 2019.

[178] H. Yu, H. D. Tuan, T. Q. Duong, Y. Fang and L. Hanzo, "Improper Gaussian signaling for integrated data and energy networking," *IEEE Trans. Commun.,* vol. 68, no. 6, pp. 3922–3934, Jun. 2020,

[179] I. Abu Mahady, E. Bedeer, S. Ikki and H. Yanikomeroglu, "Sum-rate maximization of NOMA systems under imperfect successive interference cancellation," *IEEE Commun. Lett.,* vol. 23, no. 3, pp. 474–477, Mar. 2019.

[180] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," *Lecture Notes of EE392o, Stanford University,* Autumn Quarter, 2003.

[181] W. Dinkelbach, "On nonlinear fractional programming," *Manag. Sci.,* vol. 13, no. 7, pp. 492–498, Mar. 1967.

[182] S. Sun et al., "Propagation path loss models for 5G urban micro- and macro-cellular scenarios," in *Proc. IEEE Veh. Tech. Conf.,* May 2016, pp. 1–6.

[183] Z. Yang, M. Chen, W. Saad and M. Shikh-Bahaei, "Optimization of rate allocation and power control for rate splitting multiple access (RSMA)," *IEEE Trans. Commun.,* vol. 69, no. 9, pp. 5988–6002, Sept. 2021.

Appendix A

Proofs of Chapter 4

A.1 Sum Rate Maximization Case 1: The proof and the value of $\hat{\rho}$ in (4.27)

To get the regions for $\mathcal{R}_{1,2} < \mathcal{R}_{2,2}$ or $\mathcal{R}_{2,2} < \mathcal{R}_{1,2}$, we need to find a value $\tilde{\rho}$, such that $0 < \tilde{\rho} < 1$ where each case exists. From (A.6), (A.7), we solve

$$(\hat{P}\Gamma_2 + 1)^2 - (\hat{P}\Gamma_2\xi_2)^2 < \left[\frac{\Lambda - \left[(\rho_1 P\Gamma_1\xi_1 + \rho_2 P\Gamma_1\xi_2)^2 + (\kappa\hat{P}\Gamma_s\xi_2)^2\right]}{\Upsilon - (\kappa\hat{P}\Gamma_s\xi_2)^2 - (\rho_1 P\Gamma_1\xi_1)^2}\right]. \tag{A.1}$$

After a few mathematical simplifications, we get

$$\begin{aligned}
&\rho^2\left[(P\Gamma_1)^2\left((1-\xi_2^2)((\hat{P}\Gamma_2+1)^2 - (\hat{P}\Gamma_2\xi_2)^2) + \xi_1^2 + \xi_2^2 + \xi_1\xi_2\right)\right] \\
&+ \rho\left[2(P\Gamma_1)\left((\kappa\hat{P}\Gamma_s+1)((\hat{P}\Gamma_2+1)^2 - (\hat{P}\Gamma_2\xi_2)^2) - P\Gamma_1\xi_2^2 + P\Gamma_1\xi_1\xi_2\right)\right] \\
&+ \left[((\kappa\hat{P}\Gamma_s+1)^2 - (\kappa\hat{P}\Gamma_s\xi_2)^2)((\hat{P}\Gamma_2+1)^2 - (\hat{P}\Gamma_2\xi_2)^2) + (P\Gamma_1)^2\xi_2^2 + (\kappa\hat{P}\Gamma_s)^2\xi_2^2 - (P\Gamma_1 + \kappa\hat{P}\Gamma_s + 1)^2\right] < 0.
\end{aligned} \tag{A.2}$$

By solving the quadratic equation in (A.2), we find the solution of $\hat{\rho}$ as shown below

$$\tilde{\rho} = \frac{-T + \sqrt{(T^2 - 4UV)}}{2U}, \tag{A.3}$$

where

$$U = (P\Gamma_1)^2 \times \left((1-\xi_2^2)((\hat{P}\Gamma_2+1)^2 - (\hat{P}\Gamma_2\xi_2)^2) + \xi_1^2 + \xi_2^2 + \xi_1\xi_2\right), \tag{A.4a}$$

$$V = ((\kappa\hat{P}\Gamma_s+1)^2 - (\kappa\hat{P}\Gamma_s\xi_2)^2)((\hat{P}\Gamma_2+1)^2 - (\hat{P}\Gamma_2\xi_2)^2) + (P\Gamma_1)^2\xi_2^2 + (\kappa\hat{P}\Gamma_s)^2\xi_2^2 - (P\Gamma_1 + \kappa\hat{P}\Gamma_s + 1)^2 \tag{A.4b}$$

$$T = 2(P\Gamma_1)\left((\kappa\hat{P}\Gamma_s+1)((\hat{P}\Gamma_2+1)^2 - (\hat{P}\Gamma_2\xi_2)^2) - P\Gamma_1\xi_2^2 + P\Gamma_1\xi_1\xi_2\right). \tag{A.4c}$$

Without loss of generality, and for the sake of mathematical simplicity, we assume $\angle\hat{C}_1 = \angle\hat{C}_2$ [1] and a phase difference of $\pi/2$ between $\angle\Gamma_s$ and $\angle\Gamma_1$,[2] then, the rates in (4.13), (4.15), (4.21) are derived as follows

$$\mathcal{R}_1 = \frac{1}{2}\log_2\left[\frac{\Upsilon - (\rho_1 P\Gamma_1\xi_1)^2 - \left(\kappa\hat{P}\Gamma_s\xi_2\right)^2}{\left(\kappa\hat{P}\Gamma_s+1\right)^2 - \left(\kappa\hat{P}\Gamma_s\xi_1\right)^2}\right]. \tag{A.5}$$

$$\mathcal{R}_{1,2} = \frac{1}{2}\log_2\left[\frac{\Lambda - \left[(\rho_1 P\Gamma_1\xi_1 + \rho_2 P\Gamma_1\xi_2)^2 + (\kappa\hat{P}\Gamma_s\xi_2)^2\right]}{\Upsilon - (\kappa\hat{P}\Gamma_s\xi_2)^2 - (\rho_1 P\Gamma_1\xi_1)^2}\right], \tag{A.6}$$

$$\mathcal{R}_{2,2} = \frac{1}{2}\log_2\left[\left(\hat{P}\Gamma_2\right)^2\left(1-\xi_2^2\right) + 2\left(\hat{P}\Gamma_2\right) + 1\right]. \tag{A.7}$$

It is worth mentioning that if we substitute $\xi_1 = \xi_2 = 0$ into (A.5), (A.6), and (A.7), we get well-known rate expressions seen in the case of PGS.

[1] phase of circularity of coefficient is not an optimization parameter.

[2] We assume controlled signal design, i.e., the cooperating FD-user uses a pre-defined beamforming or phase adjustment strategy to align the user channel and SI channel with a $\pi/2$ phase difference.

To find optimal ρ in *Case 1*, (4.26) results in the following equation

$$(\lambda_1+w_1)(((F\rho P\Gamma_1\xi_1^2)(\kappa\hat{P}\Gamma_s+1)^2-(\kappa\hat{P}\Gamma_s)^2\xi_2^2+F-(\kappa\hat{P}\Gamma_s\xi_2)^2-(\rho P\Gamma_1\xi_1)^2)(G-((1-\rho)\Gamma_1\xi_2)^2-(\kappa\hat{P}\Gamma_s\xi_2)^2$$
$$-(\rho P\Gamma_1\xi_1)^2)(F-(\kappa\hat{P}\Gamma_s\xi_2)^2-(\rho P\Gamma_1\xi_1)^2))+(\lambda_2+w_2)((((1-\rho)\Gamma_1\xi_2^2-\rho P\Gamma_1\xi_1^2)(F-(\kappa\hat{P}\Gamma_s\xi_2)^2-(\rho P\Gamma_1\xi_1)^2)$$
$$-(F-\rho P\Gamma_1\xi_1^2)(G-(\kappa\hat{P}\Gamma_s\xi_2)^2-((1-\rho)P\Gamma_1\xi_2)^2-(\rho P\Gamma_1\xi_1)^2))(F-(\kappa\hat{P}\Gamma_s\xi_2)^2-(\rho P\Gamma_1\xi_1)^2)((\kappa\hat{P}\Gamma_s+1)^2$$
$$-(\kappa\hat{P}\Gamma_s)^2\xi_2^2))=0, \tag{A.8}$$

where $F=(\rho P\Gamma_1+\hat{P}\Gamma_s+1)^2$, $G=((\rho P+(1-\rho)P)\Gamma_1+\hat{P}\Gamma_s+1)^2$. To continue, the derivation with respect to ξ_1 and ξ_2 in (4.25), respectively, results in

$$a\xi_1^3+b\xi_1^2+c\xi_1+d=0, \tag{A.9}$$

where

$$\begin{aligned}
a=&\rho^3(P\Gamma_1)^4(-\rho(\lambda_1+w_1)+2(\lambda_2+w_2)(1-\rho)\xi_2),\\
b=&\rho^3(1-\rho)(P\Gamma_1)^4\xi_2(-2(\lambda_1+w_1)-(\lambda_2+w_2))\\
c=&(\rho P\Gamma_1)^2((\lambda_1+w_1)(\Lambda-((1-\rho)P\Gamma_1\xi_2)^2-(\kappa\hat{P}\Gamma_s\xi_2)^2)+(\lambda_2+w_2)(\Upsilon-\Lambda+(1-\rho)P\Gamma_1\xi_2)^2)\\
d=&\rho(1-\rho)(P\Gamma_1)^2\xi_2(\Upsilon-(\kappa\hat{P}\Gamma_s\xi_2)^2)(\lambda_2+w_2).
\end{aligned}$$

and

$$-(\lambda_1+w_1)\xi_2(\kappa\hat{P}\Gamma_s)^2(\Upsilon-(\kappa\hat{P}\Gamma_s\xi_2)^2-(\rho_1P\Gamma_1\xi_1)^2)(\Lambda-(\rho_1P\Gamma_1\xi_1+\rho_2P\Gamma_1\xi_2)^2+(\kappa\hat{P}\Gamma_s\xi_2)^2)$$
$$-(\lambda_2+w_2)(((\kappa\hat{P}\Gamma_s+1)^2-(\kappa\hat{P}\Gamma_s\xi_1)^2)(\xi_2(\kappa\hat{P}\Gamma_s)^2(\Lambda-\Upsilon-(1-\rho)P\Gamma_1\xi_2)^2$$
$$-2(\rho P\Gamma_1\xi_1)(1-\rho)P\Gamma_1\xi_2)^2-((1-\rho)P\Gamma_1)(\rho P\Gamma_1\xi_1+(1-\rho)P\Gamma_1\xi_2)$$
$$(\Upsilon-((\rho P\Gamma_1\xi_1)^2+(\kappa\hat{P}\Gamma_s\xi_2)^2)))=0. \tag{A.10}$$

A.2 Sum Rate Maximization Case 2: the derivation with respect to ξ_2

From (4.25), the derivation with respect to ξ_2, after a few algebraic simplification, we get

$$\xi_2 \left\{ a_{11}\xi_2^4 + b_{11}\xi_2^2 + c_{11} \right\} = 0, \tag{A.11}$$

where a_{11}, b_{11}, c_{11} are

$$\begin{aligned}
a_{11} =& (\lambda_2 + w_2)(\Gamma_2\hat{P})^2 \left[(\Gamma_s\hat{P})^2 \right], \\
b_{11} =& (\hat{P}\Gamma_2)^2 \left[(\kappa\hat{P}\Gamma_s)^2 \right] \left\{ (\lambda_2 - \lambda_1) \left[(\rho P\Gamma_1)^2 + \kappa\hat{P}\Gamma_s + 1) \right] + 2(\lambda_2 + w_2)(\kappa\hat{P}\Gamma_s + 1)^2 \right\}, \\
c_{11} =& (\lambda_2 + w_2)(\hat{P}\Gamma_2)^2 \left[(\rho P\Gamma_1 + \kappa\hat{P}\Gamma_s + 1)^2(\kappa\hat{P}\Gamma_s + 1)^2 \right] - (\lambda_2 + w_1) \left[(\kappa\hat{P}\Gamma_s)^2 \right] \\
& \left[(\rho P\Gamma_1)^2 + \kappa\hat{P}\Gamma_s + 1) \right] (\hat{P}\Gamma_2 + 1)^2.
\end{aligned}$$

A.3 Max-MIN Case 1: the derivation with respect to ρ, ξ_i

We derive (4.33) with respect to ρ and we get (A.12) as follows

$$\begin{aligned}
\frac{\partial \mathcal{L}_{\mathcal{X}}}{\partial \rho} =& (-\lambda_1)\big((\rho(P\Gamma_1\xi_1)^2(1-\xi_1^2) + (\kappa\hat{P}\Gamma_s + 1)P\Gamma_1)((\kappa\hat{P}\Gamma_s + 1)^2 - (\kappa\hat{P}\Gamma_s\xi_2)^2) \\
& (\Lambda - ((\rho P\Gamma_1\xi_1) + (1-\rho)P\Gamma_1\xi_2)^2 + (\kappa\hat{P}\Gamma_s\xi_2)^2) \\
& (-\lambda_2)\big(- ((1-2\rho)(P\Gamma_1)^2\xi_1\xi_2 - (1-\rho)(P\Gamma_1\xi_2)^2)(\Upsilon - ((\rho P\Gamma_1\xi_1)^2 + (\kappa\hat{P}\Gamma_s\xi_2)^2)) \\
& - (\rho(P\Gamma_1)^2 + P\Gamma_1(\kappa\hat{P}\Gamma_s + 1))(\Lambda - ((\rho P\Gamma_1\xi_1) + (1-\rho)P\Gamma_1\xi_2)^2 + (\kappa\hat{P}\Gamma_s\xi_2)^2) \\
& (\rho(P\Gamma_1\xi_1)^2(\Lambda - \Upsilon - (1-\rho)P\Gamma_1\xi_2)^2 - 2(\rho P\Gamma_1\xi_1)((1-\rho)P\Gamma_1\xi_2)\big)
\end{aligned} \tag{A.12}$$

Similarly, we derive (4.33) with respect to $\xi_i, \forall i = 1, 2$ and get, respectively,

$$\begin{aligned}\frac{\partial \mathcal{L}_{\mathcal{X}}}{\partial \xi_1} =&(\lambda_1 - \lambda_2)(\rho P\Gamma_1)^2\xi_1 \left[\Lambda - \left[(\rho P\Gamma_1\xi_1 + (1-\rho)P\Gamma_1\xi_2)^2 + (\kappa\hat{P}\Gamma_s\xi_2)^2\right]\right] \\ &\lambda_2(\rho P\Gamma_1)((\rho P\Gamma_1\xi_1 + (1-\rho)P\Gamma_1\xi_2))(\Upsilon - (\kappa\hat{P}\Gamma_s\xi_2)^2 - (\rho P\Gamma_1\xi_1)^2) \qquad \text{(A.13)}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \mathcal{L}_{\mathcal{X}}}{\partial \xi_2} =& -\lambda_1\xi_2(\kappa\hat{P}\Gamma_s)^2\left[\Upsilon - (\kappa\hat{P}\Gamma_s + 1)^2 - (\rho P\Gamma_1\xi_1)^2\right]\left[(\Lambda - (\rho P\Gamma_1\xi_1 + (1-\rho)P\Gamma_1\xi_2)^2\right] \\ &-\lambda_2\left[(\kappa\hat{P}\Gamma_s + 1)^2 - (\kappa\hat{P}\Gamma_s\xi_2)^2\right]\Big(\left[\xi_2(\kappa\hat{P}\Gamma_s)^2(\Lambda - \Upsilon - (1-\rho)P\Gamma_1\xi_2)^2\right. \\ &\left.-2(1-\rho)P\Gamma_1\xi_2(\rho P\Gamma_1\xi_1)\right] - ((1-\rho)P\Gamma_1)(\rho P\Gamma_1\xi_1 + (1-\rho)P\Gamma_1\xi_2) \\ &(\Upsilon - (\kappa\hat{P}\Gamma_s\xi_2)^2 - (\rho P\Gamma_1\xi_1)^2\Big). \qquad \text{(A.14)}\end{aligned}$$

A.4 MAX-MIN CASE 2: Derivation of (4.33) with respect to ξ_2

Deriving $\mathcal{L}_{\mathcal{X}}$ in (4.33) with respect to ξ_2 gives

$$\begin{aligned}&-\lambda_1\left(2\xi_2(\kappa\hat{P}\Gamma_s)^2\left(Q - K - (\rho P\Gamma_1\xi_1)^2\right)\right)\left((\hat{P}\Gamma_2)^2(1-\xi_2^2) + 2\hat{P}\Gamma_2 + 1\right) + \lambda_2\left(2\xi_2(\hat{P}\Gamma_2)^2\right) \\ &\left(Q - ((\rho P\Gamma_1\xi_1)^2 + (\kappa\hat{P}\Gamma_s\xi_2)^2)\right)\left(K - (\kappa\hat{P}\Gamma_s\xi_2)^2\right) = 0, \qquad \text{(A.15)}\end{aligned}$$

where $Q = (\rho P\Gamma_1 + \kappa\hat{P}\Gamma_s + 1)^2, K = (\kappa\hat{P}\Gamma_s + 1)^2$.

Appendix B

Proofs of Chapter 5

B.1 Proof: R_c is decreasing function of ξ_c

By deriving (5.20) with respect to ξ_c, we get

$$\frac{\partial R_c}{\partial \xi_c} = -\frac{1}{\ln 2} \frac{\xi_c (|h_1|^2 p_c)^2}{(|h_1|^2 P + \sigma^2)^2 (|h_1|^2 p_c \xi_c)^2}. \tag{B.1}$$

Since ξ_c is between zero and one, then $\frac{\partial R_c}{\partial \xi_c}$ is always negative which proves that R_c is decreasing function of ξ_c.

B.2 Finding the bound of α_k

By substituting in (5.21) into (5.19d) and solving for α_k, we find

$$2^{2R_k^{\min}} = \frac{\left(\beta^2 \alpha_o + (1-\alpha_o) \sum_{j=1}^{K} \alpha_j + \frac{\sigma^2}{P|h_k|^2}\right)^2 - (\beta^2 \alpha_o \xi_c)^2}{\left(\beta^2 \alpha_o + (1-\alpha_o) \sum_{j=1, j\neq k}^{K} \alpha_j + \frac{\sigma^2}{P|h_k|^2}\right)^2 - (\beta^2 \alpha_o \xi_c)^2}. \tag{B.2}$$

Since $\sum_{j=1}^{K} \alpha_j = 1$ and $\sum_{j=1, j\neq k}^{K} \alpha_j = 1 - \alpha_k$, after some mathematical manipulations, (B.2) results in

$$(1-\alpha_o)^2\alpha_k^2 - 2(1-\alpha_o)\left(\beta^2\alpha_o + (1-\alpha_o) + \frac{\sigma^2}{P|h_k|^2}\right)\alpha_k - F = 0, \tag{B.3}$$

where $F = (2^{-2R_k^{\min}} - 1)\left[\left(\beta^2\alpha_o + (1-\alpha_o) + \frac{\sigma^2}{P|h_k|^2}\right)^2 - (\beta^2\alpha_o\xi_c)^2\right]$. By solving the quadratic equation in (B.3) and ensuring that $\alpha_k \geq 0$, we get (5.25), which concludes the proof.

B.3 Proof: $\sum_{k=1}^{K} R_{p_k}$ is decreasing function of α_o

By deriving (5.21) with respect to α_o, i.e. $\frac{\partial \sum_{k=1}^{K} R_{p_k}}{\partial \alpha_o}$, it results in

$$\frac{\partial \sum_{k=1}^{K} R_{p_k}}{\partial \alpha_o} = \frac{1}{\ln 2}\sum_{k=1}^{K}\frac{Q}{R \times W}, \tag{B.4}$$

where

$$Q = -(\beta^2(|h_k|^2P)^2\alpha_k)(f1) - (\sigma^2(|h_k|^2P)\alpha_k)(f2) + (\xi_c^2\alpha_o\alpha_k - 2 + \alpha_k)(f3) + (\xi_c^2 - 1)(f4), \tag{B.5}$$

where $f1 = (|h_k|^2P(1-\alpha_o))^2(1-\alpha_k) + 2\sigma^2(\beta^2|h_k|^2P\alpha_o) + (\sigma^2)^4 + \sigma^2|h_k|^2P(1-\alpha_o)(2-\alpha)$.

$f2 = (|h_k|^2P(1-\alpha_o))^2(1-\alpha_k)(|h_k|^2P(1-\alpha_o))^2(1-\alpha_k) + 2\sigma^2(\beta^2|h_k|^2P\alpha_o) + (\sigma^2)^4 + \sigma^2|h_k|^2P(1-\alpha_o)(2-\alpha) + (\sigma^2)^4 + \sigma^2|h_k|^2P(1-\alpha_o)(2-\alpha) + \beta^2(|h_k|^2P)^2\alpha_o(1-\alpha_o)(2-\alpha_k)$.

$f3 = \beta^4(|h_k|^2P)^4\alpha_o(1-\alpha_o)\alpha_k$.

$f4 = \sigma^2\beta^4(|h_k|^2P)^3\alpha_o^2\alpha_k + \beta^6(|h_k|^2P)^4\alpha_o^2\alpha_k$.

$$\begin{aligned} R = {} & (\beta^2\alpha_oP|h_k|^2)^2(1-\xi_c^2) + ((1-\alpha_o)P|h_k|^2 + \sigma^2)^2 \\ & + 2((\beta^2\alpha_oP|h_k|^2)((1-\alpha_o)P|h_k|^2 + \sigma^2)). \end{aligned} \tag{B.6}$$

$$W = (\beta^2 \alpha_o P |h_k|^2)^2 (1 - \xi_c^2) + ((1-\alpha_o)(1-\alpha_k) P |h_k|^2 + \sigma^2)^2$$
$$+ 2((\beta^2 \alpha_o P |h_k|^2)((1-\alpha_o)(1-\alpha_k) P |h_k|^2 + \sigma^2)). \quad \text{(B.7)}$$

It should be noted from the expression in (B.4) that the values of R, W are positive quantities. However, the Q components results in negative quantity since $f1, f2, f3, f4$ are positive values multiplied by negative ones which results in an overall negative value. This proves that $\sum_{k=1}^{K} R_{p_k}$ is decreasing function of α_o.

B.4 Proof: $\sum_{k=1}^{K} R_{p_k}$ is increasing function of ξ_c

$$\frac{\partial \sum_{k=1}^{K} R_{p_k}}{\partial \xi_c} = \frac{1}{\ln 2} \sum_{k=1}^{K} \xi_c (\beta^2 |h_k|^2 \alpha_o P)^2 \frac{(C-D)}{C\,D} \quad \text{(B.8)}$$

where $C = R$ in (B.6), $D = W$ in (B.7), and $(C-D) = (|h_k|^2 \alpha_k (1-\alpha_o) P)^2 + 2(|h_k|^2 \alpha_k (1-\alpha_o) P)(\beta^2 |h_k|^2 \alpha_o P + |h_k|^2 (1-\alpha_k)(1-\alpha_o) P + \sigma^2)$. Since C, D and $(C-D)$ are all positive quantities, this proves that $\sum_{k=1}^{K} R_{p_k}$ is increasing function of ξ_c.

B.5 Proof of Lemma 1

Below we use contradiction to prove this Lemma. Assume that there exist s, t such that $\alpha_s^* > \alpha_s^{\min}$ and $\alpha_t^* > \alpha_t^{\min}$ for the optimal solution α_k^*. Then, we can prove that there exist feasible power allocation, $\bar{\alpha}_s, \bar{\alpha}_t$ which yields a better objective value (5.31a). To find such $\bar{\alpha}_s, \bar{\alpha}_t$, we substitute $\alpha_j = \alpha_j^*, j \neq s, t$ into problem (5.31), which after a few manipulations results in

$$\max_{\alpha_s,\alpha_t} \quad R_k(\alpha_s) + R_k(\alpha_t) \tag{B.9a}$$

$$\text{s.t.} \quad \mathcal{C}7: \ \alpha_s + \alpha_t = 1 - \sum_{j=1,j\neq s,t}^{K} \alpha_j^*, \tag{B.9b}$$

$$\mathcal{C}8: \ \alpha_s \geq \alpha_s^{\min}, \alpha_t \geq \alpha_t^{\min}, \tag{B.9c}$$

where $R_k(\alpha_s) = 1\frac{}{2\log_2\left(\frac{(\beta^2|h_s|^2\alpha_o P+|h_s|^2P(1-\alpha_o)+\sigma^2)^2-(\beta^2|h_s|^2\alpha_o P\xi_c)^2}{(\beta^2|h_s|^2\alpha_o P+|h_s|^2P(1-\alpha_o)(1-\alpha_s)+\sigma^2)^2-(\beta^2|h_s|^2\alpha_o P\xi_c)^2}\right)}$ and $R_k(\alpha_t) = \frac{1}{2}\log_2\left(\frac{(\beta^2|h_t|^2\alpha_o P+|h_t|^2P(1-\alpha_o)+\sigma^2)^2-(\beta^2|h_t|^2\alpha_o P\xi_c)^2}{(\beta^2|h_t|^2\alpha_o P+|h_t|^2P(1-\alpha_o)(1-\alpha_t)+\sigma^2)^2-(\beta^2|h_t|^2\alpha_o P\xi_c)^2}\right)$. From (B.9b), we get

$$\alpha_s = 1 - \sum_{j=1,j\neq s,t}^{K} \alpha_j^* - \alpha_t. \tag{B.10}$$

Based on (B.9c) and (B.10), we get

$$\alpha_t^{\min} \leq \alpha_t \leq 1 - \sum_{j=1,j\neq s,t}^{K} \alpha_j^* - \alpha_s^{\min}. \tag{B.11}$$

From (B.10) and (B.11), the problem in (B.9) can be re-organized as

$$\max_{\alpha_t} \quad \hat{R}_k(\alpha_t) \tag{B.12a}$$

$$\text{s.t.} \quad \mathcal{C}9: \ \alpha_t^{\min} \leq \alpha_t \leq 1 - \sum_{j=1,j\neq s,t}^{K} \alpha_j^* - \alpha_s^{\min}, \tag{B.12b}$$

where $\hat{R}_k(\alpha_t) = -\frac{1}{2}\log_2\left((\beta^2|h_s|^2\alpha_o P + |h_s|^2P(1-\alpha_o)(\sum_{j=1,j\neq s,t}^{K}\alpha_j^* + \alpha_t) + \sigma^2)^2\right)$ $-\frac{1}{2}\log_2\left((\beta^2|h_t|^2\alpha_o P + |h_t|^2P(1-\alpha_o)(1-\alpha_t) + \sigma^2)^2\right)$. Since $-\log(x)$ is a convex function, the objective function (B.12a) is convex. It is well-known that the maximization of a convex function always occurs in the boundary of the feasible solution, i.e., the optimal

solution α_t^* of problem (B.12a) is within the following region

$$\bar{\alpha}_t \in \left\{ \alpha_t^{\min}, 1 - \sum_{j=1, j\neq s,t}^{K} \alpha_j^* - \alpha_s^{\min} \right\}. \tag{B.13}$$

From (B.10), we can also get

$$\bar{\alpha}_s = 1 - \sum_{j=1, j\neq s,t}^{K} \alpha_j^* - \bar{\alpha}_t. \tag{B.14}$$

Since problem in (B.9) is equivalent to problem (B.12), $(\bar{\alpha}_s, \bar{\alpha}_t)$ is also the optimal solution of (B.9). According to (B.13) and (B.14), $\bar{\alpha}_s = \alpha_s^{\min}$ and $\bar{\alpha}_t = \alpha_t^{\min}$ is always satisfied.

Because $(\bar{\alpha}_s, \bar{\alpha}_t)$ is the optimal solution for problem in (B.9) and $(\bar{\alpha}_s, \bar{\alpha}_t) \neq (\alpha_s^*, \alpha_t^*)$, we can claim that the solution

$$(\alpha_1^*, ..., \alpha_{s-1}^*, \bar{\alpha}_s, \alpha_{s+1}^*, ..., \alpha_{t-1}^*, \bar{\alpha}_t, \alpha_{t+1}^*, ..., \alpha_K^*) \tag{B.15}$$

is feasible with the objective function value greater than solution α_k^*, which contradicts the fact that α_k^* is the optimal solution of problem (5.31). This concludes the proof of **Lemma 1.**

B.6 Proof of Convexity of (5.31a) in terms of α_k

By deriving (5.31a) with respect to α_k, we get

$$\frac{\partial \sum_{k=1}^{K} R_{p_k}}{\partial \alpha_k} = \frac{1}{\ln 2} \sum_{k=1}^{K} \frac{|h_k|^2 (1-\alpha_o) P\mathcal{A}}{\mathcal{B}}, \tag{B.16}$$

$\mathcal{A} = \beta^2|h_k|^2\alpha_o P + |h_k|^2(1-\alpha_o)P(1-\alpha_k) + \sigma^2$ and $\mathcal{B} = (\beta^2|h_k|^2\alpha_o P + |h_k|^2(1-\alpha_o)P(1-\alpha_k) + \sigma^2)^2 - (\beta^2|h_k|^2\alpha_o P\xi_c)^2$. By taking the second derivative of (B.16), we get

$$\frac{\partial^2 \sum_{k=1}^{K} R_{p_k}}{\partial \alpha_k^2} = \frac{1}{\ln 2} \sum_{k=1}^{K} \frac{(|h_k|^2(1-\alpha_o^*)P)^2(\mathcal{A}^2 + (\beta^2|h_k|^2\alpha_o P\xi_c)^2)}{\mathcal{B}^2}. \tag{B.17}$$

From (B.17), $\frac{\partial^2 \sum_{k=1}^{K} R_{p_k}}{\partial \alpha_k^2} > 0$, which proves the convexity of (5.31a) in terms of α_k.