PERFORMANCE ANALYSIS, OPTIMISATION AND ENHANCEMENT OF NON-ORTHOGONAL MULTIPLE ACCESS

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF SCIENCE AND ENGINEERING

2023

Hamad Mohammed Ali Yahya Yahya

School of Engineering
Department of Electrical and Electronic Engineering

Contents

Word Count: 48,397

Abstract

Non-orthogonal multiple access (NOMA) is a promising candidate for future wireless networks due to its improved spectral efficiency (SE), massive connectivity and low latency. This thesis focuses on studying the error rate and throughput performances of NOMA systems considering various applications including Internet of Things (IoT) and mobile communications. Exact and asymptotic bit error rate (BER)/block error rate (BLER) expressions are derived considering different fading channels such as Rayleigh, Rician and Nakagami-m. Various system parameters such as the number of users, modulation orders, number of receiving antennas, power assignments and labelling schemes are taken into account in these expressions. The derived expressions are utilised to optimise the power assignments in order to achieve specific objectives while satisfying quality of service (QoS) constraints. Moreover, the power coefficients' bounds, which ensure users' fairness, and solve the constellation ambiguity problem, are derived for two and three user cases. In addition, the power assignment achieving equally spaced constellation points is derived for an arbitrary number of users and arbitrary modulation orders. The derived BER expressions are further used in the design cognitive NOMA, adaptive modulation NOMA and automatic repeat request (ARQ)-based non-orthogonal multiplexing (NOM) systems. While the throughput is derived analytically for the first two systems, it is only evaluated using Monte-Carlo simulations for the latter. Despite NOMA's degraded BER performance compared to the orthogonal multiple access (OMA) schemes, it gains its superiority from the significant throughput performance gains at moderate and high signal to noise ratio (SNR) values, where the throughput can be as high as the number of multiplexed users/streams per communications resource. These results offer valuable insight into the system design, where the derived expressions can be utilised by the system scheduler to make informed decisions in selecting appropriate system parameters.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

I would like to express my deepest thanks to the individuals who have been instrumental in the completion of this thesis and the accomplishment of my academic milestones during the PhD.

First, I am incredibly thankful to my supervisor, Dr Emad Alsusa. Throughout this journey, he was more than just a supervisor; he was a guiding light, a mentor, and a true friend. His unwavering support, patience, and invaluable insights have played a significant role in shaping the direction and quality of this work. I am truly blessed to have had the privilege of learning from him.

I would also like to extend my sincere thanks to Dr Arafat Al-Dweik for his valuable guidance. His expertise and willingness to lend a helping hand have been crucial in navigating the challenges of this research.

Next, I want to express my gratitude to my dear friends, who provided encouragement, laughter, and companionship during both the highs and lows of this academic pursuit. Your unwavering belief in me and the many memorable moments we shared kept me motivated throughout this journey.

Lastly, I want to put my family first and acknowledge their unwavering love and support. Their sacrifices, encouragement, and understanding have been the bedrock upon which I built my dreams. Without their constant presence and belief in me, none of this would have been possible.

List of Tables

List of Figures

List of Abbreviations

F_M	feedback message.
M-PAM	M-ary pulse amplitude modulation.
M-PSK	M-ary phase shift keying.
M-QAM	M-ary quadrature amplitude modulation.
M_n-PPM	M-ary pulse position modulation.
1G	first generation.
2G	second generation.
3G	third generation.
3GPP	third generation partnership project.
4G	fourth generation.
5G	fifth generation.
6G	sixth generation.
ACK	acknowledgment.
AI	artificial intelligence.
AMC	adaptive modulation and coding.
APA	adaptive power assignment.
ARQ	automatic repeat request.
ASK	amplitude-shift keying.
aTp	average throughput.
ATSC 3.0	advanced television systems committee 3.0.
AWGN	additive white Gaussian noise.
BCH	Bose–Chaudhuri–Hocquenghem.
BER	bit error rate.
BLER	block error rate.

BPSK	binary phase-shift keying.
BS	base station.
BSC	binary symmetric channel.
CDF	cumulative distribution function.
CDMA	code division multiple access.
ChC	chase combining.
CP	cyclic prefix.
CR	cognitive radio.
CR-NOMA	cognitive-radio inspired NOMA.
CRC	cyclic redundancy check.
CSI	channel state information.
CSIR	channel state information at the receiver.
CSIT	channel state information at the transmitter.
DL	downlink.
DRB	decision regions' boundary.
DVB-T	digital video broadcasting-terrestrial.
ED	Euclidean distance.
EH	energy harvesting.
eMBB	enhanced mobile broadband.
FDMA	frequency division multiple access.
FEC	forward error correcting.
FPA	fixed power assignment.
GFA	grant free access.
GGN	generalized Gaussian noise.
GNOM	generalized non-orthogonal multiplexing.
GSM	Global System for Mobile Communications.
H-NOMA	hybrid OMA/NOMA.
HARQ	hybrid automatic repeat request.

i.i.d	mutually independent and identically distributed.
IC	interference cancellation.
IoD	Internet of Drones.
IoT	Internet of Things.
IoV	Internet of Vehicles.
IP	Internet Protocol.
IRS	intelligent reflecting surface.
iTp	instantaneous throughput.
ITU	International Telecommunication Union.
IUI	inter-user interference.
JMLD	joint-multiuser maximum likelihood detector.
LOS	line-of-sight.
LRT	likelihood-ratio test.
LTE	Long Term Evolution.
LUT	look-up table.
MGF	moment generation function.
MIMO	multiple-input multiple-output.
MISO	multiple-input single-output.
ML	maximum likelihood.
MLD	maximum likelihood detector.
mMIMO	massive multiple-input multiple-output (MIMO).
MMS	multimedia message service.
mMTC	massive machine-type communication.
mmWave	millimeter wave.
MPPM	multiple pulse position modulation.
MRC	maximum ratio combining.
MSB	most significant bit.
MTC	machine-type communication.
MUST	multiuser superposition transmission.

NACK	negative acknowledgment.
NB	narrowband.
NCO	nearest constellation point to the origin.
NLOS	non-line-of-sight.
NOM	non-orthogonal multiplexing.
NOMA	non-orthogonal multiple access.
OFDM	orthogonal frequency division multiplexing.
OMA	orthogonal multiple access.
OOK	on-off keying.
OSTBC	orthogonal space-time block code.
OTFS	orthogonal time frequency space.
PA	power assignment.
PAM	pulse amplitude modulation.
PANOMA	power adaptive non-orthogonal multiple access.
PCB	power coefficient bound.
PD	power-domain.
PDF	probability density function.
PDR	packet drop rate.
PEP	pairwise error probability.
PER	packet error rate.
PIC	parallel interference cancellation.
PSK	phase-shift keying.
PU	primary user.
QAM	quadrature amplitude modulation.
QoS	quality of service.
QPSK	quadrature phase-shift keying.
RF	radio-frequency.
RHI	residual hardware impairment.
RSMA	rate-splitting multiple access.

SC	superposition coding.
SDR	software-defined-radio.
SE	spectral efficiency.
SER	symbol error rate.
SIC	successive interference cancellation.
SIMO	single-input-multiple-output.
SINR	signal-to-interference-plus-noise ratio.
SISO	single-input-single-output.
SMS	short messages service.
SNR	signal to noise ratio.
SPC	short packet communication.
SU	secondary user.
SW	stop-and-wait.
TDD	time-division duplexing.
TDMA	time division multiple access.
THz	terahertz.
TM	transmission mode.
UAV	unmanned areal vehicle.
UL	uplink.
UMTS	Universal Mobile Telecommunication Service.
uRLLC	ultra-reliable low latency communication.
V2I	vehicle-to-infrastructure.
V2V	vehicle-to-vehicle.
VLC	visible light communication.
WiFi	wireless fidelity.
WiMAX	worldwide interoperability for microwave access.
WPAN	wireless personal area network.
WSN	wireless sensor network.

XR extended reality.

List of Mathematical Notations

$(\cdot)^H$ Hermitian transpose

$(\cdot)^T$ Transpose

$1j = \sqrt{-1}$ Imaginary unit

$\binom{\cdot}{\cdot}$ Binomial coefficient

$\binom{\cdot}{\cdot}_s$ Coefficient from the generalized Pascal's triangle with order s

$\binom{N}{N_1,\ldots,N_n}$ Multinomial coefficient

$\lceil\cdot\rceil$ Largest integer function

$\lfloor\cdot\rfloor$ Samllest integer function

$\Leftrightarrow$ Statistical equivalence

$\mathbb{B}$ Set of binary numbers

$\mathbb{C}$ Set of complex numbers

$\mathbb{E}[\cdot]$ Statistical expectation

$\mathbb{R}$ Set of real numbers

$\mathbb{Z}$ Set of integer numbers

$\mathbf{I}_a$ Identity $a \times a$ matrix

$\mathcal{CN}(0,\sigma^2)$ Complex Gaussian random variable with a zero mean and σ^2 variance

$\mathcal{N}(0,\sigma^2)$ Gaussian random variable with a zero mean and σ^2 variance

$\mathrm{Im}[\cdot]$ Imaginary component

$\mathrm{Re}[\cdot]$	Real component
$\odot$	Hardmard element-wise produce
$\Pr(\cdot)$	Probability of an event
$F(\cdot)$	Cumulative distribution function (CDF)
$f(\cdot)$	Probability density function (PDF)
$Q(\cdot)$	Complementary Gaussian function

Chapter 1

Introduction

The emergence of revolutionary wireless applications is the driver behind the need for wider connectivity to enable new breeds of advanced technologies, in addition to the increasingly demanding traditional communication devices such as mobile phones and tablets. Examples of current new technologies include holographic applications, telemedicine, massive machine-type communication (mMTC), and autonomous systems [3,4], where the Internet of Things (IoT) is at the core of these advancements [4]. These cutting-edge technologies are highly expected to become an integral part of our daily life, particularly after the COVID-19 pandemic. They are believed to be well-suited to address the challenges of living through a pandemic. Therefore, it is crucial for the industry and academia to prioritize the development of these technologies in preparation for future outbreaks and pandemics [5].

Applications illustrating the use of these technologies include conference video calls which could evolve into a more immersive scenario with holographic calls using extended reality (XR) [6]. Holographic calls are also applicable to online teaching where the classroom environment could be replicated to let the students feel as if they are in the lecture hall. In addition, the elderly vulnerable people who have regular check-ups in hospitals could use telemedicine to avoid the risk of infection during a pandemic. Similarly, remote sessions with a practitioner psychologist would require monitoring the human senses, feelings and emotions as well. Nonetheless, interdisciplinary work is needed to measure the effectiveness of remotely monitoring human biological features. Furthermore, reducing human contact in retail stores is possible by using autonomous robots. Also, intelligent autonomous vehicles could replace the need for legacy taxis and buses offering a safer and more efficient mode of transportation. While the emergence of these technologies has the potential to revolutionize

our daily lives, their wide-scale adoption requires advancing wireless communications technology and infrastructure. Consequently, proposing tailored solutions to overcome the technical challenges associated with these data-demanding use cases is crucial.

1.1 Motivations

The demand for data access and connectivity has been growing rapidly in the past few years, driven by the increasing number of new subscribers, high data rate systems, and IoT applications. According to the Ericsson mobility report [7], mobile network data traffic grew 44% between 2020 and 2021, and reached 72 Exabyte per month generated by about 8 billion subscribers. Also, in [8], it was reported that 50% of all IoT networked devices in 2023 will be connected to cellular networks. This indicates that the wireless network is expected to support about 15 billion IoT devices in addition to 8 billion mobile users. Although the advances achieved by fifth generation (5G) wireless networks are considered substantial compared to the fourth generation (4G) ones, the 5G networks' capacity improvement is still way below the $1000\times$ increase specified by the International Mobile Telecommunications-2020 vision.

Consequently, the future wireless networks are envisioned to provide ubiquitous and unlimited wireless coverage, which requires integrating space, air, ground, and underwater networks into one large multidimensional network architecture [6]. However, spectrum scarcity is one of the main challenges for realizing such ultra-wide wireless networks that support:

- Enhanced mobile broadband (eMBB) for data-demanding applications such as XR and video streaming.
- Ultra-reliable low latency communication (uRLLC) for delay-sensitive applications such as intelligent transportation.
- mMTC for massive IoT devices deployments, for instance, in smart cities.

The responsibility to satisfy such requirements lies mostly within the multiple access functionality of the communication system. Therefore, studying the access part of the communication system is crucial to align the system design with the foreseen requirements [9]. In light of such expectations, non-orthogonal multiple access (NOMA) has attracted tremendous attention as a promising candidate for future mobile networks because of its inherent ability to provide high spectral efficiency and support massive

connectivity as well as low latency. Much research in this area was focused on the integration of NOMA in various applications, such as IoT, satellite communications, unmanned areal vehicle (UAV) communications, visible light communication (VLC) and underwater communications [10–13]. Furthermore, error rate and throughput analysis of NOMA has recently attracted much attention and motivated a massive number of researchers to evaluate the error rate of various NOMA configurations and designs. Nonetheless, the exact bit error rate (BER) analysis of NOMA for future networks with arbitrary modulation orders and number of users is an open research direction that needs investigation. Such analysis is needed to enhance the throughput performance of NOMA techniques via optimal power assignment (PA), adaptive modulation and transmission mode (TM) selection to satisfy certain quality of service (QoS) requirements.

1.2 Aims and Objectives

The first aim of the thesis is to comprehensively study the error rate performance of NOMA in dense wireless networks. To achieve this aim, the following objectives are listed:

1. Study the various error rate metrics considered in the literature to evaluate the performance of NOMA.

2. Highlight the research gaps by systematically listing the main contributions of the literature.

3. Analyse the BER performance of NOMA by providing closed-form expressions for various cellular configurations, constellation labeling schemes, PA schemes, and channel models.

4. Use Monte-Carlo simulation to corroborate the derived analytical expressions for the BER.

5. Quantify the performance gains of the proposed PA and labeling schemes in different signal to noise ratio (SNR) regions.

6. Optimise the PA to minimise the system's average BER while satisfying the BER QoS constraints for individual users.

The second aim of the thesis is to study the throughput of advanced NOMA systems while considering the BER/block error rate (BLER) QoS constraints. To achieve this aim, the following objectives are listed:

1. Review the state-of-the-art literature on cognitive radio (CR)-NOMA and adaptive modulation NOMA systems.
2. Analyse the throughput performance of NOMA by providing closed-form expressions for various TM configurations and packet lengths.
3. Use Monte-Carlo simulation to corroborate the derived analytical expressions for the throughput.
4. Optimise the PA, modulation orders and other tuning parameters to maximise the system's average throughput while satisfying the BER/BLER QoS constraints for individual users.

1.3 Key Contributions

From the previously mentioned aims and objectives, the following lists the contributions per chapter.

- The key contributions of Chapter 3 can be summarised as follows:
  - Derive closed-form exact and asymptotic BER expressions for downlink (DL) NOMA with receiver diversity and an arbitrary number of users, where each user may use binary phase-shift keying (BPSK), or quadrature amplitude modulation (QAM) with square and rectangular constellations.
  - Evaluate the BER for different PAs and provide insights into the error performance of a large number of users and high modulation orders.
  - Derive closed-form expressions for the power coefficient bounds (PCBs) to solve the constellation points ambiguity problem for two and three users.
  - Evaluate the impact of changing the modulation order of certain users on the BER of other NOMA users, which is necessary for adaptive modulation and resource allocation operations.
  - Compute the optimal PA that minimize the system's average BER for two and three users while considering the PCBs as linear and non-linear constraints.

- The key contributions of Chapter 4 can be summarised as follows:
  - Propose a data-aware PA for power-domain NOMA systems, to improve the system robustness to imperfect PAs.
  - Evaluate the impact of non-optimal PA on the NOMA and power adaptive non-orthogonal multiple access (PANOMA) systems.
  - Determine the BER performance gain of the proposed system.
  - Derive and evaluate the overall lower bound capacity and quantify the performance gain of the proposed system.
  - Find the optimal PA that minimizes the system's average BER and maximizes the system's constrained capacity.
- The key contributions of Chapter 5 can be summarised as follows:
  - Design a novel joint- multiuser Gray-mapper/ demapper for NOMA with an arbitrary rectangular QAM and arbitrary number of users.
  - Derive exact instantaneous and average BER expressions over Rician fading channels considering imperfect successive interference cancellation (SIC).
  - Quantify the performance gain of the joint-multiuser Gray-mapping over the disjoint-multiuser Gray-mapping.
  - Evaluate the impact of varying the Rician factor, number layers, and modulation orders on the BER.
  - Generate feasibility maps to highlight the flexibility of joint-multiuser Gray-mapping in terms of QoS satisfaction for various layers.
- The key contributions of Chapter 6 can be summarised as follows:
  - Propose an uncoordinated downlink cognitive-radio inspired NOMA (CR-NOMA) network with hybrid underlay-interweave CR and users' BER constraints.
  - Derive closed-form expressions of the average and instantaneous local false alarm and local miss-detection probabilities for the proposed multiple access classifier at the blind cognitive receiver.
  - Derive closed-form expressions of the average and instantaneous global false alarm and global miss-detection for various counting rules to fuse the symbol-based local observations and find the global decision on a packet.

  - Derive closed-form expressions of the NOMA and orthogonal multiple access (OMA) average and instantaneous packet error rates (PERs).
  - Derive closed-form expressions of the system's throughput while considering imperfect detection at the primary user (PU) and imperfect classification and detection at the secondary user (SU).

- The key contributions of Chapter 7 can be summarised as follows:
  - Design of a NOMA system with adaptive power and QAM modulation orders.
  - Propose an efficient low-complexity algorithm to realize the modulation and power adaptation process.
  - Propose an efficient min-max optimisation to realize the modulation adaptation process with the least number of TMs.
  - Derive exact closed-form expressions for the instantaneous throughput and closed-form approximation for the average throughput.
  - Develop a novel approach to correlate the selected modulation orders and power, and thus, no side information about the power is required at the receiver side.

- The key contributions of Chapter 8 can be summarised as follows:
  - Propose a generalized non-orthogonal multiplexing (GNOM) for automatic repeat request (ARQ)-based wireless systems to improve their power and spectral efficiencies as well as the throughput.
  - Develop a greedy algorithm that computes the optimal PA to allow stacking the maximum possible number of packets in a single transmission while satisfying a predetermined PER constraint.
  - Study the impact of changing the maximum number of multiplexed packets on the system's spectral efficiency and computational complexity.
  - Evaluate the GNOM system complexity and compare it to a non-orthogonal multiplexing (NOM) based system.

1.4 Thesis Outline

The thesis is outlined as follows:

- Chapter 1 is the introduction which provides an overview of the spectrum scarcity problem and highlights the motivations, aims and objectives of the research. Also, it summarises the key contributions per chapter and provides an outline of the thesis. Finally, it lists the publications resulting from this research.

- Chapter 2 is the background chapter that highlights the contemporary challenges in wireless communications and presents a historical view of the evolution of mobile technologies. It provides an overview of digital modulations and wireless channels. It explains the fundamentals and provides a literature review of NOMA in DL and uplink (UL).

- Chapter 3 is a work chapter that derives the exact closed-form BER expressions of the conventional power-domain NOMA in Rayleigh fading channels for an arbitrary number of users with receiver diversity and square/rectangular QAM. In addition, the PCBs are presented to allow reliable detection

- Chapter 4 is a work chapter that proposes a robust data-aware PA for NOMA to ensure reliable system's BER when the access to reliable channel state information (CSI) at the base station (BS) is limited.

- Chapter 5 is a work chapter that proposes a general joint- multiuser Gray-mapper/ demapper for an arbitrary number of users and QAM orders. Also, it derives closed-form BER expressions of NOMA with joint-multiuser Gray-mapping and quantifies its gain over the disjoint-multilayer Gray-mapping.

- Chapter 6 is a work chapter that proposes an opportunistic CR-NOMA with blind TM identification and BER constraints. It derives closed-form throughput expressions considering the imperfections of the blind cognitive receiver.

- Chapter 7 is a work chapter that derives closed-form throughput expressions of NOMA with adaptive power and modulation orders under BLER constraints. It presents integer and mixed-integer programming to solve the optimisation problem where the computational complexity is analysed.

- Chapter 8 is a work chapter that proposes a GNOM design to improve the throughput of wireless ARQ systems. It presents a greedy PA algorithm to maximise the number of multiplexed packets during transmissions/re-transmissions.
- Chapter 9 is a conclusions chapter that summarises the main findings, outlines the future plans and mentions possible potential areas of research.

1.5 List of Publications

The publications resulting from the research are listed below where "P" stands for publication.

[P1] **H. Yahya**, A. Ahmed, E. Alsusa, A. Al-Dweik and Z. Ding, "Error Rate Analysis of NOMA: Principles, Survey and Future Directions," submitted for possible publication in *IEEE Open Journal of the Communications Society*, May 2023.

[P2] **H. Yahya**, E. Alsusa and A. Al-Dweik, "Exact BER Analysis of NOMA With Arbitrary Number of Users and Modulation Orders," in *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 6330-6344, September 2021 **(Chapter 3)**.

[P3] **H. Yahya**, E. Alsusa and A. Al-Dweik, "NOMA BER and BLER Performance Evaluation Under the Received Eb/N0," in *International Symposium on Networks, Computers and Communications (ISNCC)*, Doha, Qatar, November 2023, pp. 1-6 **(Chapter 3)**.

[P4] **H. Yahya**, A. Al-Dweik and E. Alsusa, "Power-Tolerant NOMA Using Data-Aware Adaptive Power Assignment for IoT Systems," in *IEEE Internet of Things Journal*, vol. 8, no. 19, pp. 14896-14907, October 2021 **(Chapter 4)**.

[P5] **H. Yahya**, E. Alsusa and A. Al-Dweik, "Enhanced Non-orthogonal Multiple Access Using Data-aware Power Assignment," in proceedings *International Symposium on Networks, Computers and Communications (ISNCC)*, Dubai, UAE, November 2021, pp. 1-6 **(Chapter 4)**.

[P6] **H. Yahya**, E. Alsusa and A. Al-Dweik, "Joint Gray-Mapping for Multilayer NOMA based Multicasting with Arbitrary Modulation Orders," submitted for possible publication in *IEEE Transactions on Green Communications and Networking*, July 2023 **(Chapter 5)**.

[P7] **H. Yahya**, E. Alsusa, A. Al-Dweik and M. Debbah, "Cognitive NOMA With Blind Transmission-Mode Identification," in *IEEE Transactions on Communications*, vol. 71, no. 4, pp. 2042-2058, April 2023 **(Chapter 6)**.

[P8] **H. Yahya**, E. Alsusa, A. Al-Dweik and M. Debbah, "Priority-Based Dynamic IoT-Downlink Communications with Blind Transmission-Mode Recognition," in *IEEE Internet of Things Magazine*, vol. 5, no. 3, pp. 106-112, September 2022 **(Chapter 6)**.

[P9] **H. Yahya**, E. Alsusa and A. Al-Dweik, "Blind Receiver Design for Downlink Cognitive-Radio NOMA Networks," in proceedings *IEEE International Conference on Communications Workshops (ICC Workshops)*, Montreal, QC, Canada, June 2021, pp. 1-6 **(Chapter 6)**.

[P10] **H. Yahya**, E. Alsusa and A. Al-Dweik, "Design and Analysis of NOMA With Adaptive Modulation and Power Under BLER Constraints," in *IEEE Transactions on Vehicular Technology*, vol. 71, no. 10, pp. 11228-11233, October 2022 **(Chapter 7)**.

[P11] **H. Yahya**, E. Alsusa and A. Al-Dweik, "Min-Max Design and Analysis of NOMA with Adaptive Modulation Under BLER Constraints," in proceedings *IEEE Vehicular Technology Conference (VTC-Fall)*, London, UK, September 2023, pp. 1-6 **(Chapter 7)**.

[P12] **H. Yahya**, A. Al-Dweik, Y. Iraqi, E. Alsusa and A. Ahmed, "A Power and Spectrum Efficient Uplink Transmission Scheme for QoS-Constrained IoT Networks," in *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17425-17439, September 2022 **(Chapter 8)**.

Chapter 2

Background

2.1 Chapter Summary

This chapter presents a historical view of the mobile network generations' evolution from the analog first generation (1G) to the recently launched fifth generation (5G), where the services provided to end users are highlighted. Furthermore, the contemporary challenges associated with the future wireless networks, especially sixth generation (6G) are discussed, where non-orthogonal multiple access (NOMA) is presented as a promising candidate. Moreover, the chapter presents a theoretical background of digital modulation schemes and wireless channel models. In addition, it explains the fundamentals of NOMA in downlink (DL) and uplink (UL).

2.2 Mobile Communications Evolution

Mobile communications networks have been developing rapidly since the 1990s with exponential growth in the achieved peak data rates [14]. This remarkable progress is due to the consistent advancement in multiple access techniques and digital modulation schemes. These advancements lead to an elevation in the services provided roughly every decade, and hence mobile networks are categorized into different generations. In the following, a historical development view of the mobile network generations is presented and the provided services capabilities are highlighted.

2.2.1 FDMA and TDMA

The deployment of 1G took place in the 1980s. Characterized by its analog nature and based on frequency division multiple access (FDMA), the radio spectrum is divided into narrowband channels [15]. Although 1G provided the essential voice services, it lacked worldwide compatibility and encryption. The commercialisation of second generation (2G) in the 1990s represented a leap forward in all aspects. For instance, introducing digital systems led to a significant reduction in the headset size compared to 1G. Even though there were few attempts to unify the standard worldwide, 2G ended up having different regional standards. In Europe for example, 2G is known as Global System for Mobile Communications (GSM), which combines time division multiple access (TDMA) and FDMA to allow multiple users' access to the same frequency channel but in different time slots. This allowed 2G to have fewer radio-frequency (RF) chains at the base station (BS) compared to 1G. Besides improved voice services, short messages service (SMS) and multimedia message service (MMS), 2G provided a peak data rate of 384 kbps [15].

2.2.2 CDMA

The release of third generation (3G) in the early 2000s gave users limited access to the internet thanks to Internet Protocol (IP) support. Known as Universal Mobile Telecommunication Service (UMTS) in Europe, the 3G's adoption of code division multiple access (CDMA) allowed users to access wide frequency bands at any time by using a unique spreading sequence. Such signatures allow the users to distinguish their signals in a pool of interfering signals. New service paradigms were supported by 3G including email services, video streaming, and Internet surfing with an achieved peak data rate of 20 Mbps.

2.2.3 OFDM

The deployment of fourth generation (4G) in 2010 revolutionized mobile networks by introducing the first standardized mobile network globally known as Long Term Evolution (LTE). This technology is based on orthogonal frequency division multiplexing (OFDM) which divides the frequency band into orthogonal subcarriers improving the spectral efficiency compared to older technologies, where the latter uses guard-bands to avoid adjacent channel interference [15]. The connection in 4G is IP oriented and packet-switched with an achieved data rate reaching 1 Gbps and a latency of 10 ms [6].

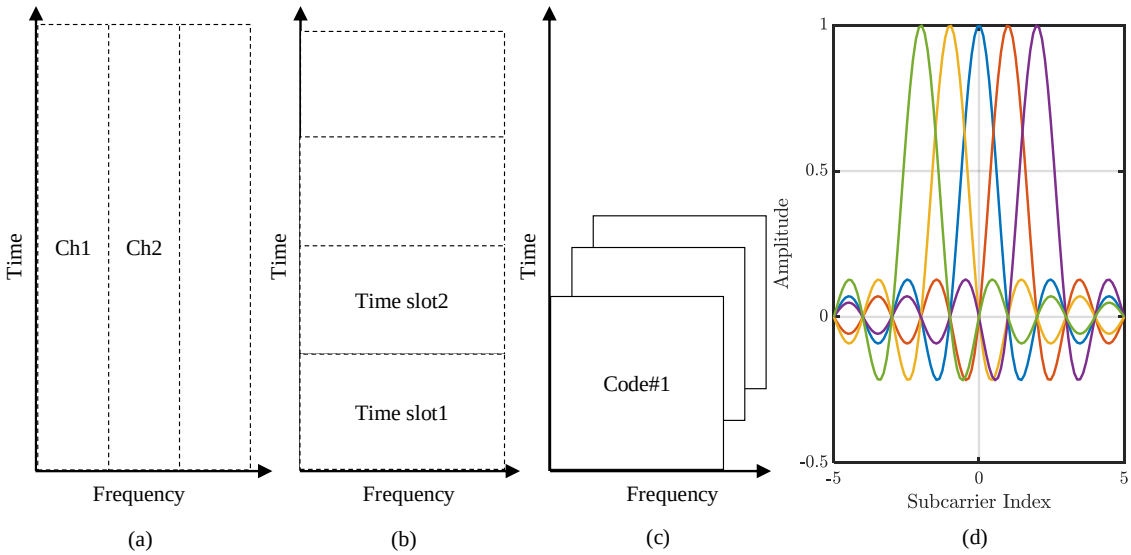


Figure 2.1: Illustration of different multiple access techniques: (a) FDMA. (b) TDMA. (c) CDMA. (d) OFDM.

In 2020, 5G was commercialised for the first time in Europe. Similar to 4G, 5G inherited OFDM technology, but it achieves 10 Gbps peak bit rate and 1 ms latency due to the larger bandwidth allocation compared to 4G. Furthermore, 5G supports different subcarrier spacing numerologies allowing flexibility to support various application scenarios. For instance, the massive machine-type communication (mMTC) is expected to provide connectivity to one million devices per square kilometre. Figure 2.1 presents the communication resources sharing strategy for the various multiple access techniques highlighted previously.

2.3 Future Wireless Systems

2.3.1 Wireless Systems Challenges

The design of future wireless systems is constrained by the challenges introduced by the increasing data demands, spectrum scarcity, and evolving technologies. In addition, the transmitted power is limited and governed by compliance regulations to ensure public safety against health complications related to using high power levels. Furthermore, mobile systems are battery-powered, and consequently, effective energy-efficient communication systems that are spectrally efficient need to be designed. Besides relying on the battery as a power primary source, energy harvesting (EH) from RF, vibration power, or solar power could be used as a secondary source. Furthermore, to overcome the spectrum scarcity and achieve high spectral efficiency, millimeter wave (mmWave)

frequency bands are used, however, the encountered losses are huge at such bands [3]. Thus, network densification is mandatory to have coverage in such cases [14]. In addition, the frequency bands safety compliance for implantable sensors and devices for instance must be ensured. Such devices are very tiny, and hence the computational complexity must be taken into consideration as the use of a fully complex communication system might not be applicable. Another solution for the coverage is having an integrated space-air-ground-underwater network that dynamically allocates the communication resources to achieve the service level agreement [4]. It is suggested that acoustic and optical communications shall be used for underwater networks [4], while unmanned areal vehicles (UAVs) could be used for air networks due to its ability to relocate [16].

2.3.2 NOMA for 6G

The 6G is envisioned to support unconventional technologies such as holographic calls which require ultra-reliability, very low latency, and enhanced mobile broadband (eMBB). Hence, it is believed that 6G should support a peak data rate of 1 Tbps, a latency of 10–100 μsec and should provide connections to ten million devices per square kilometre [6, 17]. To push 5G toward 6G, it is believed that technologies including terahertz (THz) communications, visible light communication (VLC) and blockchain, and artificial intelligence (AI) should be adopted [6]. The main limitation of 5G and the previous generations is the use orthogonal multiple access (OMA) techniques such as FDMA, TDMA, CDMA and OFDM. However, using such techniques can lead to spectrum scarcity problem when connecting a massive number of devices [18]. Therefore, NOMA is proposed as a solution to allow users to share the communication resources simultaneously, where distinct power coefficients are allocated to the users to allow interference cancellation at the receiver ends. The principle of NOMA is that the user with the highest power coefficient considers all the other users' signals to be in the noise floor. Nonetheless, the lower power coefficient users need to cancel the interference from the higher power signals and then detect their own signal [19].

2.4 Digital Modulation Schemes

$$s_n = a_n + 1jb_n \tag{2.1}$$

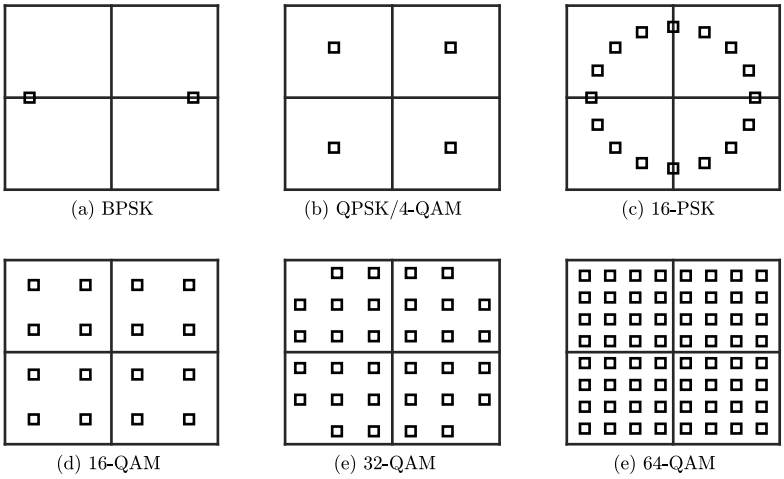


Figure 2.2: Constellation diagrams for different modulation schemes where the x-axis is the in-phase axis while the y-axis is the quadrature-phase axis.

where $1j = \sqrt{-1}$ and for a Gray mapped 4-quadrature amplitude modulation (QAM) or quadrature phase-shift keying (QPSK), $a = [1,-1,1,-1]$, $b = [1,1,-1,-1]$ and $n \in \{1,2,3,4\}$. The 4-QAM constellation points need to have an average power of 1. Hence, the normalized symbol is denoted as x such that $\mathbf{E}[|x|^2] = 1$, where $\mathbf{E}[\cdot]$ is the statistical expectation. Figure 2.2 shows the constellation diagram for different modulation schemes and different modulation orders. It can be seen that the higher the modulation order, the closer the constellation points get because of the unity average power. The demodulation process at the receiver side is said to be coherent if the receiver has knowledge about the phase of the arriving waveform such that

$$r = h\exp(-j\theta)x + w \tag{2.2}$$

where θ represents the estimated phase, h is the channel complex coefficient, x transmitted symbol and w is the additive white Gaussian noise (AWGN). When the phase is estimated accurately, the received signal can be written as

$$\check{r} = |h|x + \check{w} \tag{2.3}$$

where $\check{w}$ still has the same properties as AWGN because it is circularly symmetric. To detect the symbols with minimum possible errors, optimal detectors must be designed. The correlator which compares the received waveform with all possible waveforms is considered an optimum detector [20]. In the same context, the Euclidean distance

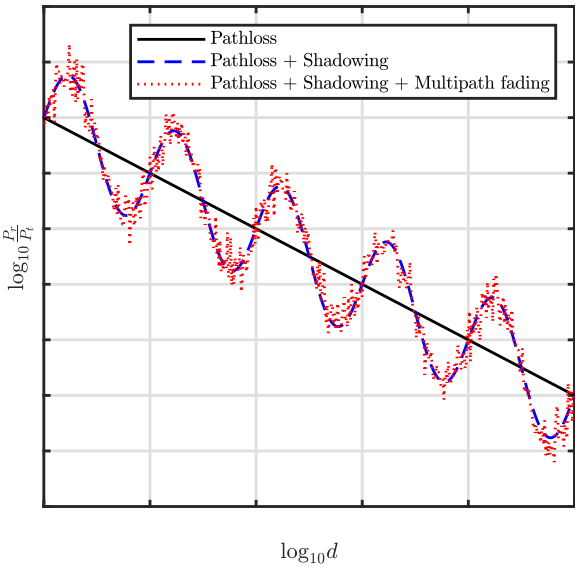


Figure 2.3: Received signal strength in a multipath fading channel.

(ED)-based detector is a maximum likelihood detector (MLD) and it is optimum as well. Studying the error rate performance is a fundamental reliability indicator of wireless communication systems is critically important, therefore, there has been a recent surge in interest in the error rate analysis of broad topics. Empirically, the bit error rate (BER) can be calculated by finding the number of detected bits in error as a fraction of the total transmitted bits. Theoretically, the BER can be derived following the probability theory and using mathematical tools such as integrals.

2.5 Wireless Channel Models

In general, the transmitted signal faces various types of impairments due to different physical phenomena. These impairments are mainly from the wireless channel and the thermal noise at the receiver. In academia, these impairments are usually modeled with statistical models but in certain scenarios measurements and empirical means are used for modeling. For instance, thermal noise is usually modeled as an AWGN. On the other hand, the wireless channel is categorized as line-of-sight (LOS) channel when a direct path between the transmitter and the receiver exists, whereas the non-line-of-sight (NLOS) channel represents the case of no direct path between the transmitter and the receiver. Hence, the received signal is a reflected or scattered version of the original signal. Furthermore, the impairments caused by the wireless channel are from: 1) Long-term shadowing. 2) Large-scale fading. 3) Small-scale fading.

Table 2.1: Multipath fading channels summary.

	$B_s < B_c$	$B_s > B_c$
$T_s > T_c$	Flat Fading Fast Fading	Frequency Selective fading Fast Fading
$T_s < T_c$	Flat Fading Slow Fading	Frequency Selective fading Slow Fading

2.5.1 Shadowing

Shadowing is defined as the deviation from the mean transmitted power. It occurs because of moving objects that might block the signal or because of environmental changes. In addition, shadowing is long-term because the environmental changes are not frequently changing. An accurate model for shadowing is log-normal shadowing which follows the log-normal distribution [21]. Moreover, shadowing is usually considered for the link budget calculation and site planning.

2.5.2 Large-Scale Fading

Large-scale fading represents the pathloss that attenuates the transmitted signal, and it is always related to the distance traveled by the signal. Some simplified models consider only the distance and the pathloss exponent. Other models consider the height of the transmitter and receiver, the gains of the antennas, and the carrier frequency. In general, the pathloss can be calculated as the ratio between the received signal and the transmitted signal, where the pathloss in the free space model is given as [21]

$$\beta = \frac{P_r}{P_t} = \left(\frac{\lambda}{4\pi d}\right)^2 \tag{2.4}$$

where P_r is the received signal power, P_t is the transmitted signal power, λ is the wavelength, d is the distance between the transmitter and the receiver, and pathloss exponent in the free space is 2. A possible simplified pathloss model is expressed as $\beta = d^{-m}$, where m is the pathloss exponent. Similarly, the large-scale fading is mainly a concern for link budget calculations and site planning.

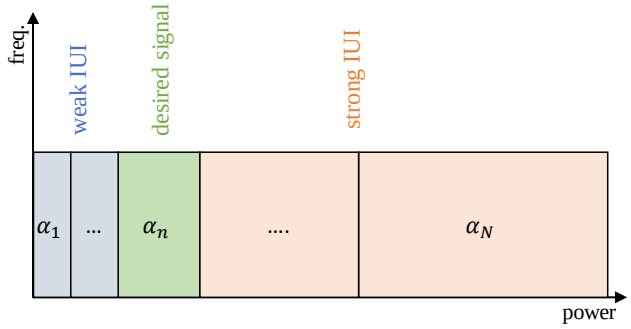


Figure 2.4: Observing DL SC at *n*th user.

2.5.3 Small-Scale Fading

The small-scale fading causes very rapid fluctuations in the received signal power. The source of the fluctuations is the multipath arrivals at the receiver which add constructively or destructively based on the headset's slight movements. Even though the slight movements might seem negligible in the human sense, they are not in the wavelength perspective. Furthermore, the small-scale fading is categorized as fast fading and slow fading based on the coherence time, T_c, which is known as the time duration in which the channel conditions remain unchanged [22]. The coherence time is directly related to the Doppler shift which is caused by mobility and related to the carrier frequency. In addition, the small-scale fading is categorized as flat fading and frequency selective fading based on the coherence bandwidth, B_c, which is related to the delay spread of the channel [22]. In general, frequency selectivity exists in ultra-wideband communication systems, whereas flat fading characterizes narrowband-like communication systems. Table 2.1 summarizes the small-scale fading channels categories where B_s is the signal bandwidth and T_s is the signal duration. It is worth noting that the flat fading LOS is modeled by Rician distribution, whereas the NLOS flat fading is modeled by Rayleigh distribution. After all, the wireless fading channel can be seen as a filter with certain coefficients where the effect of the channel can be conceived by a linear convolution of the signal with the filter coefficients. Hence, the amplitudes and the phases of the transmitted signal are altered by the channel. Figure 2.3 demonstrates how the received signal strength could vary in a multipath fading channel [21].

2.6 NOMA Principles

2.6.1 Downlink NOMA

Unlike OMA, NOMA allows the users to share the same resource block in a non-orthogonal manner. For example, the transmitter in DL-NOMA multiplexes the data of N users in the power-domain by allocating each user a distinct power coefficient, $\alpha_n, n \in \{1,\ldots,N\}$. This process is commonly known as superposition coding (SC) which can be mathematically described as $x_{sc} = \sum_{n=1}^{N} \sqrt{\alpha_n} x_n$, where $\sum_{n=1}^{N} \alpha_n = 1$ to normalize the transmitted power, x_n is the nth user information symbol, which has a unity average power, i.e. $\mathbb{E}[|x_n|^2]$. The information symbols for each user are selected from a symbol alphabet χ_n that corresponds to a certain constellation set, such as M-ary phase shift keying (M-PSK) or M-ary quadrature amplitude modulation (M-QAM). The process of allocating the power coefficients and modulation orders is performed by the scheduler at the BS, which is typically configured to ensure fairness and efficient use of resources. Figure 2.4 shows the SC process at the BS for the case of N users where $\alpha_1 < \alpha_2 \cdots < \alpha_N$. The received signal at the nth user is $y_n = h_n x_{sc} + w_n$, where h_n is the channel coefficient between the BS and nth user, and $w_n \sim \mathcal{CN}(0, N_0)$ is the AWGN. The received signal at the nth user can be written as

$$y_n = \underbrace{h_n \sum_{i=1}^{n-1} \sqrt{\alpha_i} x_i}_{\text{weak IUI}} + \underbrace{h_n \sqrt{\alpha_n} x_n}_{\text{desired signal}} + \underbrace{h_n \sum_{i=n+1}^{N} \sqrt{\alpha_i} x_i}_{\text{strong IUI}} + w_n. \tag{2.5}$$

As can be noted from (2.5), the users experience two types of interference, weak inter-user interference (IUI) and strong IUI, except for the first and Nth users whose received signals suffer only from one interference type.

The IUI is classified as strong or weak based on its power as compared to the desired signal power as shown in Figure 2.4, where the signals of users 1 to $n-1$, i.e. $s_i = \sqrt{\alpha_i} x_i$, are considered as weak IUIs to user n, while the signals of users $n+1$ to N are considered as strong IUIs to user n. Hence, s_1 to s_{n-1} are treated as unknown additive noise when detecting s_n. Moreover, the user with the maximum power, i.e., user N is the interference-limited user, and its symbols can be detected directly by all users using a single-user MLD such that

$$\hat{x}_N = \arg\min_{x_N \in \chi_N} \left| y_N - \hat{h}_N \sqrt{\alpha_N} x_N \right|^2 \tag{2.6}$$

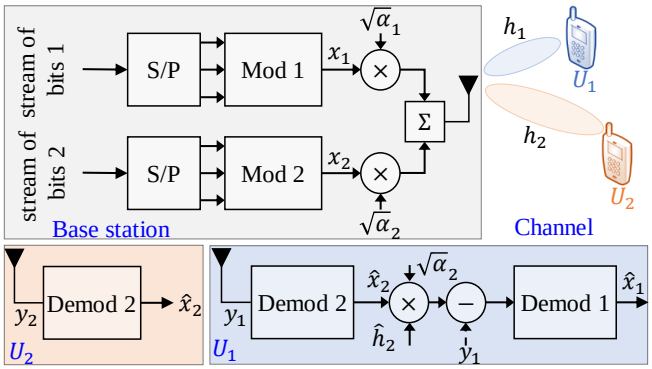


Figure 2.5: SC and SIC schematic for $N = 2$ case.

where $\hat{h}_n$ is the estimated channel coefficient, which can be written as $\hat{h}_n = h_n + \varepsilon_n$, ε_n is the channel estimation error. The MLD selects $\hat{x}_N$ as the trial value of x_N that minimizes the ED.

Furthermore, successive interference cancellation (SIC) can be used to cancel the strong IUIs where residual interference is caused when erroneous estimates of strong IUIs are used. The detection of the nth user signal can be written as

$$\widehat{x_n} = \arg\min_{x_n \in \chi_n} \left| y_n - \hat{h}_n \underbrace{\sum_{i=n+1}^{N} \sqrt{\alpha_i}\hat{x}_i}_{\text{residual interference}} - \hat{h}_n \sqrt{\alpha_n} x_n \right|^2 \tag{2.7}$$

Figure 2.5 shows a schematic of the SC process at the BS for $N = 2$ and the detection process at the users as explained in (2.6)–(3.6). It is worth mentioning that the SIC chain will get longer for $N > 2$, and hence complexity and delay will increase [23]. Alternatively, a joint-multiuser maximum likelihood detector (JMLD) can be applied to detect the signals jointly [24, Eq. (5)]. While Figure 2.5 refers to the baseband processes at the NOMA BS and receivers, interested readers can refer to [25, Figure 1] for the baseband-passband chain of processes of NOMA which is critical to conduct experimental studies.

2.6.2 Uplink NOMA

On the other hand, NOMA in the UL uses the channel gains to multiplex the different users' signals as seen in Figure 2.6. Assuming that $|h_1|^2 \alpha_1 < \ldots < |h_N|^2 \alpha_N$, the received signal at the BS for the synchronized UL-NOMA with respect to the nth user is

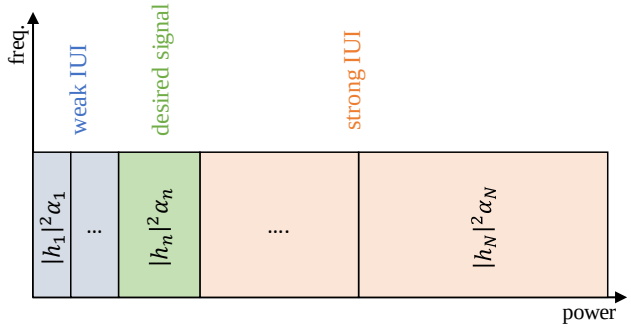


Figure 2.6: Observing UL SC at the BS.

given as

$$y = \underbrace{\sum_{i=1}^{n-1} h_i\sqrt{\alpha_i}x_i}_{\text{weak IUI}} + \underbrace{h_n\sqrt{\alpha_n}x_n}_{\text{desired signal}} + \underbrace{\sum_{i=n+1}^{N} h_i\sqrt{\alpha_i}x_i}_{\text{strong IUI}} + w. \tag{2.8}$$

Hence, the nth user received signal power at the BS depends on the $|h_n|^2\alpha_n$. Therefore, the SIC decoding order depends on the received signal power. Therefore, detecting user N signal can be performed as follows,

$$\hat{x}_N = \arg\min_{x_N \in \chi_N} \left| y - \hat{h}_N\sqrt{\alpha_N}x_N \right|^2 \tag{2.9}$$

while the detection of the nth user signal can be performed using SIC such that

$$\hat{x}_n = \arg\min_{x_n \in \chi_n} \left| y - \underbrace{\sum_{i=n+1}^{N} \hat{h}_i\sqrt{\alpha_i}\hat{x}_i}_{\text{residual interference}} - \hat{h}_n\sqrt{\alpha_n}x_n \right|^2. \tag{2.10}$$

2.6.3 Power Assignment

The power assignment (PA) scheme affects the error rate performance, and hence, it is a crucial aspect to discuss. Generally speaking, there are two main PA schemes, which are fixed PA and adaptive PA. The former ensures minimum signaling overhead and low complexity owing to its non-optimality, whereas the latter can ensure optimality or sub-optimality but with higher signaling overhead and complexity [24]. A number of issues should be considered when designing both schemes. First, ambiguity at the receivers. For example, ambiguity can be caused at the receivers for two NOMA users with binary phase-shift keying (BPSK) if equal power coefficients are given when $x_1 = -x_2$. Hence, two NOMA symbols will coincide on the same constellation point.

Moreover, enabling reliable detection using SIC requires abiding by certain power coefficient bounds (PCBs) [24, 26–29]. Considering the previous example, when $\alpha_1 < \alpha_2$ the order of the NOMA symbols is as follows from the negative in-phase axis: 11, 01, 10, 00. In such a scenario, MLD can detect b_2 reliably because $b_2 = 0$ is always in the positive side. Hence, SIC can cancel the interference caused by b_2 and detect b_1 reliably. In contrast, when $\alpha_1 > \alpha_2$, the two middle symbols swap order causing $b_2 = 0$ to fall in two different regions. Therefore, MLD will not be able to detect b_2 correctly, and hence the SIC performance will be unreliable. It is worth noting that this problem is not encountered when JMLD is used.

2.6.4 User Pairing

User pairing is the scheduler's process of selecting a set of users to share the same communication resource. In a DL/UL coordinated scenario, the BS acts as a centralized unit to coordinate the user pairing process to achieve a certain objective. For example, the objective could be the sum rate maximization while satisfying certain quality of service (QoS) constraints such as minimum rate per user or maximum BER per user. Also, user pairing can aim to ensure fairness among users. Furthermore, the satisfaction of such objectives and constraints depends on the network knowledge available at the scheduler. For instance, user pairing can satisfy certain criteria instantaneously if the channel state information (CSI) knowledge of all links is available instantaneously. Alternatively, such criteria can be satisfied on average if statistical CSI knowledge is available. For instance, it is reported that pairing users with the most distinct channel gains would maximize the sum rate in NOMA systems. The concept of prioritizing users with different priorities is known as cognitive radio (CR)-NOMA in which the primary user is given the highest priority while the secondary user is served opportunistically. Moreover, CR-NOMA tends to pair the best channel user with the second best channel user [30]. On the other hand, the UL decentralized scenario usually refers to the grant free access (GFA) in which the users are allowed random access to the channel without resource allocation. Hence, collisions might occur where the BS needs to detect collisions and resolve them in order to detect the messages.

2.7 NOMA Literature Review

The error rate performance of conventional NOMA has been widely studied in the literature, focusing mainly on DL with less attention to UL. Additionally, power allocation and modulation order optimisation based on error performance metrics has attracted the interest of many researchers. This section reviews the work that considered the DL and UL conventional NOMA, as well as its optimisation.

2.7.1 Downlink

The main work that considers DL-NOMA with a various number of users and modulation orders is given in [23, 24, 26, 29–60].

2.7.1.1 Various number of users

The error rate performance of two-user NOMA is extensively considered in the literature [26, 30–42], unlike the three-user [43–48, 58] and an arbitrary number of users cases [24, 29, 49–51]. All the aforementioned references consider single-input-single-output (SISO) channels except [24] which considers the case of multiple receiving antennas. Table 2.2 provides a comprehensive summary of the work that considered DL-NOMA. For each of the listed references, the table shows the error metric, antenna setup, channel model, modulation scheme and order M, number of users N, receiver model and CSI availability if applicable when considering fading channels. It is worth mentioning that this table structure will be adopted throughout the chapter. Furthermore, a certain superscript is adopted to distinguish the modulation schemes. For example, $\star$ is for on-off keying (OOK), $\bullet$ is for M-ary pulse amplitude modulation (M-PAM), $*$ is for M-PSK, $\diamond$ is for M-ary pulse position modulation (M_n-PPM), $\dagger$ is for multiple pulse position modulation (MPPM), and finally no superscript corresponds to M-QAM.

Two-user: Starting off with the two-user scenario, the closed-form symbol error rate (SER) expressions are derived for the imperfect SIC case in [26, 31–33]. The authors of [26, 31] consider rectangular M-QAM, whereas [33] considers hexagonal M-QAM. Meanwhile M-PAM is considered in [32]. It is worth noting that the average SER is derived for Rayleigh fading channel in [26] and Nakagami-m fading channel in [33], while the instantaneous SER is derived in [31, 32]. For the BER, Assaf *et al.* [34] derive the exact closed-form BER expressions for imperfect SIC considering rectangular M-QAM. Kara and Kaya [35] derive the BER of the near and far-users

Table 2.2: Summary of DL-NOMA work.

[#]	**Metric**	**Antennas**	**Channel**	M	N	**Receiver**	CSI
[26]	SER	SISO	Rayleigh	$\forall M$	2	Imperf. SIC	Perf.
[31]	SER	SISO	AWGN	$\forall M$	2	Imperf. SIC	-
[32]	SER	SISO	AWGN	$\forall M^{\bullet}$	2	Imperf. SIC	-
[33]	SER	SISO	Nakagami-m	$\forall M$	2	Imperf. SIC	Perf.
[34]	BER	SISO	Rayleigh	$\forall M$	2	Impref. SIC	Perf.
[35]	BER	SISO	Rayleigh	4 + 2	2	Imperf. SIC	Perf.
[36]	BER	SISO	Rayleigh	≤ 4	2	Imperf. SIC	Perf.
[37]	BER	SISO	AWGN	4	2	Imperf. SIC	Perf.
[38]	BER	SISO	Rayleigh	2	2	Perf. SIC	Imperf.
[39, 40]	BER	SISO	Rayleigh	2	2	Perf. SIC	Perf.
[41]	BER	SISO,MISO	Double Nakagami-m	$\forall M^{\bullet}$, M^{*}	2	Imperf. SIC	Perf.
[42]	PEP	2×1	Rayleigh	2, 4	2	Imperf. SIC	Perf.
[30]	PER	SISO	Nakagami-m	2	2	Imperf. SIC	Perf.
[43]	PEP	SISO	Nakagami-m	$\forall M$	2, 3	Imperf. SIC	Perf.
[44]	PEP	SISO	Rayleigh	$\forall M$	2, 3	Imperf. SIC	Perf.
[45]	PEP	SISO	Rayleigh	≤ 4	2, 3	Imperf. SIC	Perf.
[46]	PEP	SISO	Rician	2	2, 3	Imperf. SIC	Imperf.
[47]	SER	SISO	Nakagami-m	$\forall M^{\bullet}$, $\forall M$	2, 3	Imperf. SIC	Perf.
[48]	BER	SISO	Nakagami-m	4	2, 3	Imperf. SIC	Perf.
[24]	BER	SIMO	Rayleigh	$\forall M$	$\forall N$	MRC–JMLD	Perf.
[49]	BER	SISO	AWGN	4	$\forall N$	Imperf. SIC	-
[50]	BER	SISO	Rayleigh	2	$\forall N$	Perf. SIC	Perf.
[51]	PEP	SISO	Nakagami-m	$\forall M$	$\forall N$	JMLD, Imperf. SIC,	Perf.
[29]	PEP	SISO	Nakagami-m	$\forall M^{*}$ (Identical)	$\forall N$	JMLD	Perf.

which use QPSK and BPSK, respectively. Furthermore, the authors of [36] extend the work in [35] to consider the identical BPSK and QPSK cases. The authors of [37] study the high-power amplifier non-linear distortions in a DL OFDM-NOMA system, where the distortions are modeled with memory polynomial model. The exact closed-form BER expressions are derived for the two-user case with QPSK considering perfect and imperfect SIC. Chung [38–40] studies the BER performance assuming BPSK for both users and perfect SIC. For example, [38] quantifies the impact of imperfect CSI, while [39, 40] consider users with correlated information symbols, where [40] focuses on the negative correlation mapping performance. On the other hand, Jaiswal *et al.* [41] investigates a two-user vehicle-to-vehicle (V2V) communications system with NOMA. The proposed model considers SISO and multiple-input single-output (MISO) scenarios. Due to the vehicular nature of the nodes, this work considers an independent and non-identical double Nakagami-m fading channels. The source broadcasts the NOMA signal through an opportunistically selected antenna. The proposed system performance is assessed in terms of the average BER in the SISO and MISO cases. Exact expressions for both users under SISO are derived in which the infinite summation is presumed to be convergent to allow tractable analysis. Finally, the average BER is calculated by utilizing the average BER expressions of both users. A

further in-depth analysis is carried out of the system's average BER performance for the two proposed transmitting antenna selection procedures, which were verified using numerical and simulation results. Reference [42] explores the error performance of orthogonal space-time block code (OSTBC) in a DL-NOMA system. The system consists of a transmitter with two antennas and two NOMA users, each equipped with a single receiving antenna. The study evaluates the SER performance using pairwise error probability (PEP) considering imperfect SIC for QPSK-BPSK and BPSK-BPSK modulations at the near-user and far-user over Rayleigh fading channels. The authors of [30] study a two-user cognitive NOMA system, where opportunistic transmissions are considered based on the users' priorities. While assuming imperfect detection at the users' ends, closed-form packet error rate (PER) expressions are derived for BPSK case over Nakagami-m fading channels.

Two and three-user: References [43, 48] consider ordered Nakagami-m fading channels, while references [44, 45, 47] consider Rayleigh fading channels. Meanwhile reference [46] studied Rician fading channels. The authors of [43–46] derive upper bound BER expressions using the union bound of the PEPs assuming imperfect SIC. Assaf *et al.* [48] derive exact closed-form BER expressions considering QPSK for all users. The authors of [45, 46] propose integrating physical layer security with NOMA to degrade the performance of internal unknown eavesdroppers without affecting the legitimate users' performance. The scheme is based on utilizing the random phase of the channel between the users and the base station, which is known to both ends but not to illegitimate users. Hence, the scheme introduces phase shifts of multiples of $2\pi/M$ to the users' transmitted symbols based on their instantaneous channel phases before SC. Therefore, the eavesdroppers will not be able to detect the symbols, as each user symbol will appear as a different symbol in the constellation diagram. Furthermore, privacy is ensured between legitimate users. In [45], the exact closed-form union bound of the PEPs in insecure conventional NOMA is derived for the unknown eavesdroppers considering arbitrary location. Furthermore, the worst-case scenario union bound of the PEPs is derived for the proposed scheme. Reference [46] derives the exact closed-form union bound on the PEPs for the proposed scheme considering the unknown eavesdroppers to the intended users, while the impact of imperfect channel phase estimation is evaluated at one of the users. Reference [47] considers the SER analysis under equally-spaced constellations condition because only the outer constellation points are considered in the analysis. Nonetheless, the equal PA for such condition is only found for M-PAM with $M_1 = M_2 = 2$. Exact and approximate expressions

Table 2.3: Summary of the work considering various channel models for DL SISO NOMA.

[#]	**Metric**	**Channel**	M	N	**Receiver**	CSI
[52]	BER	Double Rayleigh	$\forall M^{\bullet}, \forall M^{*}$	2	Perf. SIC	Perf.
[53]	BER	κ-μ fading	4 + 2	2	Imperf. SIC	Perf.
[54]	BER	α-η-μ fading	2	2	JMLD	Perf.
[55]	BER	κ-μ shadowed fading	$\forall M$	2	Perf. SIC	Perf.
[56]	SER	Shadowed Rician	$\forall M^{*}$	2	Perf. SIC	Perf.
[57]	BER	α-μ fading	$\forall M$	2, 3	Imperf. SIC	Perf.
[23]	BER	Rician	4	2, 3	JMLD	Perf.
[58]	BER	Two-wave with diffused power	2	2, 3	Imperf. SIC	Perf.
[59]	PEP	GGN with Rayleigh	2	3	Imperf. SIC	Perf.
[60]	SER	Generalized K	2	$\forall N$	Imperf. SIC	Perf.

of the SER are presented for the two-user case considering M-PAM with $M_n \in \{2,4\}$ and M-QAM with $M_n \in \{4,16\}$, while for the three-user case the expressions are presented for M-PAM with $M_1 = 2$, $M_2 = 4$ and $M_3 = 8$ only.

Arbitrary number of users: References [24,49,50] derive exact closed-form BER expressions of an arbitrary number of users, whereas references [29, 51] considers an upper bound using PEPs. The authors of [24] consider users with rectangular M-QAM. Reference [51] derives PEP assuming M-QAM while [29] assumes users with identical M-PSK orders. Garnier *et al.* [49] consider users with QPSK. The BER is derived considering the real part only, as the in-phase and quadrature-phase components are independent in QPSK and have identical analyses. Additionally, an iterative algorithm using gradient descent is developed to find optimal power coefficients to minimize the overall system BER for two and three-user cases. The theoretical results are verified by Monte-Carlo simulations and over-the-air experimental results through a software-defined-radio (SDR) testbed. Aldababsa *et al.* [50] consider users employing BPSK and perfect SIC in Rayleigh fading channels. SDR experiments are conducted to validate the derived analytical expressions. Besides [49, 50], several articles have considered the implementation and testing of NOMA over-the-air. Interested readers can refer to Qi *et al.* [25] which provides a comprehensive review of experimental works related to NOMA. It is noted from the literature that while the analysis is generally tractable and results in closed-form solutions for the DL scenario, it becomes cumbersome and tedious for a higher number of users and higher modulation orders. Therefore, deriving a closed-form expression that is applicable to all cases is challenging.

2.7.1.2 Various channel models

Furthermore, various channel models are considered in the literature. For example, the BER of the two-user case is derived for the κ-μ fading [53], α-η-μ fading [54] and κ-μ shadowed fading [55]. In addition, the two-user SER for shadowed Rician fading is

Table 2.4: Summary of UL-NOMA work.

[#]	**Metric**	**Antennas**	**Channel**	M	N	**Receiver**	CSI
[61]	BER	SISO	AWGN	4	2	Imperf. SIC	-
[62]	BER	SISO	AWGN	4	2	Imperf. SIC	-
[63]	BER	SISO	AWGN	4 + 2	2	Imperf. SIC	-
[64]	BER	SISO	Rician	2	2	JMLD	Perf.
[65]	BER	SIMO	Rayleigh	4	2	JMLD	Perf.
[66]	PEP, SER	SISO	AWGN	4	2	Imperf. SIC	-
[67]	BER	SISO	Rayleigh	4	2	Perf. SIC	Perf.
[68]	PEP	SIMO	Rayleigh	$\forall M^*$	$\forall N$	JMLD	Perf.
[69]	PEP	SIMO	Rayleigh	$\forall M$	$\forall N$	JMLD	Perf.

derived in [56]. The BER of two and three-user scenarios is considered for α-μ fading channels in [57], for Rician fading channels in [23], and for two-wave with diffused power channel model in [58]. Bariah *et al.* [59] derived upper bound BER expressions using the union bound of PEPs while considering generalized Gaussian noise (GGN) with Rayleigh fading.

Aslan and Gucluoglu [60] derive approximate and asymptotic SER expressions over the generalized K fading channel by replacing the signal to noise ratio (SNR) of the single-user instantaneous SER with the signal-to-interference-plus-noise ratio (SINR) of multiuser NOMA. The probability density function (PDF) of the SINR is derived, but it is limited by a boundary condition. The SER analysis is validated using Monte-Carlo simulation for BPSK with an arbitrary number of users, and the results show a close match to the exact solution. However, this approach is not valid for higher modulation orders. Similarly, the authors of [52] derive the BER of two-user NOMA over double Rayleigh fading channels by integrating the BER expression over the PDF of the SINR. Table 2.3 provides a summary of various channel models used for DL-NOMA with SISO configuration.

2.7.2 Uplink

Table 2.4 provides a comprehensive summary of the conventional NOMA work in the UL direction. Synchronous UL-NOMA is studied while considering the two-user scenario [61–67] and arbitrary number of users [68, 69]. In [61–63], closed-form BER expressions are derived for SISO setup over AWGN channels, where imperfect SIC is assumed. For example, both users use QPSK in [61,62], whereas QPSK and BPSK are assigned to the near and far-users to account for channel asymmetry in [63]. Furthermore, accurate BER expressions over fading channels are derived for JMLD in [64,65], where [64] considers SISO Rician channel and BPSK, while [65] considers single-input-multiple-output (SIMO) Rayleigh fading channel and QPSK.

Table 2.5: Summary of the conventional NOMA optimisation work.

[#]	**Objective**	**Metric**	**Direction**	**Channel**	M	N	**Receiver**	CSI
[70]	Max. Throughput	BER	DL	Rayleigh	$\forall M$	2	Imperf. SIC	Perf.
[71]	Max. Throughput	BER	DL	Rayleigh	$\forall M$	2	Imperf. SIC	Perf.
[1,72]	Max. Throughput	BLER	DL	Rayleigh	$\forall M$	2	Imperf. SIC	Perf.
[73]	Max. Throughput	BER	DL	Rayleigh	$\forall M$	2	Imperf. SIC	Perf.
[74]	Max. Throughput	BER	UL	AWGN, Rayleigh	≤ 4	2	Imperf. SIC	Perf.
[75]	Min. SER	SER	DL	AWGN	$\forall M^{\bullet}, \forall M$	2	Imperf. SIC	-
[76]	Min. BER	SER, BER	DL	AWGN	4	2	Imperf. SIC	-
[77]	Min. BER	SER	DL	AWGN	$\forall M$	2	Imperf. SIC	-
[78]	Min. SER	SER	DL	Rayleigh	$\forall M$	$\forall N$	Perf. SIC	Imperf.
[79]	Min. ED	BER	UL	Rayleigh	$\forall M$	2	JMLD	Perf.
[80]	SINR balancing	BER	DL	Rayleigh	-	2	JMLD	Imperf.

Liu and Beaulieu [66] derive closed-form union bounds on the SER and BER of the two-user NOMA with QPSK for arbitrary relative phase offset in AWGN channel. These expressions are valid for both UL and DL, where intentional phase rotation can be applied for the former, and the channel can add phase rotation in the UL. Additionally, exact single-integral-from of the SER and BER expressions are derived for a specific power ratio. The authors of [67] study a multicarrier underlay CR-NOMA system in the UL. By employing perfect SIC at the receiver, analytical expressions are derived over Rayleigh fading channels for the SER per sub-carrier of both the primary and secondary users with QPSK. Furthermore, closed-form approximations for these expressions are provided.

An upper bound of the BER for an arbitrary number of UL users with M-PSK in Rayleigh fading channels is derived in [68] considering JMLD and multiple antennas at the base station. It is found that JMLD overcomes the error floor problem of the SIC. Comprehensive performance analysis of an UL-NOMA with adaptive M-QAM in the presence of Rayleigh fading channels is proposed in [69]. The system consists of N users, each equipped with a single antenna, transmitting data to a BS with multiple receiving antennas. The BS employs maximum ratio combining (MRC) to combine the received signals from all antennas and JMLD for jointly estimating the symbols of each user. The system's performance is evaluated in terms of PEP. The utilization of MRC-JMLD effectively eliminates the error floor.

2.7.3 Optimisation

To improve the performance of NOMA systems, finding the optimal transmission parameters such as the power coefficients and modulation orders is considered for different objectives and reliability constraints in [1, 70–80]. A comprehensive summary of the conventional NOMA optimisation work is shown in Table 2.5.

2.7.3.1 Maximum throughput design

For example, references [1, 70–73] study resource allocation of two-user DL-NOMA system to maximize the sum-rate using practical modulation schemes such as M-QAM while satisfying the BER/block error rate (BLER) constraints. Unlike single carrier systems [73], maximizing the sum-rate in multicarrier systems is a non-deterministic polynomial-time hard problem that requires highly complex exhaustive search [70,71]. Therefore, it is important to design efficient algorithms. Cejudo *et al.* [70] design an efficient resource allocation algorithm to maximize the sum-rate while satisfying the BER constraints. Accurate instantaneous BER expressions are presented for square M-QAM considering imperfect SIC. These expressions are used to derive the exact optimal channel gain ratios between a pair of NOMA users, maximizing the achievable sum-rate for a given BER constraint. To ensure reliable SIC detection, the power coefficient for equally-spaced constellation points or highly separable constellation points groups is presented in closed-form. The exact optimal channel gain ratios and the approximated BER expressions over Rayleigh flat fading channel are used to design an efficient iterative resource allocation algorithm which involves user pairing algorithm and continuous power and rate allocation. It is worth mentioning that when NOMA is infeasible, the subcarrier is not used. Moreover, the control channel is used to inform the users about the rate, power and subcarrier assignment.

Assaf *et al.* [71] design an efficient iterative greedy algorithm to maximize the system throughput while satisfying the users' BER constraints. Adaptive modulation, fixed power allocation and hybrid OMA/NOMA are utilized by the design. It is noteworthy to mention that relying on exact rather than approximate NOMA BER expressions induce high computational complexity. The authors of [1] design an adaptive modulation and PA scheme for packet-based NOMA to maximize the throughput. The design considers optimising the transmission mode (TM) by the BS which tunes the modulation orders and power coefficients while satisfying the BLER requirement for each user. Two PA schemes are considered: 1) Fixed PA based on linear mapping. 2) Adaptive PA. The former optimisation is solved using integer programming while the latter is solved using mixed-integer programming. In addition, the computational complexity is reduced using an efficient terminating criterion, segment-line search, and quantized signal to noise ratios SNRs. Approximated closed-form BLER expressions are derived over Rayleigh fading channels. The authors extend their work in [72] by proposing the min-max method to solve the fixed PA optimisation. Such a method reduces the complexity and the number of TMs significantly while improving

the throughput of the system.

The authors of [73] propose a joint adaptive power and modulation order to maximize the SER without sacrificing the BER performance, where the BER expressions of NOMA over Rayleigh fading are approximated as OMA counterpart. This is applicable only when assuming perfect SIC. Two adaptive power schemes are proposed to guarantee the minimum target rate for one user while maximizing the rate of the other user. Continuous and discrete adaptive rate modulation are considered to adapt to the channel fade level. Adaptive modulation is studied for UL-NOMA in [74], where an asymmetric adaptive modulation algorithm is designed for a two-user UL-NOMA system to maximize the system throughput for a given average SNRs and BER thresholds. The BER expressions are derived in closed-form for AWGN channels considering imperfect SIC and M-PSK with $M_n \in \{2,4\}$. Lower bound BER expressions are derived for Rayleigh fading channels while ignoring the channel phase effect. The boundary value effect is studied, where to have reliable SIC performance, the modulation orders are selected based on ratio of the users SNRs and the boundary value.

2.7.3.2 Minimum BER design

References [75–78] formulate several optimisation problems to minimize the error performance of DL-NOMA systems. For example, the objective function in [75] is to minimize the instantaneous SER of the strong user while satisfying the SER requirement for the weak user. The instantaneous SER is derived in closed-form for both users considering M-QAM and M-PAM. Nonetheless, a closed-form suboptimal PA is derived. The authors of [76] propose a line search algorithm for a two-user system to find the optimal power coefficients that minimize the un/coded NOMA system overall error performance in AWGN channels. BER is considered and derived for the uncoded QPSK, while the short packet communication (SPC) BER approximation is considered for the coded system. The work is extended in [77] to consider square M-QAM, where the BER is approximated using the SER expressions derived for AWGN channels. These expressions are used to find the optimal power coefficients that minimize the instantaneous overall system BER. The optimisation problem is simplified by obtaining upper and lower bounds of the overall BER.

Dutta [78] approximates the closed-form BER expressions using the SER expressions that are derived for an arbitrary number of DL-NOMA users, considering arbitrary QAM and perfect SIC. These expressions have been used for power allocation optimisation such that the average BER among all users is minimized while satisfying

certain minimum rate requirements per user. Since the problem is non-convex, an upper bound approximation is used for the objective function to make it convex, while the concave rate is transformed to be convex. As the probability of error considers all the possible realizations of the input vector, it is computationally difficult to handle the objective function. Hence, a faster approach is developed to find the optimal power coefficients, which is based on splitting the data sets into subsets and running the optimisation for all subsets in parallel. Optimising the error rate performance of the UL-NOMA system is studied in [79] which derives closed-form optimal power coefficients and phases to maximize the minimum ED of the sum of the constellations of two-user NOMA with square M-QAM, while JMLD is used at the receiver. The resulting sum constellation is found to be a uniform M-QAM constellation of a larger size. Mixed-integer optimisation is used to formulate the problem, where the complex Rayleigh fading channel is decomposed into symmetrical real and imaginary parts. Only one part is analysed due to symmetry. In a different category, optimal closed-form SINR-balancing is derived in [80] for two-user case assuming imperfect CSI. This scheme can provide fairness between the users, however, the performance is not guaranteed.

Chapter 3

Exact BER Analysis of NOMA

3.1 Chapter Introduction

Non-orthogonal multiple access (NOMA) is a promising candidate for future mobile networks as it enables improved spectral efficiency, massive connectivity, and low latency. In this chapter, the exact and asymptotic bit error rate (BER) expressions are derived under Rayleigh fading channels for NOMA systems with an arbitrary number of users and an arbitrary number of receiving antennas and modulation orders, including binary phase-shift keying and rectangular/square quadrature amplitude modulation (QAM). Furthermore, the power coefficient bounds (PCBs), which ensure users' fairness, and solve the constellation ambiguity problem, are derived for $N = 2$ and 3 users cases with any modulation orders. In addition, this work determines the optimal power assignment that minimizes the system's average BER. These results provide valuable insight into the system's BER performance and power assignment granularity. For instance, it is shown that the feasible power coefficients range becomes significantly small as the modulation order, or N, increases, where the BER performance degrades due to the increased inter-user interference (IUI). Hence, the derived expressions can be crucial for the system scheduler in allowing it to make accurate decisions of selecting appropriate N, modulation orders, and power coefficients to satisfy the users' requirements. The presented expressions are corroborated via Monte Carlo simulations.

3.1.1 Chapter Organisation

The rest of the chapter is organised as follows. In Section 3.2, related work is presented. In Section 3.3, the system and the channel models are introduced. In Section 3.4, the BER expression is derived for successive interference cancellation (SIC) and joint-multiuser maximum likelihood detector (JMLD) detectors as well as the BER for two and three users considering binary phase-shift keying (BPSK). Then in Section 3.5 the generalized BER expressions are derived for N NOMA users using arbitrary modulation orders considering single-input-single-output (SISO) and single-input-multiple-output (SIMO) systems. Section 3.6 demonstrates the analysis of the PCBs. Section 3.7 presents the analytical and Monte Carlo simulation results as well as the optimal power assignments results. Finally, Section 3.8 concludes the chapter with a summary of the main findings.

3.1.2 Notations

The notations used throughout the work are as follows. Boldface uppercase and lowercase symbols, such as $\mathbf{X}$ and such as $\mathbf{x}$, will denote matrices and row/column vectors, respectively. The transpose is denoted by $(\cdot)^T$, the Hermitian transpose is denoted by $(\cdot)^H$, and the $\odot$ denotes the Hadamard element-wise product. The real, complex, integer domains are denoted by $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{Z}$, respectively. Moreover, $\mathbb{B}$ represents the set of binary numbers. $\Pr(\cdot)$ is the probability of an event, $f(\cdot)$ is the probability density function (PDF) of a random variable, $\mathbb{E}[\cdot]$ is the statistical expectation, $\Leftrightarrow$ indicates statistical equivalence, $|\cdot|$ and $||\cdot||$ are the absolute value and the Euclidean norm, $\mathrm{Re}[\cdot]$ and $\mathrm{Im}[\cdot]$ denotes the real and imaginary components, $\binom{n}{k}$ denotes the binomial coefficients, and the imaginary number is $1j = \sqrt{-1}$. The identity $a \times a$ matrix is denoted as $\mathbf{I}_a$, and the complex Gaussian random variable with a zero mean and σ^2 variance is denoted as $\mathcal{CN}(0, \sigma^2)$.

3.2 Related Work

NOMA has attracted tremendous attention as a promising candidate for future mobile networks because of its ability to provide high spectral efficiency, massive connectivity, and low latency [19, 81–85]. Hence, much research was focused on the integration of

NOMA in various applications, such as the Internet of Things (IoT), satellite communication, unmanned areal vehicle (UAV) communications, and underwater communication [10–13, 86, 86]. For example, Perez *et al.* [86] studied NOMA for IoT networks to provide reliable secure short packet communication for downlink (DL) and uplink (UL). The work in [10, 86] investigated the application of NOMA in the forward link of multibeam satellite, whereas [13] studied its performance in underwater channels. Furthermore, a framework for UAVs serving ground users using NOMA is studied in [11], while the integration of NOMA with visible light communication (VLC) systems for indoor environments is discussed in [12].

The widely considered power-domain (PD)-NOMA, denoted as NOMA for short, is based on utilizing the power-domain to multiplex different users' signals through superposition coding (SC), where distinct power coefficients are allocated to the users before combining their signals [9]. This lends itself well to scenarios where the users require different levels of quality of service (QoS) [83]. While combining signals in this manner can improve the network capacity, the absence of orthogonality between users' signals introduces IUI which causes performance degradation to all users [87]. Therefore, BER and symbol error rate (SER) analysis of NOMA has received increased attention [12, 13] [26, 31] [34–36] [48, 50, 55, 57, 61] [63–65] [75, 76] [88–95]. For example, Cejudo *et al.* [31] attempted to approximate the BER using SER expressions for a two-user DL NOMA, where each user may use a different modulation scheme. The considered modulations schemes are BPSK, quadrature phase-shift keying (QPSK), and QAM. Nonetheless, the SER analysis is limited to additive white Gaussian noise (AWGN) channels. On the other hand, the authors of [61] derived the exact closed-form BER expressions for UL two-user NOMA with QPSK modulation over AWGN channels. They assumed that this model is perfectly synchronized while considering imperfect SIC at the base station. In [96], exact BER expressions are derived for VLC-NOMA system with an arbitrary number of users employing on-off keying (OOK).

On the other hand, the BER for a SISO Rayleigh fading wireless channel is considered in [35], where exact closed-form BER expressions are derived for the DL while approximate expressions are derived for the UL. However, these expressions are limited to a two-user NOMA considering QPSK for the near user and BPSK for the far user. The authors in [43] derived closed-form expressions for the union bound on the BER of DL NOMA with imperfect SIC over Nakagami-m fading channels. The tightness of the derived bounds varies based on various system parameters, and the gap

between the bound and exact BER may exceed 3 dB. Furthermore, analytical expressions of the pairwise error probability (PEP) are given in [51] for an arbitrary number of users and modulation orders while considering imperfect SIC over Nakagami-m fading channels. The presented analytical and simulation results show that the gap between the exact BER and PEP can be substantial, particularly for low and moderate signal to noise ratios (SNRs). Assaf *et al.* [48] derived exact BER closed-form expressions for DL NOMA over SISO Nakagami-m fading channels for the two and three-user scenarios with QPSK. In [93], closed-form BER expressions are derived for a two-user DL NOMA-VLC system while considering a limited set of modulation orders for phase-shift keying (PSK), pulse amplitude modulation (PAM), and QAM. In addition, Alqahtani and Alsusa [57] derived exact closed-form BER expressions for a two-user case employing BPSK in flat fading channels that are modeled by α-η-μ fading distribution to study the significance of different fading parameters on the BER performance.

Aldababsa *et al.* [50] presented closed-form BER expressions for an arbitrary number of users employing BPSK in Rayleigh flat fading, assuming perfect SIC. They also derived the range of proper power assignment for each user to ensure reliable BER performance. In [34], Assaf *et al.* extended the work in [48] for a two-user NOMA, where each user may use square QAM with arbitrary modulation. Additionally, proper power assignment was formulated to ensure fairness between the users and to avoid constellation points overlap. Besides the fact that the work is limited to the two-user scenario, modulation schemes such as BPSK and 8-QAM modulation orders are not considered. In [13] exact closed-form BER expressions are derived for VLC-NOMA system consisting of two users with OOK modulation in underwater environments. Analytical SER expressions for NOMA are given in [26, 31, 75, 76, 90, 91]. The authors of [26] considered the two-user case in DL NOMA using arbitrary QAM with imperfect SIC. In addition, the condition for proper power assignment is considered for the two users case. The approximated BER using SER is found to be inaccurate for high modulation orders, or at low SNR values [34]. Moreover, the authors of [91] considered a threshold detector instead of SIC. It is found that the analytical performance of the proposed detector is very close to the SIC detector. Nonetheless, the SIC detector outperforms the threshold detector at low SNRs. A comprehensive survey of work that considers BER and SER of NOMA is given in Table 3.1. In this table, SIC refers to imperfect SIC, while Approx. SIC represents perfect SIC.

Table 3.1: Error rate performance analysis survey

Metric	Year	[#]	Direction	Antennas	Channel	M	N	Receiver
SER	2017	[31]	DL	SISO	AWGN	$\forall M$,2^*	2	SIC
	2019	[26]			Rayleigh	$\forall M$		SIC
		[75]			AWGN	$\forall M^\bullet$		Approx. SIC
		[76]			AWGN	4		SIC
	2020	[90]			VLC	$\forall M$		SIC
		[91]		SIMO	Rayleigh	$\forall M$		Threshold based
BER	2017	[61]	UL	SISO	AWGN	4	2	SIC
	2019	[64]			Rician	2^*		JMLD
		[65]		SIMO	Rayleigh	4		JMLD
	2020	[63]		SISO	AWGN	$4 + 2^*$		SIC
	2017	[96]	DL		VLC	$2^\star$	$\forall N$	
	2018	[35]	DL, UL		Rayleigh	$4 + 2^*$	2	
		[92]	DL		VLC	$\leq 16^*$	2	
	2019	[48]			Nakagami-m	4	2,3	
		[93]			VLC	$\leq 16^*$, $\forall M^\bullet$	2	
		[94]		SIMO	Rayleigh	4	2	SIC
	2020	[35]		SISO	Rician	4	2,3	JMLD
		[34]			Rayleigh	$\forall M$	2	SIC
		[50]			Rayleigh	2^*	$\forall N$	Approx. SIC
		[57]			α-η-μ	2^*	2	JMLD
		[36]			Rayleigh	$\leq 4^*$		SIC
		[13]			Underwater VLC	$2^\star$		SIC
		[55]			κ-μ	$\forall M$		Approx. SIC
	Current			SIMO	Rayleigh	$\forall M$	$\forall N$	JMLD

$\star$: OOK, $*$: PSK, $\bullet$: PAM and QAM, No sign: QAM.

3.2.1 Motivations and Contributions

As can be noted from the surveyed literature, and to the best of the authors' knowledge, the existing work that considers BER analysis for DL NOMA has one or more constraints in terms of the number of users, modulation order or accuracy. However, the availability of analytical BER analysis tools is indispensable for efficient system design and optimization. Hence, this work considers the BER performance analysis of NOMA with an arbitrary number of users, where each user employs an arbitrary modulation order. The considered modulation schemes are BPSK and M-QAM with square and rectangular constellations. The obtained BER analysis is crucial for various applications such as adaptive modulation, resource allocation, user pairing, optimal power allocation and QoS requirements' satisfaction. The main contributions of this work can be summarized as follows:

1. Derive closed-form BER expressions for DL NOMA with an arbitrary number of users, where each user may use BPSK, or M-QAM with square and rectangular constellations. The analysis is applicable to NOMA systems with receiver

diversity as well.

2. Derived the asymptotic BER to simplify the BER calculation at high SNRs.
3. Evaluated the BER for different power assignments and provided insights into the error performance of a large number of users and high modulation orders.
4. Derive closed-form expressions for the PCBs to solve the constellation points ambiguity problem for $N \in \{2,3\}$, where arbitrary modulation orders are also considered.
5. Evaluated the impact of changing the modulation order of certain users on the BER of other NOMA users, which is necessary for adaptive modulation and resource allocation operations.
6. Compute the optimal power assignments that minimize the system's average BER for $N = 2$ and 3 cases while considering the PCBs as linear and non-linear constraints.

3.3 System and Channel Models

In DL NOMA, the base station multiplexes the information symbols of N users using the same radio resources by assigning each user a distinct power coefficient based on its channel conditions. Without loss of generality, we assume that N users are ordered in ascending order based on their average channel gain, i.e. $\mathbb{E}[|h_1|^2] > \mathbb{E}[|h_2|^2] > \cdots > \mathbb{E}[|h_N|^2]$, where h_n is the channel frequency response of the link between the base station and the nth user, i.e. U_n, $n \in \{1,2,\ldots,N\}$. Therefore, the power assignment is performed such that a user with severe fading conditions is assigned higher power than a user with good channel conditions [43, 48]. Consequently, the power coefficients $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N]$ are assigned such that $\alpha_1 < \alpha_2 < \cdots < \alpha_N$, where $\sum_{n=1}^{N} \alpha_n = 1$. Figure 3.1 shows an illustrative diagram of the system model for a single cell with JMLD receivers. Therefore, the NOMA symbol is described by

$$x_{SC} = \sum_{n=1}^{N} \sqrt{\alpha_n} x_n \tag{3.1}$$

where x_n is the information symbol of the nth user, which is drawn uniformly from a BPSK or M-QAM constellation χ_n. The nth user modulation order is M_n, $\mathcal{M}_n = \log_2 M_n$ and $\mathbf{m} = [M_1, M_2, \ldots, M_N]$.

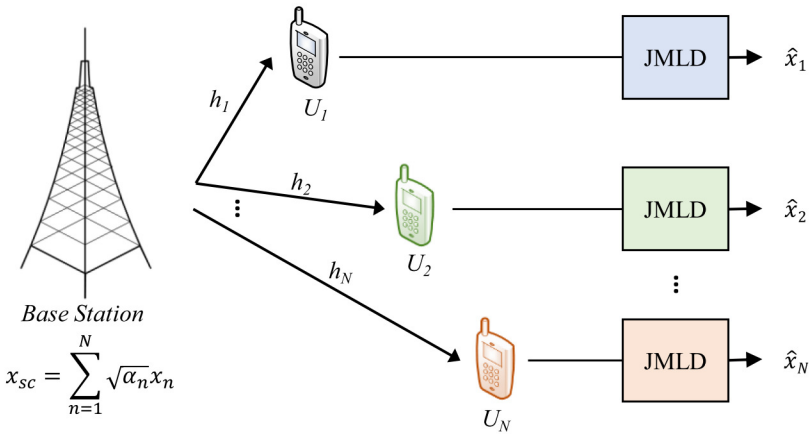


Figure 3.1: Illustrative diagram of the system model.

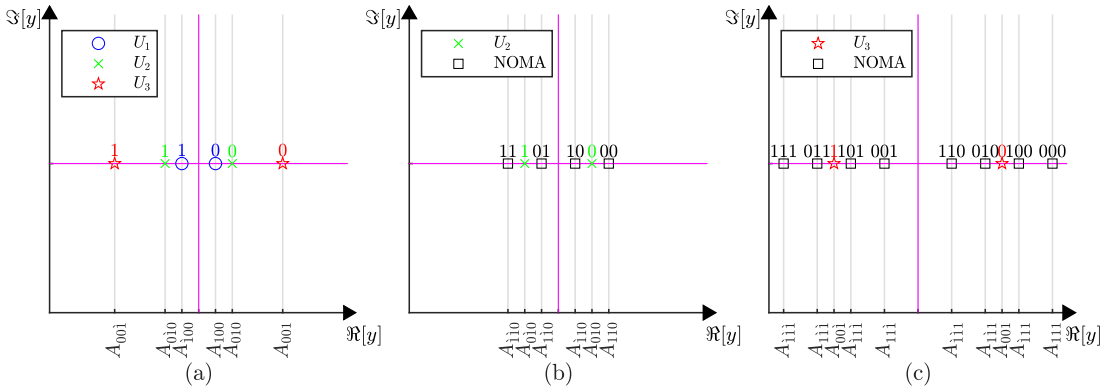


Figure 3.2: Constellation points of: (a) All users without superposition coding. (b) U_2 and the superposition coding of U_1 and U_2. (c) U_3 and the superposition coding of U_1, U_2 and U_3.

For QAM signals, the information symbols typically have $\mathbb{E}[x_n] = 0$ and $\mathbb{E}[|x_n|^2] = 1 \quad \forall n$, consequently, $\mathbb{E}[|x_{SC}|^2] = 1$. Without loss of generality, the real and imaginary components of x_{SC} are denoted by $A_{\nu_1,\nu_2,\ldots,\nu_N}$, $\nu_n \in \{0, \pm 1, \pm 3, \ldots, \pm \Lambda_n\}$, where $\Lambda_n = 2^{\lceil \log_2 \sqrt{M_n} \rceil} - 1$ and the real and imaginary components of x_{SC} are related to the individual users symbols, and for notational simplicity we define $\grave{\nu}_n \triangleq -\nu_n$. Therefore, the real and imaginary components of x_{SC} can be expressed as $A_{\nu_1,\nu_2,\ldots,\nu_N} = \sum_{n=1}^{N} \nu_n \sqrt{\frac{\alpha_n}{\kappa_n}}$ and κ_n is a scaling factor that is used to normalize the data symbols such that $\mathbb{E}[|x_n|^2] = 1$. For the special case of square QAM, $\Lambda_n \triangleq \sqrt{M_n} - 1$, and

$$\kappa_n \triangleq \frac{2}{3}(M_n - 1). \tag{3.2}$$

Figure 3.2 shows an example for $N = 3$, where $\mathbf{m} = [2, 2, 2]$. In Figure 3.2a, the constellation diagram for each user is shown separately and the x_{SC} real and imaginary components are annotated accordingly. It is worth noting that the constellation for

each user is presented after scaling with its respective power coefficient α_n. Figure 3.2b shows the resultant constellation after the superposition of U_1 and U_2 symbols. In Figure 3.2c, the overall NOMA constellation is presented showing the real and imaginary components for each constellation point. There is no imaginary component of the NOMA symbol as all users use BPSK. On the other hand, the maximum real component is $A_{111} = \sqrt{\alpha_1} + \sqrt{\alpha_2} + \sqrt{\alpha_3}$. The other real and imaginary components can be found by considering the different combinations of $\nu_1, \nu_2, \ldots, \nu_N$.

The bit-to-symbol mapping considered in this work follows the widely used model, where only the individual user bit mapping is based on Gray mapping as shown in Figure 3.2a. Therefore, the NOMA constellation will not be Gray mapped. The constellation diagrams in Figure 3.2 are labeled using binary numbers that represent the symbol values in bits. The NOMA bit-word is denoted by $\mathbf{b} = [b_1, b_2, \ldots, b_q]$, where $q = \sum_{i=1}^{N} \mathcal{M}_i$ and b_1 is the most significant bit (MSB). The nth individual user bits can be expressed as $\mathbf{b}_n = [b_{O_n}, b_{O_n+1}, \ldots, b_{O_n+\mathcal{M}_n-1}]$, where b_{O_n} is user's MSB, and

$$O_n = \begin{cases} 1, & n = 1 \\ 1 + \sum_{i=1}^{n-1} \mathcal{M}_{n-i}, & n > 1 \end{cases} . \tag{3.3}$$

For the example in Figure 3.2c, $\mathbf{b}_1 = [b_1]$ belongs to U_1, $\mathbf{b}_2 = [b_2]$ belongs to U_2, and $\mathbf{b}_3 = [b_3]$ belongs to U_3. It is worth noting that using nonlinear mapping for the NOMA constellation may provide some error rate performance improvement, however, the gained improvement is generally small and increases the receiver complexity [97, 98]. The BER analysis for Gray mapped NOMA generally follows the same approach used for linear mapping. At the receiver side, the received baseband signal in flat fading channels is written as

$$y_n = h_n x_{SC} + w_n \tag{3.4}$$

where $w_n \sim \mathcal{CN}(0, N_0)$ is the AWGN, and $\mathrm{Re}[w_n] \Leftrightarrow \mathrm{Im}[w_n] \triangleq \tilde{w}_n \sim \mathcal{N}(0, N_0/2)$, while the noise spectral density is given as $\frac{N_0}{2} = \sigma^2_{w_n}$. In channels with small scale Rayleigh fading and large scale pathloss, the channel gain can be decomposed as $h_n = \sqrt{\beta_n} \times \hbar_n$, where $\hbar_n \sim \mathcal{CN}(0, \sigma^2_{\hbar_n})$, $\beta_n = \Upsilon_n^{-\lambda}$, Υ_n is the distance between the base station and U_n, and λ is the pathloss exponent. The coefficients $\hbar_1, \hbar_2, \ldots, \hbar_N$ are mutually independent and identically distributed (i.i.d) random variables.

The most common schemes for multi-user detection of NOMA signals are the SIC and the JMLD detectors [34, 35]. The main difference between SIC and JMLD is that the former attempts to cancel the interference of other users, while the latter detects the

users' signals jointly without interference cancellation. Nevertheless, Assaf *et al.* [23] proved that the BER performance of SIC and JMLD is identical for the DL NOMA under perfect knowledge of channel state information (CSI).

Considering SIC, user N symbols can be detected directly using a single-user maximum likelihood detector (MLD) such that

$$\hat{x}_N = \arg\min_{x_N \in \chi_N} |y_N - h_N \sqrt{\alpha_N} x_N|^2 \tag{3.5}$$

where x_i represents the trail symbols for the ith user taken from the symbol alphabet χ_i.

Furthermore, the strong IUIs will be cancelled for the nth user signal which can be detected as

$$\widehat{x_n} = \arg\min_{x_n \in \chi_n} \left| y_n - h_n \underbrace{\sum_{i=n+1}^{N} \sqrt{\alpha_i} \hat{x}_i}_{\text{residual interference}} - h_n \sqrt{\alpha_n} x_n \right|^2 \tag{3.6}$$

On the other hand, JMLD can recover the information symbols of all users jointly such that

$$\{\widehat{x_1}, \widehat{x_2}, ..., \widehat{x_N}\} = \arg\min_{x_i \in \chi_i} \left| y_n - h_n \sum_{i=1}^{N} \sqrt{\alpha_i} x_i \right|^2 \tag{3.7}$$

where $\{\widehat{x_1}, \widehat{x_2}, ..., \widehat{x_N}\}$ are the jointly detected N users' symbols, and x_i represents the trail symbols for the ith user taken from the symbol alphabet χ_i. In this work, we consider JMLD to enable a compact systematic analysis.

3.4 BER Analysis: Different Examples

The BER can be obtained by considering all the possible NOMA constellation points transmission. For equally probable symbols, the conditional BER can be found by considering the rightmost symbols due to symmetry. Hence the conditional BER can be written as

$$P_{B_n}|\gamma_n = \frac{1}{2^{N-1}} \sum_{i=1}^{2^{q-2}} P_{B_n}^{(i)} \tag{3.8}$$

where γ_n is the instantaneous which is defined as $\gamma_n \triangleq \frac{2|h_n|^2}{N_0}$ and $q = \sum_{i=1}^{N} \mathcal{M}_i$. To prove that JMLD has identical BER performance to SIC under perfect CSI assumption, the performance of both detectors needs to be derived [34, 48].

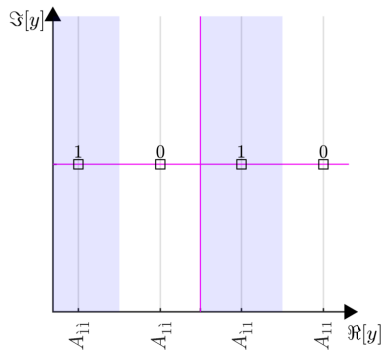


Figure 3.3: NOMA constellation points for $N = 2$, $\mathbf{m} = [2,2]$ where the decision regions boundaries for U_1 are shown.

3.4.1 JMLD and SIC Analysis

The case of $N = 2$ is considered where BPSK is used by both users. It is noted that the number of possible NOMA constellation points is 4 which is proportional to the number of users N where the total number of NOMA constellation points is given as 2^N. Moreover, the near user, U_1 could use SIC or JMLD to detect its own signal whereas the far user, U_2, detects it signal directly using hard decision considering U_1 signal to be in the noise floor. The hard decision is represented by the MLD which can be implemented using (3.7). Hence, the analysis of U_2 is not in the scope of this subsection.

3.4.1.1 JMLD analysis

Case 1: The transmitted NOMA word is, $\mathcal{X} = 00$, where the amplitude in this case is A_{11}. The transmitted symbol is detected erroneously if the AWGN level shifts the transmitted symbol to one of the decision regions where the bit value is flipped. The NOMA constellation points and the decision regions can be seen in Figure 3.3. Therefore,

$$P_{B_1}^{(1)} = \Pr\left(\tilde{w} < -\left\{A_{11} - \frac{A_{11}+A_{\grave{1}1}}{2}\right\}\right) - \Pr\left(-\left\{A_{11} + \frac{A_{11}+A_{\grave{1}1}}{2}\right\} < \tilde{w} < -A_{11}\right) \tag{3.9}$$

where the user index is dropped from $\tilde{w}$ for notational simplicity. Thus,

$$\begin{aligned} P_{B_1}^{(1)} &= \frac{1}{\sqrt{2\pi\sigma_w^2}} \int_{-\infty}^{-\frac{A_{11}+A_{\grave{1}1}}{2}} \exp\left(\frac{-\tilde{w}^2}{2\sigma_w^2}\right) d\tilde{w} - \frac{1}{\sqrt{2\pi\sigma_w^2}} \int_{-\frac{3A_{11}-A_{\grave{1}1}}{2}}^{-A_{11}} \exp\left(\frac{-\tilde{w}^2}{2\sigma_w^2}\right) d\tilde{w} \\ &= Q\left(\sqrt{\mathcal{A}_{1,1}}\right) + Q\left(\sqrt{\mathcal{A}_{2,1}}\right) - Q\left(\sqrt{\mathcal{A}_{4,1}}\right) \end{aligned} \tag{3.10}$$

where $Q(.)$ is the Gaussian function, and the definitions of $\mathcal{A}_{l,n}$ for $N = 2$ case can be found in Table A.1 in the Appendix.

Case 2: The transmitted NOMA word is, $\mathcal{X} = 10$, where the amplitude is $A_{\grave{1}1}$ and the error event can be computed as the sum of two events, i.e., $P_{B_1}^{(2)} = P_{B_1}^{(2a)} + P_{B_1}^{(2b)}$, where

$$P_{B_1}^{(2a)} = \Pr\left(\tilde{w} > \left\{\frac{A_{11} + A_{\grave{1}1}}{2} - A_{\grave{1}1}\right\}\right) = Q\left(\sqrt{\mathcal{A}_{1,1}}\right) \tag{3.11}$$

and

$$\begin{aligned} P_{B_1}^{(2b)} &= \Pr(\tilde{w} < -A_{\grave{1}1}) - \Pr\left(\tilde{w} < -\left\{A_{\grave{1}1} + \frac{A_{11} + A_{\grave{1}1}}{2}\right\}\right) \\ &= Q\left(\sqrt{\mathcal{A}_{3,1}}\right) - Q\left(\sqrt{\mathcal{A}_{5,1}}\right). \end{aligned} \tag{3.12}$$

Thus,

$$P_{B_1}^{(2)} = Q\left(\sqrt{\mathcal{A}_{1,1}}\right) + Q\left(\sqrt{\mathcal{A}_{3,1}}\right) - Q\left(\sqrt{\mathcal{A}_{5,1}}\right). \tag{3.13}$$

For the remaining cases, ***Cases 3*** and ***4***, the BER is equal in ***Cases 2*** and ***1***, respectively. Therefore, the conditional BER of U_1 is given by

$$P_{B_1}|\gamma_1 = \sum_{i=1}^{5} c_i Q\left(\sqrt{\mathcal{A}_{i,1}}\right) \tag{3.14}$$

where $\mathbf{c} = \frac{1}{2}[2, 1, 1, -1, -1]$.

3.4.1.2 SIC analysis

There are two assumptions in the literature for SIC process, one assumes that SIC is perfect and it cancels the interference perfectly. The other assumption is the imperfect SIC which is based on the fact the AWGN and the wireless channel could result in wrong interference cancellation at the receiver. Here the second assumption is assumed where to get the BER, the case of correct SIC and incorrect SIC most be considered.

Case 1: The transmitted amplitude in this case is A_{11}. By splitting this case to two sub-cases, correct SIC and incorrect SIC, the error occurs in two ways. First, when

SIC is correct, the detected amplitude is A_{10} and the probability of error is

$$\begin{aligned} P_{B_1}^{(1a)} &= \Pr\left(\tilde{w} < -A_{10} \cap \tilde{w} > -A_{11}\right) \\ &= \Pr\left(-A_{11} < \tilde{w} < -A_{10}\right) \\ &= Q\left(\sqrt{\mathcal{A}_{1,1}}\right) - Q\left(\sqrt{\mathcal{A}_{4,1}}\right) \end{aligned} \tag{3.15}$$

whereas when SIC is incorrect, the detected amplitude is A_{12} and the probability of error can be written as

$$P_{B_1}^{(1b)} = \Pr\left(\tilde{w} < -A_{12} \cap \tilde{w} < -A_{11}\right) = \Pr\left(\tilde{w} < -A_{12}\right) = Q\left(\sqrt{\mathcal{A}_{2,1}}\right) \tag{3.16}$$

Hence, the BER for this can be written as

$$P_{B_1}^{(1)} = Q\left(\sqrt{\mathcal{A}_{1,1}}\right) + Q\left(\sqrt{\mathcal{A}_{2,1}}\right) - Q\left(\sqrt{\mathcal{A}_{4,1}}\right) \tag{3.17}$$

Case 2: The transmitted amplitude in this case is $A_{\grave{1}1}$. By splitting this case to two sub-cases, correct SIC and incorrect SIC, the error occurs in two ways. First, when SIC is correct, the detected amplitude is $A_{\grave{1}0}$ and the probability of error is

$$P_{B_1}^{(2a)} = \Pr\left(\tilde{w} > A_{10} \cap \tilde{w} > -A_{\grave{1}1}\right) = \Pr\left(\tilde{w} > A_{10}\right) = Q\left(\sqrt{\mathcal{A}_{1,1}}\right) \tag{3.18}$$

whereas when SIC is incorrect, the detected amplitude is $A_{\grave{1}2}$ and the probability of error can be written as

$$\begin{aligned} P_{B_1}^{(2b)} &= \Pr\left(\tilde{w} > -A_{\grave{1}2} \cap \tilde{w} < -A_{\grave{1}1}\right) \\ &= \Pr\left(A_{\grave{1}1} < \tilde{w} < A_{\grave{1}2}\right) \\ &= Q\left(\sqrt{\mathcal{A}_{3,1}}\right) - Q\left(\sqrt{\mathcal{A}_{5,1}}\right) \end{aligned} \tag{3.19}$$

Hence, the BER for this can be written as

$$P_{B_1}^{(2)} = Q\left(\sqrt{\mathcal{A}_{1,1}}\right) + Q\left(\sqrt{\mathcal{A}_{3,1}}\right) - Q\left(\sqrt{\mathcal{A}_{5,1}}\right) \tag{3.20}$$

By averaging (3.17) and (3.20) the conditional BER for SIC can be found. It can be seen that the SIC performance is identical to JMLD AWGN channel (3.14). In addition, the performances are identical in fading channels if the CSI is perfectly known at the receiver.

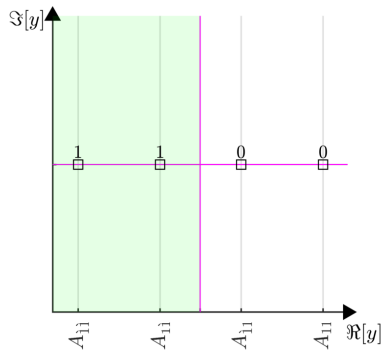


Figure 3.4: NOMA constellation points for $N = 2$, $\mathbf{m} = [2,2]$ where the decision region boundary for U_2 are shown.

3.4.2 Conditional BER Analysis: Two Users Case

In this subsection, the analysis of the conditional BER for two users case using BPSK will be shown based on the JMLD detector. Since the analysis of U_1 was shown in the previous subsection, it is omitted here and the analysis of U_2 is only shown. The only decision region boundary for U_2 is shown in Figure 3.4.

Case 1: The same transmitted amplitude as in U_1. However, its decision boundary is zero. Therefore, the conditional BER is

$$P_{B_2}^{(1)} = \Pr\left(\tilde{w} < -A_{11}\right) = Q\left(\sqrt{\mathcal{A}_{1,2}}\right). \tag{3.21}$$

Case 2: The transmitted amplitude is $A_{\grave{1}1}$. Thus, the conditional BER is

$$P_{B_2}^{(2)} = \Pr(\tilde{w} < -A_{\grave{1}1}) = Q\left(\sqrt{\mathcal{A}_{2,2}}\right). \tag{3.22}$$

The remaining cases, ***Cases 3*** and ***4***, have identical BER to their peers, ***Cases 2*** and ***1***, respectively. Therefore, it sufficient to only consider the positive side. Consequently, the conditional BER is

$$P_{B_2}|\gamma_2 = \sum_{i=1}^{2} c_i Q\left(\sqrt{\mathcal{A}_{i,2}}\right) \tag{3.23}$$

where $\mathbf{c} = \frac{1}{2}[1,1]$.

3.4.3 Conditional BER Analysis: Three Users Case

In this subsection, the analysis of the conditional BER for three users case using BPSK will be demonstrated for the JMLD detector. The decision regions boundaries for the three users are shown in Figure 3.5. Similar to the previous subsection, however, the

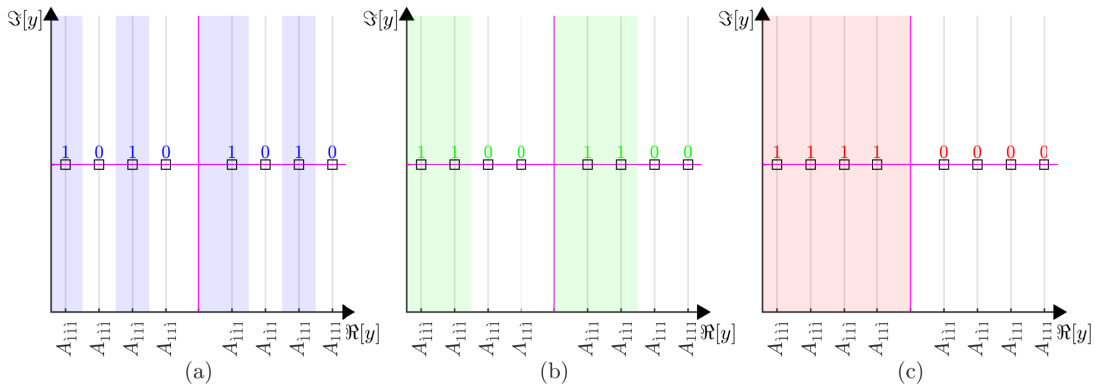


Figure 3.5: NOMA constellation points for $N = 3$ with the decision regions boundaries of: (a) U_1. (b) U_2. (c) U_3.

number of users is 3. Hence, the amplitude of the transmitted symbol is determined by the three power coefficients, α_1, α_2 and α_3. The same strategy to get the BER will be followed here as well.

3.4.3.1 Near user analysis

Case 1: The transmitted NOMA word is, $\mathcal{X} = 000$, where the amplitude in this case is A_{111}. The BER of this case is shown in (3.24). Note that the definitions of $\mathcal{A}_{l,n}$ for $N = 3$ case can be found in Table A.2 in the Appendix.

$$\begin{aligned}
P_{B_1}^{(1)} &= \Pr\left(\tilde{w} < -\left\{A_{111} - \frac{A_{111} + A_{\grave{1}11}}{2}\right\}\right) \\
&- \Pr\left(-\left\{A_{111} - \frac{A_{1\grave{1}1} + A_{\grave{1}\grave{1}1}}{2}\right\} < \tilde{w} < -\left\{A_{111} - \frac{A_{\grave{1}11} + A_{1\grave{1}1}}{2}\right\}\right) \\
&- \Pr\left(-\left\{A_{111} + \frac{A_{1\grave{1}1} + A_{\grave{1}\grave{1}1}}{2}\right\} < \tilde{w} < -A_{111}\right) \\
&- \Pr\left(-\left\{A_{111} + \frac{A_{111} + A_{\grave{1}11}}{2}\right\} < \tilde{w} < -\left\{A_{111} + \frac{A_{\grave{1}11} + A_{1\grave{1}1}}{2}\right\}\right) \\
&= Q\left(\sqrt{\mathcal{A}_{1,1}}\right) + Q\left(\sqrt{\mathcal{A}_{2,1}}\right) + Q\left(\sqrt{\mathcal{A}_{3,1}}\right) + Q\left(\sqrt{\mathcal{A}_{4,1}}\right) \\
&- Q\left(\sqrt{\mathcal{A}_{14,1}}\right) - Q\left(\sqrt{\mathcal{A}_{15,1}}\right) - Q\left(\sqrt{\mathcal{A}_{16,1}}\right). \qquad (3.24)
\end{aligned}$$

Case 2: The transmitted NOMA word is, $\mathcal{X} = 100$, where the amplitude in this case is $A_{\grave{1}11}$. The error can be computed as the sum of two events, i.e., $P_{B_1}^{(2)} = P_{B_1}^{(2a)} + P_{B_1}^{(2b)}$, where

$$P_{B_1}^{(2a)} = \Pr\left(\tilde{w} > \left\{\frac{A_{111} + A_{\grave{1}11}}{2} - A_{\grave{1}11}\right\}\right) = Q\left(\sqrt{\mathcal{A}_{1,1}}\right) \qquad (3.25)$$

and

$$\begin{aligned}
P_{B_1}^{(2b)} &= \Pr\left(\tilde{w} < -\left\{A_{\grave{1}11} - \frac{A_{\grave{1}11}+A_{1\grave{1}1}}{2}\right\}\right) \\
&- \Pr\left(-A_{\grave{1}11} < \tilde{w} < -\left\{A_{\grave{1}11} - \frac{A_{1\grave{1}1}+A_{\grave{1}\grave{1}1}}{2}\right\}\right) \\
&- \Pr\left(-\left\{A_{\grave{1}11} + \frac{A_{\grave{1}11}+A_{1\grave{1}1}}{2}\right\} < \tilde{w} < -\left\{A_{\grave{1}11} + \frac{A_{1\grave{1}1}+A_{\grave{1}\grave{1}1}}{2}\right\}\right) \\
&- \Pr\left(\tilde{w} < -\left\{A_{\grave{1}11} + \frac{A_{111}+A_{\grave{1}11}}{2}\right\}\right) \\
&= Q\left(\sqrt{\mathcal{A}_{5,1}}\right) + Q\left(\sqrt{\mathcal{A}_{6,1}}\right) + Q\left(\sqrt{\mathcal{A}_{7,1}}\right) \\
&- Q\left(\sqrt{\mathcal{A}_{17,1}}\right) - Q\left(\sqrt{\mathcal{A}_{18,1}}\right) - Q\left(\sqrt{\mathcal{A}_{19,1}}\right). \qquad (3.26)
\end{aligned}$$

Thus,

$$\begin{aligned}
P_{B_1}^{(2)} &= Q\left(\sqrt{\mathcal{A}_{1,1}}\right) + Q\left(\sqrt{\mathcal{A}_{5,1}}\right) + Q\left(\sqrt{\mathcal{A}_{6,1}}\right) + Q\left(\sqrt{\mathcal{A}_{7,1}}\right) \\
&- Q\left(\sqrt{\mathcal{A}_{17,1}}\right) - Q\left(\sqrt{\mathcal{A}_{18,1}}\right) - Q\left(\sqrt{\mathcal{A}_{19,1}}\right). \qquad (3.27)
\end{aligned}$$

Case 3: The transmitted NOMA word is, $\mathcal{X} = 010$, where the amplitude is $A_{1\grave{1}1}$. Similarly, the error can be computed as the sum of two events, i.e., $P_{B_1}^{(3)} = P_{B_1}^{(3a)} + P_{B_1}^{(3b)}$, where

$$\begin{aligned}
P_{B_1}^{(3a)} &= \Pr\left(\tilde{w} > \left\{\frac{A_{\grave{1}11}+A_{1\grave{1}1}}{2} - A_{1\grave{1}1}\right\}\right) - \Pr\left(\tilde{w} > \left\{\frac{A_{111}+A_{\grave{1}11}}{2} - A_{1\grave{1}1}\right\}\right) \\
&= Q\left(\sqrt{\mathcal{A}_{5,1}}\right) - Q\left(\sqrt{\mathcal{A}_{20,1}}\right) \qquad (3.28)
\end{aligned}$$

and

$$\begin{aligned}
P_{B_1}^{(3b)} &= \Pr\left(\tilde{w} < -\left\{A_{1\grave{1}1} - \frac{A_{1\grave{1}1}+A_{\grave{1}\grave{1}1}}{2}\right\}\right) \\
&- \Pr\left(-\left\{A_{1\grave{1}1} + \frac{A_{1\grave{1}1}+A_{\grave{1}\grave{1}1}}{2}\right\} < \tilde{w} < -A_{1\grave{1}1}\right) \\
&- \Pr\left(-\left\{A_{1\grave{1}1} + \frac{A_{111}+A_{\grave{1}11}}{2}\right\} < \tilde{w} < -\left\{A_{1\grave{1}1} + \frac{A_{\grave{1}11}+A_{1\grave{1}1}}{2}\right\}\right) \\
&= Q\left(\sqrt{\mathcal{A}_{8,1}}\right) + Q\left(\sqrt{\mathcal{A}_{9,1}}\right) + Q\left(\sqrt{\mathcal{A}_{10,1}}\right) - Q\left(\sqrt{\mathcal{A}_{21,1}}\right) - Q\left(\sqrt{\mathcal{A}_{22,1}}\right). \qquad (3.29)
\end{aligned}$$

Thus,

$$P_{B_1}^{(3)} = Q\left(\sqrt{\mathcal{A}_{5,1}}\right) + Q\left(\sqrt{\mathcal{A}_{8,1}}\right) + Q\left(\sqrt{\mathcal{A}_{9,1}}\right) + Q\left(\sqrt{\mathcal{A}_{10,1}}\right) - Q\left(\sqrt{\mathcal{A}_{20,1}}\right) - Q\left(\sqrt{\mathcal{A}_{21,1}}\right) - Q\left(\sqrt{\mathcal{A}_{22,1}}\right). \tag{3.30}$$

Case 4: The transmitted NOMA word is, $\mathcal{X} = 110$, where the amplitude is $A_{\grave{1}\grave{1}1}$. The BER can be computed as the sum of two events, i.e., $P_{B_1}^{(4)} = P_{B_1}^{(4a)} + P_{B_1}^{(4b)}$, where

$$\begin{aligned} P_{B_1}^{(4a)} &= \Pr\left(\tilde{w} > \left\{\frac{A_{1\grave{1}1} + A_{\grave{1}\grave{1}1}}{2} - A_{\grave{1}\grave{1}1}\right\}\right) \\ &\quad - \Pr\left(\left\{\frac{A_{\grave{1}11} + A_{1\grave{1}1}}{2} - A_{\grave{1}\grave{1}1}\right\} < \tilde{w} < \left\{\frac{A_{111} + A_{\grave{1}11}}{2} - A_{\grave{1}\grave{1}1}\right\}\right) \\ &= Q\left(\sqrt{\mathcal{A}_{8,1}}\right) + Q\left(\sqrt{\mathcal{A}_{11,1}}\right) - Q\left(\sqrt{\mathcal{A}_{23,1}}\right) \end{aligned} \tag{3.31}$$

and

$$\begin{aligned} P_{B_1}^{(4b)} &= \Pr\left(\tilde{w} < -A_{\grave{1}\grave{1}1}\right) \\ &\quad - \Pr\left(-\left\{A_{\grave{1}\grave{1}1} + \frac{A_{\grave{1}11} + A_{1\grave{1}1}}{2}\right\} < \tilde{w} < -\left\{A_{\grave{1}\grave{1}1} + \frac{A_{1\grave{1}1} + A_{\grave{1}\grave{1}1}}{2}\right\}\right) \\ &\quad - \Pr\left(\tilde{w} < -\left\{A_{\grave{1}\grave{1}1} + \frac{A_{111} + A_{\grave{1}11}}{2}\right\}\right) \\ &= Q\left(\sqrt{\mathcal{A}_{12,1}}\right) + Q\left(\sqrt{\mathcal{A}_{13,1}}\right) - Q\left(\sqrt{\mathcal{A}_{24,1}}\right) - Q\left(\sqrt{\mathcal{A}_{25,1}}\right). \end{aligned} \tag{3.32}$$

Thus,

$$P_{B_1}^{(4)} = Q\left(\sqrt{\mathcal{A}_{8,1}}\right) + Q\left(\sqrt{\mathcal{A}_{11,1}}\right) + Q\left(\sqrt{\mathcal{A}_{12,1}}\right) + Q\left(\sqrt{\mathcal{A}_{13,1}}\right) - Q\left(\sqrt{\mathcal{A}_{23,1}}\right) - Q\left(\sqrt{\mathcal{A}_{24,1}}\right) - Q\left(\sqrt{\mathcal{A}_{25,1}}\right). \tag{3.33}$$

The cases from 5 – 8 are dropped for brevity. Note that the analysis is identical for both sides. Therefore, the conditional BER can be expressed by

$$P_{B_1}|\gamma_1 = \sum_{i=1}^{25} c_i Q\left(\sqrt{\mathcal{A}_{i,1}}\right) \tag{3.34}$$

where $\mathbf{c} = \frac{1}{4}[2,1,1,1,2,1,1,2,1,1,1,1,1,-1,\ldots,-1]$.

3.4.3.2 Middle user analysis

Case 1: The BER of this case is

$$\begin{aligned} P_{B_2}^{(1)} &= \Pr\left(\tilde{w} < -\left\{A_{111} - \frac{A_{\grave{1}11} + A_{1\grave{1}1}}{2}\right\}\right) \\ &\quad - \Pr\left(-\left\{A_{111} + \frac{A_{\grave{1}11} + A_{1\grave{1}1}}{2}\right\} < \tilde{w} < -A_{111}\right) \\ &= Q\left(\sqrt{\mathcal{A}_{1,2}}\right) + Q\left(\sqrt{\mathcal{A}_{2,2}}\right) - Q\left(\sqrt{\mathcal{A}_{8,2}}\right). \end{aligned} \tag{3.35}$$

Case 2: The BER of this case is

$$\begin{aligned} P_{B_2}^{(2)} &= \Pr\left(\tilde{w} < -\left\{A_{\grave{1}11} - \frac{A_{\grave{1}11} + A_{1\grave{1}1}}{2}\right\}\right) \\ &\quad - \Pr\left(-\left\{A_{\grave{1}11} + \frac{A_{\grave{1}11} + A_{1\grave{1}1}}{2}\right\} < \tilde{w} < -A_{\grave{1}11}\right) \\ &= Q\left(\sqrt{\mathcal{A}_{3,2}}\right) + Q\left(\sqrt{\mathcal{A}_{4,2}}\right) - Q\left(\sqrt{\mathcal{A}_{9,2}}\right). \end{aligned} \tag{3.36}$$

Case 3: The BER can be computed as the sum of two events, i.e., $P_{B_2}^{(3)} = P_{B_2}^{(3a)} + P_{B_2}^{(3b)}$, where in

$$P_{B_2}^{(3a)} = \Pr\left(\tilde{w} > \left\{\frac{A_{\grave{1}11} + A_{1\grave{1}1}}{2} - A_{1\grave{1}1}\right\}\right) = Q\left(\sqrt{\mathcal{A}_{3,2}}\right) \tag{3.37}$$

and

$$\begin{aligned} P_{B_2}^{(3b)} &= \Pr\left(\tilde{w} < -A_{1\grave{1}1}\right) - \Pr\left(\tilde{w} < -\left\{A_{1\grave{1}1} + \frac{A_{\grave{1}11} + A_{1\grave{1}1}}{2}\right\}\right) \\ &= Q\left(\sqrt{\mathcal{A}_{5,2}}\right) - Q\left(\sqrt{\mathcal{A}_{10,2}}\right). \end{aligned} \tag{3.38}$$

Thus,

$$P_{B_2}^{(3)} = Q\left(\sqrt{\mathcal{A}_{3,2}}\right) + Q\left(\sqrt{\mathcal{A}_{5,2}}\right) - Q\left(\sqrt{\mathcal{A}_{10,2}}\right). \tag{3.39}$$

Case 4: Similarly, the BER, can be computed as the sum of two events, i.e., $P_{B_2}^{(4)} = P_{B_2}^{(4a)} + P_{B_2}^{(4b)}$, where

$$P_{B_2}^{(4a)} = \Pr\left(\tilde{w} > \left\{\frac{A_{\grave{1}11} + A_{1\grave{1}1}}{2} - A_{\grave{1}\grave{1}1}\right\}\right) = Q\left(\sqrt{\mathcal{A}_{6,2}}\right) \tag{3.40}$$

and

$$P_{B_2}^{(4b)} = \Pr\left(\tilde{w} < -A_{\grave{1}\grave{1}1}\right) - \Pr\left(\tilde{w} < -\left\{A_{\grave{1}\grave{1}1} + \frac{A_{\grave{1}11} + A_{1\grave{1}1}}{2}\right\}\right)$$
$$= Q\left(\sqrt{\mathcal{A}_{7,2}}\right) - Q\left(\sqrt{\mathcal{A}_{11,2}}\right). \tag{3.41}$$

Thus,

$$P_{B_2}^{(4)} = Q\left(\sqrt{\mathcal{A}_{6,2}}\right) + Q\left(\sqrt{\mathcal{A}_{7,2}}\right) - Q\left(\sqrt{\mathcal{A}_{11,2}}\right). \tag{3.42}$$

Finally, the conditional BER can be written as

$$P_{B_2}|\gamma_2 = \sum_{i=1}^{11} c_i Q\left(\sqrt{\mathcal{A}_{i,2}}\right) \tag{3.43}$$

where $\mathbf{c} = \frac{1}{4}[1,1,2,1,1,1,1,-1,\ldots,-1]$.

3.4.3.3 Far user analysis

Similarly, the bit error probability for cases 1, 2, 3 and 4 are given in (3.44-(3.47), respectively.

$$P_{B_3}^{(1)} = \Pr\left(\tilde{w} < -A_{111}\right) = Q\left(\sqrt{\mathcal{A}_{1,3}}\right) \tag{3.44}$$

$$P_{B_3}^{(2)} = \Pr\left(\tilde{w} < -A_{\grave{1}11}\right) = Q\left(\sqrt{\mathcal{A}_{2,3}}\right) \tag{3.45}$$

$$P_{B_3}^{(3)} = \Pr\left(\tilde{w} < -A_{1\grave{1}1}\right) = Q\left(\sqrt{\mathcal{A}_{3,3}}\right) \tag{3.46}$$

$$P_{B_3}^{(4)} = \Pr\left(\tilde{w} < -A_{\grave{1}\grave{1}1}\right) = Q\left(\sqrt{\mathcal{A}_{4,3}}\right). \tag{3.47}$$

Thus, the conditional BER for the far user is

$$P_{B_3}|\gamma_3 = \sum_{i=1}^{4} c_i Q\left(\sqrt{\mathcal{A}_{i,3}}\right) \tag{3.48}$$

where $\mathbf{c} = \frac{1}{4}[1,1,1,1]$.

3.4.4 Unconditional BER Analysis

The conditioning can be eliminated by averaging the conditional BER over the PDF of γ_n

$$P_{B_n} = \int_0^\infty P_{B_n}|\gamma_n \; p(\gamma_n)\, d\gamma_n. \tag{3.49}$$

By substituting the general representation of P_{B_n} in (3.49) we obtain,

$$P_{B_n} = \sum_i c_i \int_0^\infty Q\left(\sqrt{\mathcal{A}_{i,n}}\right) p(\gamma_n)\, d\gamma_n \tag{3.50}$$

where $\mathcal{A}_{i,n} = \frac{2|h_n|^2 g_{i,n}}{N_0} = g_{i,n}\gamma_n$ and $g_{i,n}$ is a constant. By noting that h_n is a Gaussian complex random variable, then $|h_n|$ is Rayleigh distributed, γ_n is exponentially distributed. Therefore, the PDF of γ_n is exponential as well,

$$p(\gamma_n) = \frac{1}{\bar{\gamma}_n}\exp\left(\frac{-\gamma_n}{\bar{\gamma}_n}\right) \tag{3.51}$$

where $\bar{\gamma}_n = \frac{2\mathbb{E}[|h_n|^2]}{N_0}$. Using the Craig representation of the Q-function

$$Q(x) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left(\frac{-x^2}{2\sin^2\theta}\right) d\theta \tag{3.52}$$

and the moment generation function (MGF) of the exponential PDF [22, Eq. (5.3)], the BER expression can be written as,

$$P_{B_n} = \sum_i \frac{1}{\pi} c_i \int_0^{\pi/2} M_{\gamma_n}\left(\frac{-\mathcal{A}_{i,n}}{2\sin^2\theta}\right) d\theta = \frac{1}{2}\sum_i c_i\, \zeta(\mathcal{A}_{i,n}) \tag{3.53}$$

where $M_{\gamma_n}(-s) = 1/(1+s\bar{\gamma}_n)$ and c_i is the coefficient multiplied by the Gaussian function in the conditional BER. Also, $\zeta(\mathcal{A}_{i,n}) = 1 - \sqrt{\overline{\mathcal{A}}_{i,n}/(2+\overline{\mathcal{A}}_{i,n})}$ and $\overline{\mathcal{A}}_{i,n} = \frac{2\mathbb{E}[|h_n|^2]g_{i,n}}{N_0}$.

3.5 Generalized BER Analysis

To simplify the BER analysis in this section, a simple example shown in Figure 3.6 is considered where $\mathbf{m} = [4,4,2]$. The analysis for this example is generic where the steps are based on the principles shown in Section 3.4. These steps can be applied for

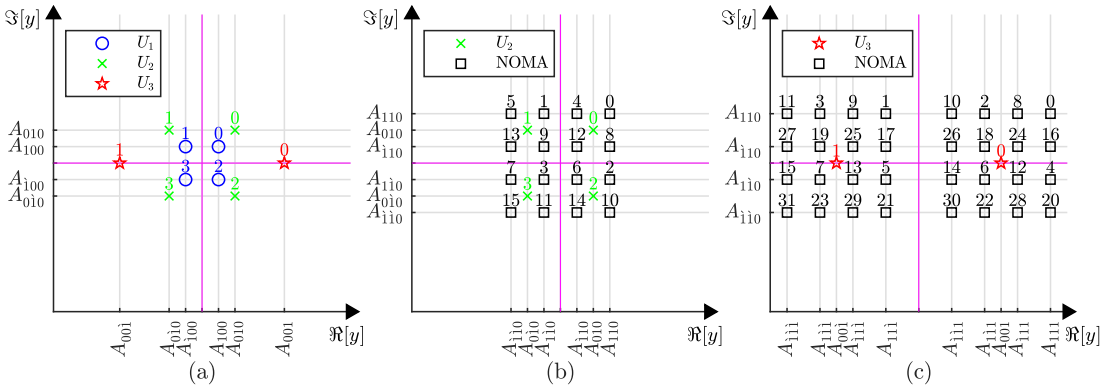


Figure 3.6: Constellation points of: (a) All users without superposition coding. (b) U_2 and the superposition coding of U_1 and U_2. (c) U_3 and the superposition coding of U_1, U_2 and U_3.

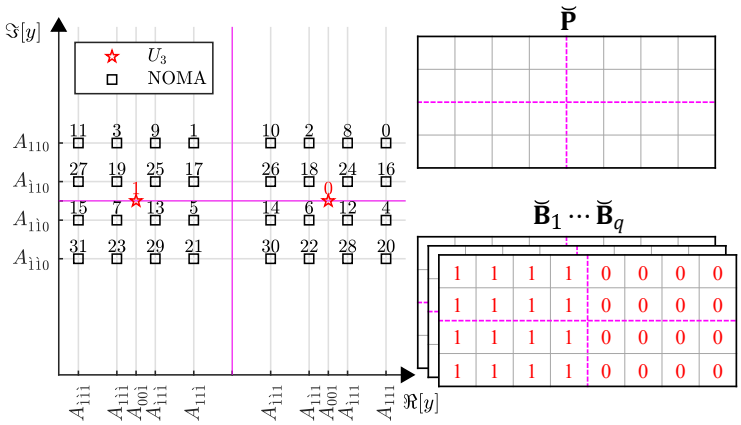


Figure 3.7: Mapping NOMA constellation points to scatter matrices $\breve{\mathbf{P}}$ and $\breve{\mathbf{B}}_k$ for $\mathbf{m} = [4,4,2]$.

an arbitrary number of users and arbitrary modulation orders. In this section, the conditional BER expressions will be derived and then the unconditional BER expressions will be shown. To get the BER, all possible transmitted symbols must be considered. Nonetheless for equally probable symbols, the BER can be calculated considering only the first quadrant in the space diagram due to symmetry. The analysis of the conditional BER is based on the probability that a transmitted NOMA symbol leads to a erroneously received bit. Therefore, finding the decision regions for each bit is crucial. Once these decision regions are determined, the BER calculation for each transmitted symbol can be calculated by finding the Euclidean distance to the decision regions' boundarys (DRBs).

3.5.1 Decision Regions' Boundaries

To evaluate the BER, the DRBs for each bit should be specified. This can be achieved by segmenting the NOMA constellation into q constellation diagrams each of which

Algorithm 3.1: Generation of scatter matrix $\breve{\mathbf{P}}$.

Input: $\mathbf{p}$
Output: $\breve{\mathbf{P}}$

$u = \text{length}(\text{unique}(\text{Re}[\mathbf{p}]))$
$v = \text{length}(\text{unique}(\text{Im}[\mathbf{p}]))$
$\mathbf{p}_{\text{Re}} = \text{sort}(\text{Re}[\mathbf{p}], \text{“ascend”})$
$\mathbf{P}_{\text{Re}} = \text{reshape}(\mathbf{p}_{\text{Re}}, [u, v])$
$\mathbf{p}^{*}_{\text{Im}} = \text{sort}(\text{Im}[\mathbf{p}], \text{“descend”})$
$\mathbf{p}^{**}_{\text{Im}} = \text{unique}(\mathbf{p}^{*}_{\text{Im}})$
$\mathbf{p}_{\text{Im}} = \text{reshape}(\mathbf{p}^{**}_{\text{Im}}, [u, 1])$
$\breve{\mathbf{P}} = \mathbf{P}_{\text{Re}} + 1j \times \mathbf{p}_{\text{Im}}$

corresponds to a particular bit. For example, the binary representation of the top-left symbol in Figure 3.7 is 01011 since $q = 5$. Consequently, the top-left bit in each of the 5 constellations will be 0, 1, 0, 1, 1, respectively. The remaining points in all bit constellations can be obtained by following the same approach. To generate the constellations more systematically, we define $\mathbf{p} \in \mathbb{C}^{1\times 2^q}$ as a vector that contains all possible NOMA symbols. Then vector $\dot{\mathbf{b}}$ is generated by converting the symbols in $\mathbf{p}$ into binary, and $\dot{\mathbf{b}}_k \in \mathbb{B}^{1\times 2^q}$, $k \in \{1, 2, \ldots, q\}$, is obtained by segmenting $\dot{\mathbf{b}}$ into q vectors $\dot{\mathbf{b}}_1, \dot{\mathbf{b}}_2, \ldots, \dot{\mathbf{b}}_q$, where $\dot{\mathbf{b}}_1$ contains MSB of all symbols.

The next step is to generate the scatter matrix $\breve{\mathbf{P}} \in \mathbb{C}^{u\times v}$, which can be performed using Algorithm 3.1. The elements of matrix $\breve{\mathbf{P}}$ are the NOMA symbols arranged exactly according to the constellation diagram. For the example in Figure 3.7, $\breve{P}_{1,1} = A_{\grave{1}\grave{1}\grave{1}} + 1jA_{110}$. The values of u and v are given in Algorithm 3.1. Similarly, each kth bit vector, $\dot{\mathbf{b}}_k$, will have a scatter matrix $\breve{\mathbf{B}}_k \in \mathbb{B}^{u\times v}$. The algorithms to produce the scatter matrices $\breve{\mathbf{P}}$ and $\breve{\mathbf{B}}_k$, $\forall k$ are given in Algorithms 3.1 and 3.2, respectively. Algorithm 3.1 is based on finding the real and imaginary amplitude levels and creating the scatter matrix $\breve{\mathbf{P}}$ by moving from left to right and top to bottom. Furthermore, Algorithm 3.2 is mainly based on the results from Algorithm 3.1, where symbol to binary mapping is done to find the kth bit for each constellation point in $\breve{\mathbf{P}}$. Note that the functions used in Algorithm 3.1 can be found in advanced mathematical software packages such as Matlab, where length$(\cdot)$ finds the length of a vector, unique$(\cdot)$ returns the unique elements of a vector, sort$(\cdot)$ orders the elements of a vector, and reshape$(\cdot)$ transforms the array size. tbh!

Once $\breve{\mathbf{P}}$ is calculated, the real and imaginary primary DRBs, $\mathbf{d}_{\text{Re}} \in \mathbb{R}^{1\times u-1}$ and $\mathbf{d}_{\text{Im}} \in \mathbb{R}^{1\times v-1}$, can be found by using a sliding window averaging filter with a window

Algorithm 3.2: Generation of scatter matrices $\breve{\mathbf{B}}_k$, $\forall k$.

Input: $\dot{\mathbf{b}}_k$, $\mathbf{p}$, $\breve{\mathbf{P}}$
Output: $\breve{\mathbf{B}}_k$, $\forall k$
for $i = 1 : u$ **do**
  for $j = 1 : v$ **do**
    $\breve{P}_{i,j} \xrightarrow{\text{to binary}} \mathbf{b}$
    for $k = 1 : q$ **do**
      $\breve{B}_{i,j}^{(k)} = b_k$

of size 2 whose output can be written as

$$d_U^{\text{Re}} = \frac{1}{2}\text{Re}\left[\breve{P}_{1,U} + \breve{P}_{1,U+1}\right] \tag{3.54}$$

and

$$d_V^{\text{Im}} = \frac{1}{2}\text{Im}\left[\breve{P}_{V,1} + \breve{P}_{V+1,1}\right] \tag{3.55}$$

where $U \in \{1, 2, \ldots, u-1\}$ and $V \in \{1, 2, \ldots, v-1\}$. The kth bit DRBs can be computed using the scatter matrix $\breve{\mathbf{B}}_k$, where the DRBs appear if there is a bit flip. Therefore, exclusive-OR operator can be used over one row or one column of $\breve{\mathbf{B}}_k$ to find the bit flip location. This can be expressed as follows,

$$t_{\text{Re}_U}^{(k)} = \breve{B}_{1,U}^{(k)} \oplus \breve{B}_{1,U+1}^{(k)} \tag{3.56}$$

and

$$t_{\text{Im}_V}^{(k)} = \breve{B}_{V,1}^{(k)} \oplus \breve{B}_{V+1,1}^{(k)} \tag{3.57}$$

where $\mathbf{t}_k^{\text{Re}} \in \mathbb{B}^{1\times u-1}$ and $\mathbf{t}_k^{\text{Im}} \in \mathbb{B}^{1\times v-1}$. Using the constellation diagrams in Figure 3.6, it can be noted that the kth bit flips in either $\mathbf{t}_k^{\text{Re}}$ or $\mathbf{t}_k^{\text{Im}}$, but not in both at the same time. The total number of DRBs for the kth bit is given by

$$\vartheta_k = \begin{cases} \vartheta_k^{\text{Re}}, & \vartheta_k^{\text{Im}} = 0 \\ \vartheta_k^{\text{Im}}, & \vartheta_k^{\text{Re}} = 0 \end{cases} \tag{3.58}$$

where $\vartheta_k^{\text{Re}} = \sum_{i=1}^{u-1} t_{\text{Re}_i}^{(k)}$ and $\vartheta_k^{\text{Im}} = \sum_{i=1}^{v-1} t_{\text{Im}_i}^{(k)}$. Therefore, the kth bit's DRBs, $\dot{\mathbf{d}}_k \in \mathbb{R}^{1\times\vartheta_k}$, can be found by considering the indices of $\mathbf{t}_k^{\text{Re}}$ or $\mathbf{t}_k^{\text{Im}}$ where the entry is 1. These indices

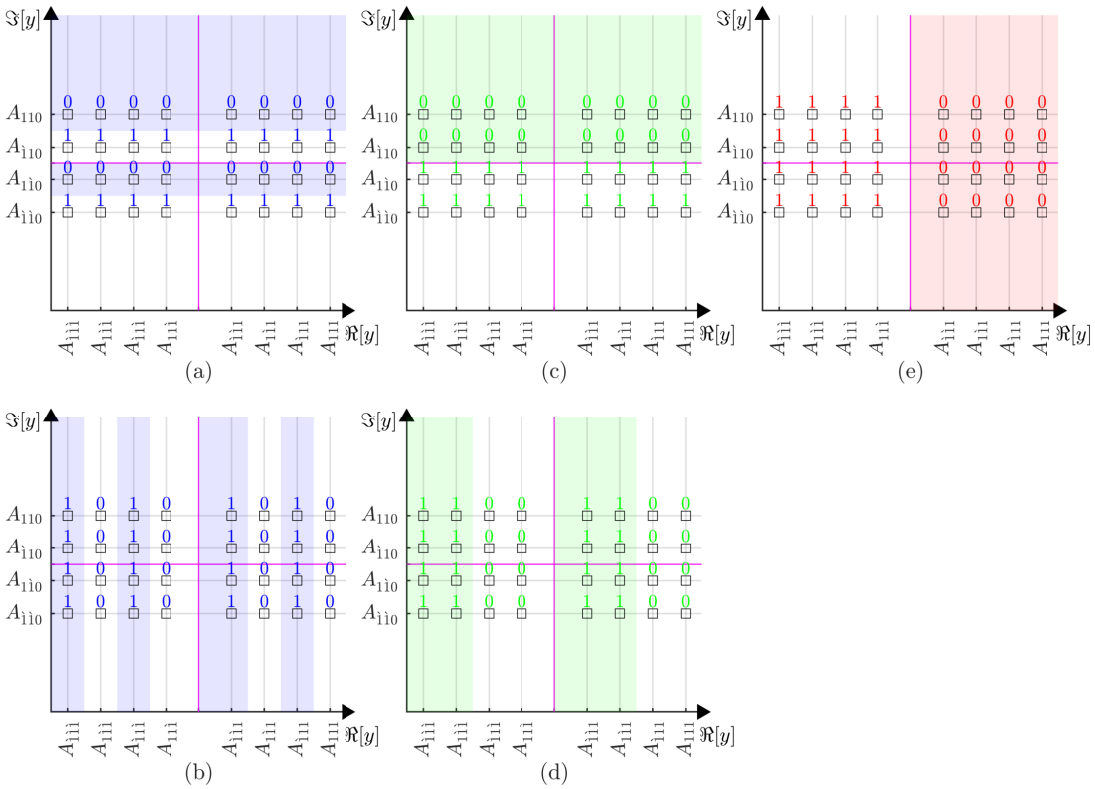


Figure 3.8: NOMA constellation points for $\mathbf{m} = [4,4,2]$ showing the decision regions for: (a) b_1. (b) b_2. (c) b_3. (d) b_4. (e) b_5.

are stored in $\mathbf{z}_k \in \mathbb{R}^{1\times\vartheta_k}$, and hence $\dot{\mathbf{d}}_k$ elements can be found by

$$\dot{d}_i^{(k)} = \begin{cases} d^{\mathrm{Re}}_{z_i^{(k)}}, & \vartheta_k^{\mathrm{Im}} = 0 \\ d^{\mathrm{Im}}_{z_i^{(k)}}, & \vartheta_k^{\mathrm{Re}} = 0 \end{cases} . \tag{3.59}$$

Note that $\breve{\mathbf{d}}_k$ sorts the elements of $\dot{\mathbf{d}}_k$ in a descending order, which is required to find the bit error probability as well as the coefficients matrix which will be explained in the following subsection. The DRBs of all NOMA bits for $\mathbf{m} = [4,4,2]$ are shown in Figure 3.8 using different color shading.

3.5.2 Euclidean Distance Computation

The BER expressions can be obtained by computing the Euclidean distance between the constellation points and the DRBs in $\breve{\mathbf{d}}_k$ $\forall k$. However, due to constellation diagram symmetry, it is sufficient to consider only the first quadrant. Thus, we are interested in $\mathbf{p}^+$ which contains the first quadrant symbols of the scatter matrix $\breve{\mathbf{P}}$, where $\mathbf{p}^+ \in \mathbb{C}^{1\times 2^{q-2}}$. Therefore, the displacement matrix for the kth bit, $\Delta_k \in \mathbb{R}^{\vartheta_k \times 2^{q-2}}$, computes

the displacement between the first quadrant constellation points and the DRBs which can be expressed by (3.60). Note that each column in Δ_k corresponds to a specific constellation point in $\tilde{\mathbf{p}}_k^+$, which is defined in (3.61), while each row corresponds to a specific DRB in $\breve{\mathbf{d}}_k$.

$$\Delta_k = [\delta_1^{(k)}, \delta_2^{(k)}, \ldots, \delta_{2^{q-2}}^{(k)}] = \tilde{\mathbf{p}}_k^+ - \breve{\mathbf{d}}_k^T \tag{3.60}$$

$$\tilde{\mathbf{p}}_k^+ = \begin{cases} \mathrm{Re}[\mathbf{p}^+], & \vartheta_k^{\mathrm{Im}} = 0 \\ \mathrm{Im}[\mathbf{p}^+], & \vartheta_k^{\mathrm{Re}} = 0 \end{cases} . \tag{3.61}$$

3.5.3 Conditional BER Analysis

The conditional BER can be derived by considering all transmitted and received bit combinations and their relation to the DRBs. After exhaustive manipulations, the conditional BER per bit can be expressed as

$$P_B^{(k)} | \gamma_k^* = \frac{1}{2^{q-2}} \sum_{i=1}^{\vartheta_k} \sum_{j=1}^{2^{q-2}} c_{i,j}^{(k)} Q\left(\sqrt{\mathcal{A}_{i,j}^{(k)}}\right) \tag{3.62}$$

where

$$\boldsymbol{\mathcal{A}}_k = \frac{2\left|h_k^*\right|^2 \mathbf{E}_k}{N_0} = \mathbf{E}_k \gamma_k^* \tag{3.63}$$

and the squared Euclidean distance matrix $\mathbf{E}_k$ can be calculated using Δ_k, where $E_{i,j}^{(k)} = \left|\Delta_{i,j}^{(k)}\right|^2$, $\boldsymbol{\mathcal{A}}_k = \left[\mathbf{a}_1^{(k)}, \mathbf{a}_2^{(k)}, \ldots, \mathbf{a}_{2^{q-2}}^{(k)}\right]$, $h_k^* = h_n$ and $\gamma_k^* = \gamma_n$ iff $O_n \leq k < O_n + \mathcal{M}_n$ and $n \in \{1, 2, \ldots, N\}$. Note that the user index is dropped for notational simplicity. The coefficients matrix is defined as $\mathbf{C}_k = \left[\mathbf{c}_1^{(k)}, \mathbf{c}_2^{(k)}, \ldots, \mathbf{c}_{2^{q-2}}^{(k)}\right]$, $\mathbf{C}_k \in \mathbb{Z}^{\vartheta_k \times 2^{q-2}}$. For $\vartheta_k = 1$, $\mathbf{C}_k$ will reduce to a row vector of length 2^{q-2} where $c_{1,j}^{(k)} = 1 \ \forall j$. For $\vartheta_k > 1$, the elements of $\mathbf{C}_k$ can be calculated as

$$c_{i,j}^{(k)} = \begin{cases} (-1)^{i+1}, & g_j^{(k)} = 0 \\ \varphi\left(i, g_j^{(k)}\right), & \textit{Otherwise} \end{cases} \tag{3.64}$$

where $g_j^{(k)} = \sum_{i=1}^{\vartheta_k} \psi_{i,j}^{(k)}$ and

$$\psi_{i,j}^{(k)} = \begin{cases} 1, & \Delta_{i,j}^{(k)} < 0 \\ 0, & \textit{Otherwise} \end{cases} \tag{3.65}$$

$$\mathbf{C}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \grave{1} & 1 & \grave{1} & 1 & \grave{1} & 1 & \grave{1} & 1 \\ 1 & \grave{1} & 1 & \grave{1} & 1 & \grave{1} & 1 & \grave{1} \end{bmatrix}$$

$$\mathbf{C}_2 = \begin{bmatrix} 1 & 1 & \grave{1} & \grave{1} & 1 & 1 & 1 & 1 \\ \grave{1} & \grave{1} & 1 & 1 & 1 & 1 & \grave{1} & \grave{1} \\ 1 & 1 & 1 & 1 & \grave{1} & \grave{1} & 1 & 1 \\ 1 & 1 & \grave{1} & \grave{1} & 1 & 1 & \grave{1} & \grave{1} \\ \grave{1} & \grave{1} & 1 & 1 & \grave{1} & \grave{1} & 1 & 1 \\ 1 & 1 & \grave{1} & \grave{1} & 1 & 1 & \grave{1} & \grave{1} \\ \grave{1} & \grave{1} & 1 & 1 & \grave{1} & \grave{1} & 1 & 1 \end{bmatrix}$$

$$\mathbf{C}_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{C}_4 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & \grave{1} & \grave{1} & \grave{1} & \grave{1} \\ \grave{1} & \grave{1} & \grave{1} & \grave{1} & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{C}_5 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{p}^{+} = \begin{bmatrix} \breve{p}_{1,5} & \breve{p}_{2,5} & \breve{p}_{1,6} & \breve{p}_{2,6} & \breve{p}_{1,7} & \breve{p}_{2,7} & \breve{p}_{1,8} & \breve{p}_{2,8} \end{bmatrix}$$

Figure 3.9: $\mathbf{C}_k$ and $\mathbf{p}^{+}$ for the example of $\mathbf{m} = [4,4,2]$.

and

$$\varphi\left(i, g_j^{(k)}\right) = \begin{cases} (-1)^{i+1}, & g_j^{(k)}\colon \{\geq i, \text{odd}\} \text{ or } \{< i, \text{even}\} \\ (-1)^{i}, & \textit{Otherwise}. \end{cases} \tag{3.66}$$

The expression in (3.62) considers all the constellation points in the first quadrant of the space diagram. Thus, the weighting factor of $1/2^{q-2}$ is considered as these constellation points are equally probable. For the special case of identical BPSK modulation orders, this weighting factor becomes $1/2^{N-1}$. Furthermore, each column in $\boldsymbol{\mathcal{A}}_k$ corresponds to a specific constellation point where the sum over that column gives the probability of error for that constellation point.

To demonstrate (3.60)–(3.66), the example shown in Figures 3.6–3.8 is considered and $\mathbf{C}_k, \forall k$ is found and shown in Figure 3.9. For brevity, $P_B^{(2)}|\gamma_2^*$ is computed for the NOMA word $\mathbf{b} = \mathbf{b}^E = [0,1,0,1,0]$. It can be seen from Figure 3.8b that b_2^E does not flip by moving vertically, i.e., it flips only by moving horizontally. The displacement between the NOMA word $\mathbf{b}^E$ and $\breve{\mathbf{d}}_2$ can be calculated as

$$\delta_1^{(2)} = \tilde{p}_1^{(2)^{+}} - \breve{\mathbf{d}}_2^T = \left[\Delta_{1,1}^{(2)}, \Delta_{2,1}^{(2)} \ldots, \Delta_{\vartheta_2,1}^{(2)}\right]^T. \tag{3.67}$$

The squared Euclidean distance for (3.67) can be calculated by squaring the vector elements, i.e. $\mathbf{e}_1^{(2)} = \delta_1^{(2)} \odot \delta_1^{(2)} = \left[E_{1,1}^{(2)}, E_{2,1}^{(2)}, \ldots, E_{\vartheta_2,1}^{(2)}\right]^T$. Thus, the first column of $\boldsymbol{\mathcal{A}}_k$ in (3.63) can be written as

$$\mathbf{a}_1^{(2)} = \frac{2\,|h_2^*|^2\,\mathbf{e}_1^{(2)}}{N_0} = \left[\mathcal{A}_{1,1}^{(2)}, \mathcal{A}_{2,1}^{(2)}, \ldots, \mathcal{A}_{\vartheta_2,1}^{(2)}\right]^T. \tag{3.68}$$

The probability that b_2 is detected erroneously, given that the NOMA word $\mathbf{b}^E$ is transmitted, can be calculated as

$$\Pr\left(\widehat{b}_2 \neq b_2 | \mathbf{b} = \mathbf{b}^E\right) = \Pr\left(\tilde{w} > \left|\Delta_{1,1}^{(2)}\right|\right) + \Pr\left(\left|\Delta_{3,1}^{(2)}\right| < \tilde{w} < \left|\Delta_{2,1}^{(2)}\right|\right) + \Pr\left(-\left|\Delta_{5,1}^{(2)}\right| < \tilde{w} < -\left|\Delta_{4,1}^{(2)}\right|\right) + \Pr\left(-\left|\Delta_{7,1}^{(2)}\right| < \tilde{w} < -\left|\Delta_{6,1}^{(2)}\right|\right). \tag{3.69}$$

By noting that $\tilde{w}_n \sim \mathcal{N}(0, N_0/2)$, then it is straightforward to show that

$$\Pr\left(\widehat{b}_2 \neq b_2 | \mathbf{b} = \mathbf{b}^E\right) = \sum_{i=1}^{7} c_{i,1}^{(2)} Q\left(\sqrt{\mathcal{A}_{i,1}^{(2)}}\right). \tag{3.70}$$

The coefficients $c_{i,1}^{(2)}$, $\forall i$ can be calculated by noting that $g_1^{(2)} = 3$ for this case. Therefore, using (3.64) gives $\mathbf{c}_1^{(2)} = [+1, -1, +1, +1, -1, +1, -1]^T$. Consequently, the conditional BER is calculated for b_2 given that the NOMA word $\mathbf{b}^E$ is transmitted, and the same approach should be repeated for all constellation points in $\tilde{\mathbf{p}}_k^+$ to compute the overall conditional BER per bit (3.62). Finally, the nth user conditional BER can be found by averaging its $P_B^{(k)}|\gamma_k^*$ expressions (3.62). This can be written as

$$P_{B_n}|\gamma_n = \frac{1}{\mathcal{M}_n} \sum_{k=O_n}^{O_n+\mathcal{M}_n} P_B^{(k)}|\gamma_k^*. \tag{3.71}$$

3.5.4 BER Analysis without Receiver Diversity

Similar to Section 3.4 the conditioning can be eliminated by averaging the conditional BER over the PDF of γ_k. Thus the unconditional BER per user over a Rayleigh fading channel can be written as

$$P_{B_n} = \frac{1}{2^{q-1}\mathcal{M}_n} \sum_{k=O_n}^{O_n+\mathcal{M}_n} \sum_{i=1}^{\vartheta_k} \sum_{j=1}^{2^{q-2}} c_{i,j}^{(k)} \psi_{i,j}^{(k)} \tag{3.72}$$

where $\psi_{i,j}^{(k)} \triangleq \zeta(\mathcal{A}_{i,j}^{(k)}) = 1 - \sqrt{\overline{\mathcal{A}}_{i,j}^{(k)} / (2 + \overline{\mathcal{A}}_{i,j}^{(k)})}$ and $\overline{\mathcal{A}}_{i,j}^{(k)} = \frac{2\mathbb{E}[|h_k^*|^2]E_{i,j}^{(k)}}{N_0}$. It is worth noting that using the binomial series expansion, the asymptotic BER can be obtained by substituting $\psi_{i,j}^{(k)} \to \frac{1}{\overline{\mathcal{A}}_{i,j}^{(k)}}$ into (3.72) [21, p. 185].

3.5.5 BER Analysis with Adaptive Power Assignment

The power assignment for NOMA has a major impact on the users' BER, and thus, it has been considered widely in the literature [26, 34, 35, 43, 48, 75, 76]. The power assignment for NOMA can be generally classified into three types, which are fixed power assignment [26, 34, 35], adaptive power assignment based on channel statistical information [35, 43, 48, 51], and adaptive power assignment based on the instantaneous fading coefficients [75, 76]. The adaptive power assignment is usually performed to minimize the system's average error probability [43, 48, 76], or provide certain error probability for each user [75]. For the fixed power allocation, the conditional and average BER expressions are given by (3.71) and (3.72), respectively.

In the case of adaptive power assignment based on the channel statistics, the conditional BER in (3.71) can be used to derive the average BER for any channel model by using the corresponding PDF of γ_n in (3.63). It is worth noting that the power adaptation process is performed after avenging over the PDF of γ_n. The analysis for the Rayleigh channel model is given by (3.53). For the third type, the BER should be conditioned on both, the channel instantaneous fading coefficients and the used power coefficients. In such scenarios, the power coefficients depend on the fading coefficients. Hence, they can be written as $\alpha = \mathcal{F}(h_1, h_2, \ldots, h_N)$, where $\mathcal{F}(\cdot)$ is a general function that depends on the adopted optimization criterion. Therefore, the BER can be evaluated by replacing α with $\mathcal{F}(h_1, h_2, \ldots, h_N)$ in (3.71), and then averaging over the PDF of γ_n. However, deriving the function $\mathcal{F}(\cdot)$ and evaluating the average BER for this power assignment strategy is highly challenging because the relation between the power and channel coefficients is highly nonlinear [75, 76]. In such scenarios, the desired power coefficients can be obtained using a particular numerical search method using (3.71).

3.5.6 BER Analysis with Receiver Diversity

In this work, we consider that the nth user receiver is equipped with L_n receiving antennas, and the channels between the base station and all receiving antennas are i.i.d Therefore, the received signal for the nth user is

$$\mathbf{y}_n = \mathbf{h}_n x_{SC} + \mathbf{w}_n \tag{3.73}$$

where $\{\mathbf{y}_n, \mathbf{h}_n, \hbar_n, \mathbf{w}_n\} \in \mathbb{C}^{L_n \times 1}$, and the entries of these vectors are defined similar to those of (3.4). The detector in this case can be realized as a maximum ratio combining

(MRC) followed by the JMLD. Therefore,

$$\{\widehat{x}_1,\widehat{x}_2,...,\widehat{x}_N\} = \arg\min_{x_i\in\chi_i}\left|\mathbf{h}_n^H\mathbf{y}_n - \|\mathbf{h}_n\|^2\sum_{i=1}^{N}\sqrt{\alpha_i}x_i\right|^2. \tag{3.74}$$

However, the BER for the SIMO case is similar to the SISO one except that the matrix $\boldsymbol{\mathcal{A}}_k$ will be replaced by $\boldsymbol{\mathcal{B}}_k = \sum_{i=1}^{L_k^*}\boldsymbol{\mathcal{A}}_k^{(i)}$, where $\boldsymbol{\mathcal{A}}_k^{(i)}$ is equal to $\boldsymbol{\mathcal{A}}_k$ per receiving antenna. Therefore, the PDF of $\mathcal{B}_{i,j}^{(k)}$, $\forall k$ is a Chi-square with $2L_k^*$ degrees of freedom [21, pp. 214-215],

$$f_{\mathcal{B}_{i,j}^{(k)}}\left(\mathcal{A}_{i,j}^{(k)}\right) = \frac{\left(\mathcal{A}_{i,j}^{(k)}\right)^{L_k^*-1}}{\left(\overline{\mathcal{A}}_{i,j}^{(k)}\right)^{L_k^*}(L_k^*-1)!}\exp\left(-\frac{\mathcal{A}_{i,j}^{(k)}}{\overline{\mathcal{A}}_{i,j}^{(k)}}\right) \tag{3.75}$$

where $L_k^* = L_n$ iff $O_n \le k < O_n + \mathcal{M}_n$. Based on (3.75) and [94, Eq. (23)], the BER per bit can be evaluated as

$$P_B^{(k)} = \frac{1}{2^{q-1}}\sum_{i=1}^{\vartheta_k}\sum_{j=1}^{2^{q-2}}c_{i,j}^{(k)}\left[1-\psi_{i,j}^{(k)}\sum_{r=0}^{L_k^*-1}\binom{2r}{r}\left(\frac{1-(\psi_{i,j}^{(k)})^r}{4}\right)^r\right]. \tag{3.76}$$

The nth user average BER can be computed by averaging its $P_B^{(k)}$ expressions similar to (3.71). Moreover, a tight asymptotic BER per bit can be calculated as follows [22, pp. 326-327],

$$P_{B,\infty}^{(k)} = \frac{1}{2^{q+L_k^*-1}}\binom{2L_k^*}{L_k^*}\sum_{i=1}^{\vartheta_k}\sum_{j=1}^{2^{q-2}}c_{i,j}^{(k)}\left(\frac{1}{\overline{\mathcal{A}}_{i,j}^{(k)}}\right)^{L_k^*}. \tag{3.77}$$

While considering the most dominant term in (3.77), the diversity order is L_k^* and the coding gain is the most dominant $E_{i,j}^{(k)}$ term [22, p. 798].

3.6 PCB Analysis

To enable reliable detection of NOMA symbols using SIC, the power coefficient for each user should be selected such that the constellation for each user does not overlap with the other users' constellations [26, 27, 34, 50]. The PCB can be generally derived by noting that the nearest constellation point to the origin (NCO) in the first quadrant of the x-y plane should not cross the x or y axes. However, due to axes symmetry, it is sufficient to consider the real part of the NCO. Therefore, for $N = 2$, we have

$\sqrt{\frac{\alpha_2}{\kappa_2}} - \Lambda_1\sqrt{\frac{\alpha_1}{\kappa_1}} > 0$. Consequently, the PCB can be written as $\frac{\alpha_1}{\alpha_2} < \frac{\kappa_1}{\kappa_2}\frac{1}{\Lambda_1^2}$ and the maximum possible power coefficient for U_1 can be obtained by noting that $\alpha_2 = 1-\alpha_1$, and thus $\alpha_{1,\max}^{(N=2)} = \frac{\kappa_1}{\kappa_1+\kappa_2\Lambda_1^2}$. For $M_n = 2,4,8,16,64$ the factors $\kappa_n = 1,2,6,10,42$, and $\Lambda_n = 1,1,3,3,7$, respectively.

The PCB for $N = 3$ can be evaluated recursively by considering first the $N = 2$ case where U_1 and U_2 will be combined. Then the constellation that resulted from combining U_1 and U_2 constellations is considered as one constellation that will be combined with that of U_3. Therefore, the first constraint that should be satisfied is identical to $N = 2$ case which is

$$\alpha_2 > \rho(\alpha_1) = \frac{\kappa_2}{\kappa_1}\Lambda_1^2\alpha_1 \triangleq \rho(\alpha_1). \tag{3.78}$$

The second step is to ensure that the NCO in the combined constellation does not cross the y-axis to the negative side. Therefore, the constraint can be written as

$$\sqrt{\frac{1-\alpha_1-\alpha_2}{\kappa_3}} - \left[\Lambda_1\sqrt{\frac{\alpha_1}{\kappa_1}} + \Lambda_2\sqrt{\frac{\alpha_2}{\kappa_2}}\right] > 0. \tag{3.79}$$

Solving the inequality in (3.79) for either α_1 or α_2 results in two solutions, $\eta(\alpha_n)$ and $\varepsilon(\alpha_n)$ for $n \in \{1,2\}$, but one of them is not applicable. The desired solution of (3.79) with respect to α_2 will be denoted as $\eta(\alpha_1)$. Therefore, the second constraint becomes $\alpha_2 < \eta(\alpha_1)$. As an example, consider the case of $\mathbf{m} = [4,4,2]$ in Figure 3.6, where the obtained solution is $\eta(\alpha_1) = \frac{2}{3} - \frac{7}{9}\alpha_1 - \frac{2}{9}\sqrt{\alpha_1}\sqrt{6-8\alpha_1}$. Figure 3.10 visualizes (3.78) and (3.79) which makes it easier to infer the PCBs. The intersection between (3.78) and $\eta(\alpha_1)$ reflects the maximum possible power coefficient for U_1, i.e., $\alpha_{1,\max}^{(N=3)}$. Therefore, the pair (α_1,α_2) that satisfies the PCBs must be inside the region having the bounds of $\alpha_2 > \alpha_1$ and $\alpha_2 < \frac{2}{3} - \frac{7}{9}\alpha_1 - \frac{2}{9}\sqrt{\alpha_1}\sqrt{6-8\alpha_1}$.

Generally, the condition in (3.78) could intersect with the desired solution in (3.79) more than once. However, the desired $\alpha_{1,\max}^{(N=3)}$ is the intersection that gives the minimum value of α_1. For the arbitrary modulation orders cases, the PCBs are given by (3.78) and

$$\eta(\alpha_1) = \frac{\kappa_2}{\kappa_1\varpi_1^2}\left(\varpi_1[\varpi_2+\varpi_3] - 2\Lambda_1\Lambda_2\kappa_3\sqrt{\alpha_1[\varpi_3-\kappa_2\varpi_2]}\right) \tag{3.80}$$

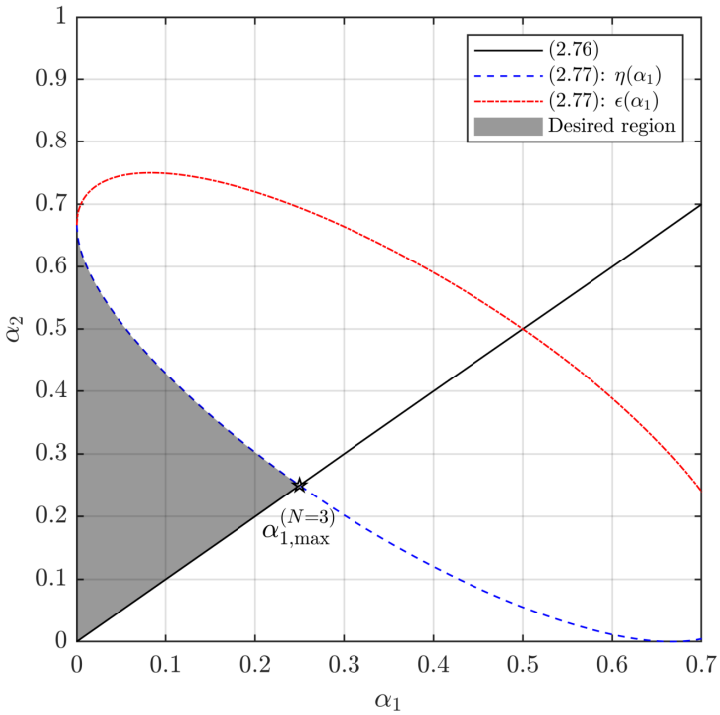


Figure 3.10: Visualizing the PCBs for $\mathbf{m} = [4,4,2]$.

Table 3.2: The PCBs for selected modulation orders, $N = 3$

$\mathbf{m}$	$\alpha_{1,\max}^{(N=3)}$	$\rho(\alpha_1)$	$\eta(\alpha_1)$
$[2,2,2]$	$\frac{1}{6}$	α_1	$\frac{1}{2}-\sqrt{\alpha_1}\left(\frac{1}{2}\sqrt{2-3\alpha_1}+\frac{1}{2}\sqrt{\alpha_1}\right)$
$[4,4,4]$	$\frac{1}{6}$	α_1	$\frac{1}{2}-\sqrt{\alpha_1}\left(\frac{1}{2}\sqrt{2-3\alpha_1}+\frac{1}{2}\sqrt{\alpha_1}\right)$
$[8,8,8]$	$\frac{1}{154}$	$9\alpha_1$	$\frac{1}{10}-\frac{9}{5}\sqrt{\alpha_1}\left(\frac{1}{10}\sqrt{10-19\alpha_1}-\frac{9}{10}\sqrt{\alpha_1}\right)-\alpha_1$
$[16,16,16]$	$\frac{1}{154}$	$9\alpha_1$	$\frac{1}{10}-\frac{9}{5}\sqrt{\alpha_1}\left(\frac{1}{10}\sqrt{10-19\alpha_1}-\frac{9}{10}\sqrt{\alpha_1}\right)-\alpha_1$
$[64,64,64]$	$\frac{1}{3186}$	$49\alpha_1$	$\frac{1}{50}-\frac{49}{25}\sqrt{\alpha_1}\left(\frac{1}{50}\sqrt{50-99\alpha_1}-\frac{49}{50}\sqrt{\alpha_1}\right)-\alpha_1$

where $\varpi_1=\Lambda_2^2\kappa_3+\kappa_2$, $\varpi_2=\Lambda_1^2\alpha_1\kappa_3$, $\varpi_3=\kappa_1(1-\alpha_1)$. Additionally,

$$\alpha_{1,\max}^{(N=3)} = \frac{\kappa_1}{\Lambda_1^2\Lambda_2\kappa_3(\Lambda_2+2)+\Lambda_1^2(\kappa_2+\kappa_3)+\kappa_1}. \tag{3.81}$$

Table 3.2 summarizes the PCBs for the identical modulation schemes. It can be seen that the power coefficients space becomes smaller as the modulation order increases. Also, it is worth mentioning that the space of the power coefficients for $N = 3$ becomes narrower compared to $N = 2$ because of the extra PCB introduced. Following the same approach, the PCBs for other values of N can be derived utilizing the conditions in [27].

These closed-form PCBs expressions can be used as linear and non-linear constraints while solving minimization or maximization optimization problems. For example, by noting that the average BER of an N users NOMA system is given by $P_{B,\text{Avg.}}^{(N)} = \frac{1}{N}\sum_{n=1}^{N} P_{B_n}$. Then the optimal power assignment that minimizes $P_{B,\text{Avg.}}^{(N=2)}$ is formulated as

$$\min_{\alpha} P_{B,\text{Avg}}^{(N=2)} \tag{3.82a}$$

subject to,

$$\alpha_1 < \alpha_{1,\max}^{(N=2)} \tag{3.82b}$$

$$\alpha_1 + \alpha_2 = 1 \tag{3.82c}$$

where (3.82b) satisfies the PCB condition and (3.82c) is considered to ensure that the transmitted power is normalized to unity. The objective function in (3.82a) is non-linear, and hence, it is difficult to find a closed-form analytical solution for this problem. Therefore, the problem can be solved using interior-point optimization, which provides near-optimal solutions [48]. Similarly, the optimization problem can be extended to $N = 3$ case. As such, the optimization problem is formulated as

$$\min_{\alpha} P_{B,\text{Avg}}^{(N=3)} \tag{3.83a}$$

subject to,

$$\alpha_1 < \alpha_{1,\max}^{(N=3)} \tag{3.83b}$$

$$\rho(\alpha_1) < \alpha_2 < \eta(\alpha_1) \tag{3.83c}$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1 \tag{3.83d}$$

where (3.83b) is linear, and it is similar to $N = 2$ case. However, (3.83c) is an additional inequality constraint that is introduced for $N = 3$, where its upper bound is non-linear, whereas the lower bound is linear. In addition, (3.83d) is to ensure normalized transmission power.

3.7 Results and Discussions

This section presents the exact and asymptotic BER results of a DL NOMA system using various number of users and modulation orders. The BER and asymptotic BER

are computed analytically using the generic expressions in Section 3.5, while the BER is validated by Monte Carlo simulations. In addition, the optimal power assignments that minimize the system's average BER for $N = 2$ and 3 cases are computed, where the PCBs, derived in Section 3.6, are used as constraints to solve the non-linear optimization problem using the interior-point algorithm [48]. It is worth noting that the PCB constraints increase from one constraint for $N = 2$, to three constraints for $N = 3$. Moreover, the small scale fading is considered to be flat and it follows the Rayleigh distribution with $\sigma^2_{\hbar_n} = 1$. The large scale fading is considered as fixed pathloss with an exponent of $\lambda = 2.7$, where the users are at a normalized distance of $\Upsilon_n = 10^{\frac{3}{5\lambda}(n-1)}$ from the base station. The AWGN variance σ_{w_n} is assumed to be common for all users, which corresponds to the transmit SNR $\triangleq 1/2\sigma^2_w$ [34, 35, 43, 48, 95]. The base station and all users are assumed to be equipped with a single antenna unless stated otherwise. The power assignment for $N = 2$ is performed such that (3.78) is satisfied for all the considered modulation orders, where the worst-case scenario is when $\mathbf{m} = [8, 64]$, and thus, $\alpha = [1 \times 10^{-2}, 0.99]$. For $N = 3$, the power assignment should satisfy (3.78) and (3.79) simultaneously for all the considered modulation orders, thus, $\alpha = [1 \times 10^{-4}, 1 \times 10^{-2}, 0.9899]$. The legends in all figures, where $2 \leq N \leq 4$ represent the modulation order vectors $\mathbf{m} = [M_1, M_2, \ldots, M_n]$. Furthermore, for $N > 3$, the power coefficients are selected using linear search such that the constellation points do not overlap and the power coefficients order is maintained.

Figure 3.11 shows the analytical and simulation BER results for $N = 2$, where both users adopt identical modulation orders. The figure shows that the analytical and simulation results match very well for all the considered scenarios. Moreover, the asymptotic BER can be considered as an accurate approximation at high SNR values. As expected, increasing the modulation order degrades the BER performance, and the degradation generally follows the case of orthogonal multiple access. For example, QPSK modulation for both users requires about 3 dB additional power as compared to BPSK to achieve a BER of 10^{-2}. Moreover, when comparing BPSK and 64-QAM, the latter requires about 14.4 and 17.5 dB for U_1 and U_2, respectively, to achieve BERs of 10^{-2}.

Figure 3.12 shows the BER for U_1 and U_2, where the modulation orders are not necessarily identical. It can be noted that at high SNR values, the BER for each user does not generally depend on the modulation order of the other user, except for $\mathbf{m} = [8, 64]$. It is worth noting that the selected power coefficients for this case are close to the PCB

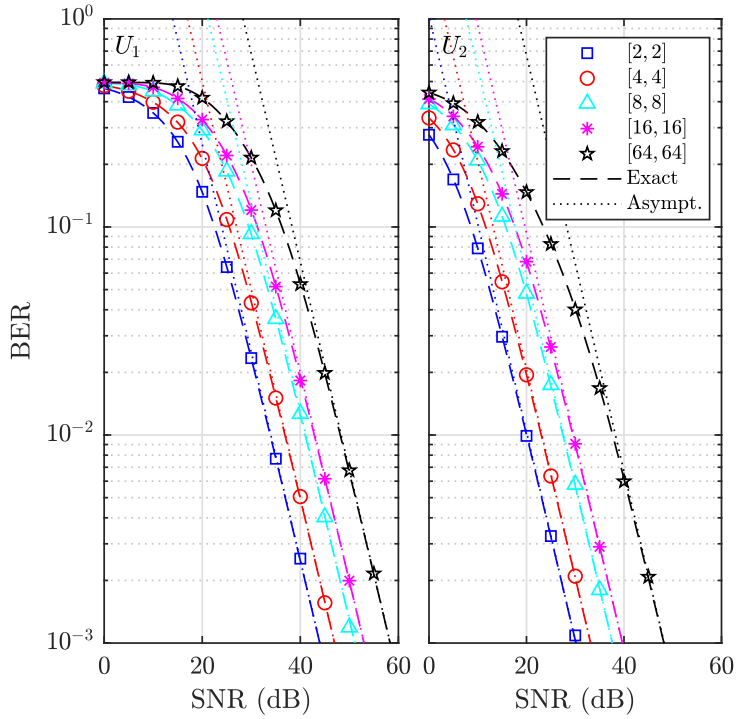


Figure 3.11: Analytical (dashed lines), simulated (markers) and asymptotic BER (dotted lines) results, where $N = 2$ and $M_1 = M_2$.

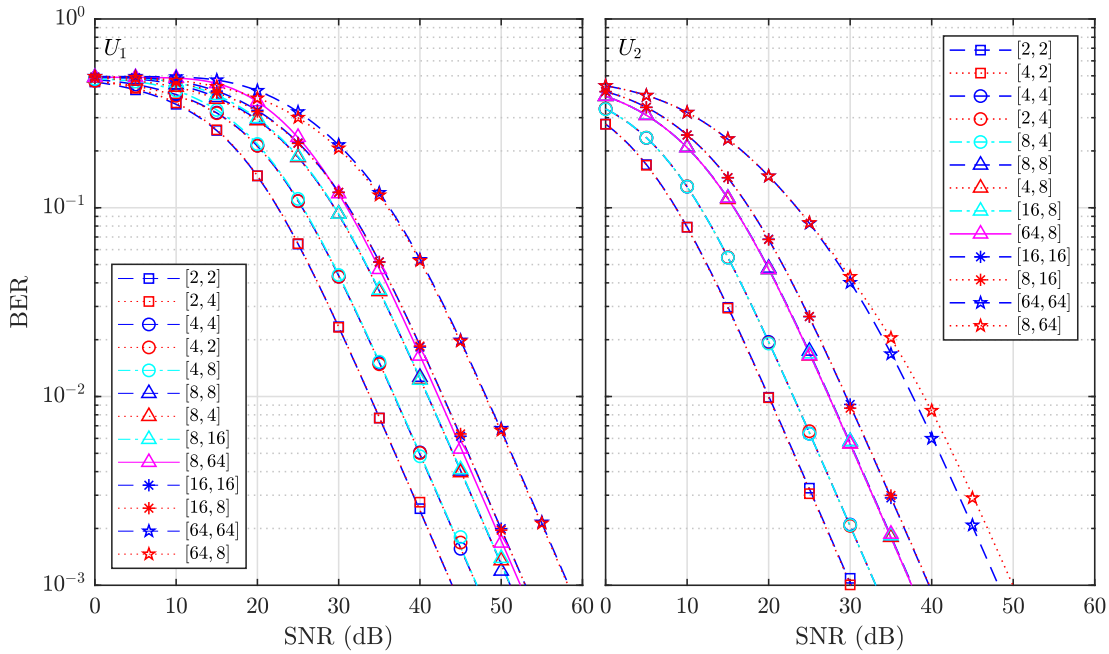


Figure 3.12: U_1 and U_2 BER for various modulation schemes, where $N = 2$.

(3.78), while for the rest of modulation orders combinations, the selected power coefficients are relatively close to the center of the PCB. Therefore, the IUI can cause significant BER variations based on the assigned power for each user. The exact BER change that results by changing the modulation orders is given in Figure 3.13, which is

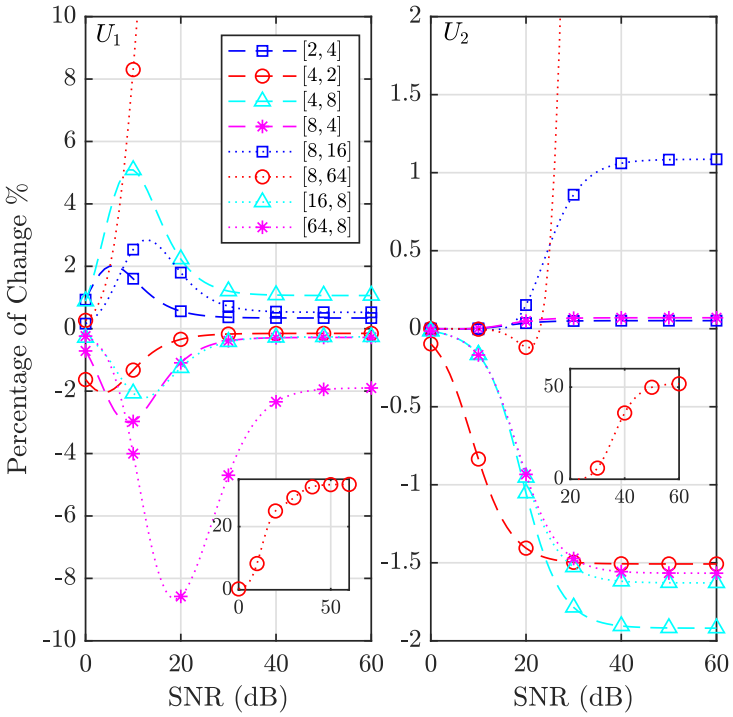


Figure 3.13: BER percentage of change with respect to the identical modulation case for $N = 2$.

computed as the percentage of BER change relative to the identical modulation order case. For example, the BER percentage of change for U_1 with $\mathbf{m} = [8,4]$ is calculated as $(P_{B_1}|\mathbf{m} = [8,4] - P_{B_1}|\mathbf{m} = [8,8])/P_{B_1}|\mathbf{m} = [8,8]$. However, for U_2 with $\mathbf{m} = [8,4]$, the percentage of change is calculated as $(P_{B_2}|\mathbf{m} = [8,4] - P_{B_2}|\mathbf{m} = [4,4])/P_{B_2}|\mathbf{m} = [4,4]$. As can be noted from the figure, the percentage of change converges to a constant value at high SNR values. In addition, the percentage of change for U_1 at high SNR values is within $\pm 2\%$ for all modulation orders, except for $\mathbf{m} = [8,64]$, which saturates at $+30\%$. Similarly, the percentage of change for U_2 is within $\pm 2\%$, except for $\mathbf{m} = [8,64]$, and it saturates at $+50\%$.

Figure 3.14 shows the BER using various modulation orders for $N = 3$. Similar to the $N = 2$ case, the BER increase when higher modulation orders are used. However, the overall BER performance degrades when compared to $N = 2$ because the power budget is shared by three users, and additional interference is introduced by U_3. Due to the equal power assignment for all modulation orders, U_1 is assigned a very low power coefficient, which results in poor BER performance. The power coefficient assigned to U_2 is higher than U_1 by 20 dB, however, the BER advantage is reduced by 6 dB, which corresponds to the large scale fading. The same observations apply to U_3 since it has 20 dB power advantage over U_2. Nonetheless, the power advantage is reduced by the 6 dB relative pathloss difference. The figure also shows the impact of changing the

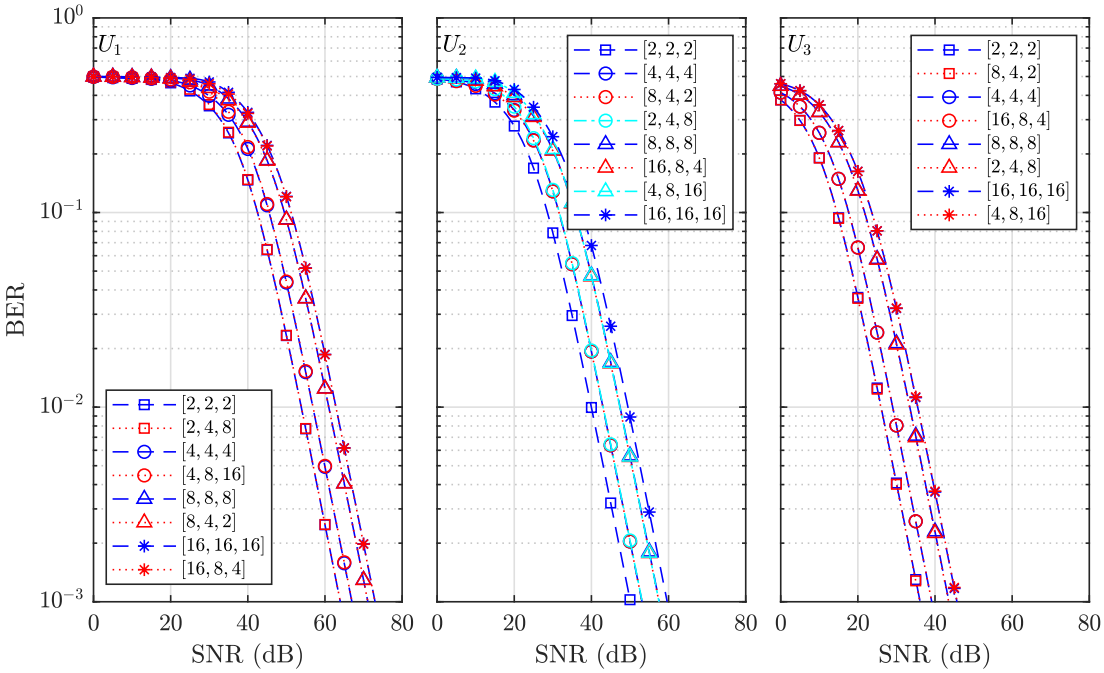


Figure 3.14: BER for various modulation schemes, where $N = 3$.

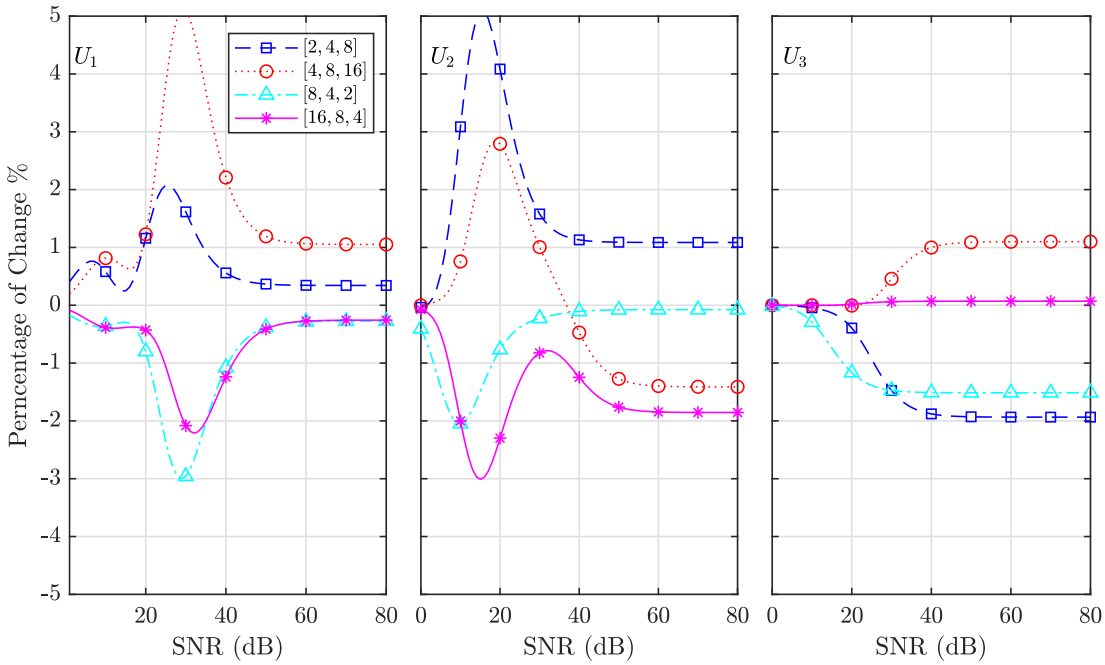


Figure 3.15: BER percentage of change with respect to the identical modulation case for $N = 3$.

modulation order for certain users on the BER of other users. More specifically, it can be noted that changing the modulation orders for any two users will have a negligible effect on the BER of the other users.

Figure 3.15 quantifies the BER variation due to the modulation orders change. As can be noted from the figure, the change at high SNR values is roughly within $\pm 2\%$

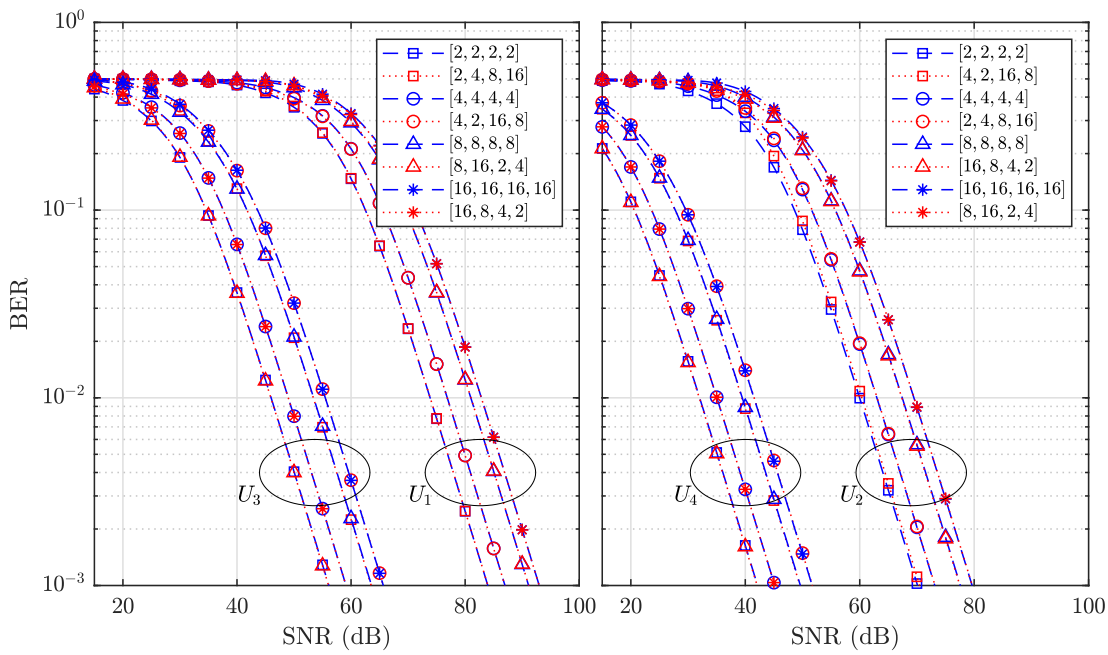


Figure 3.16: BER for various modulation schemes, where $N = 4$.

for all users. At low and moderate SNR values, the change roughly less than $\pm 5\%$.

Figure 3.16 presents the BER for the case of $N = 4$. The power coefficients used are $\alpha_1 = 10^{-6}$, $\alpha_2 = 10^{-4}$, $\alpha_3 = 10^{-2}$ and $\alpha_4 = 0.989899$. Consequently, such power coefficients require extremely accurate power control at the base station. Similar to the $N = 2$ and 3 values, the BER for a particular user is roughly independent of other users' modulation orders. As can be noted from the figure, the small power coefficients, IUI and the large scale fading lead to a degraded BER when compared to cases with a lower number of users. As depicted in Figure 3.17, the impact of changing the modulation orders of other users on a particular user's BER generally follows the other considered cases, where the change of the BER with respect to the identical modulation order is bounded by $\pm 2\%$, at high SNR values for all users and modulation orders, except for $\mathbf{m} = [4, 2, 16, 8]$ for U_2, which saturates in the range of $+10\%$. At low and moderate SNR values, the change is within $\pm 5\%$ for all users and modulation orders, except for $\mathbf{m} = [4, 2, 16, 8]$ for U_2, where it is about $+19\%$.

Figure 3.18 shows the BER for $N = 2, 3, \ldots, 7$, where QPSK modulation is adopted for all users. The power coefficients for each case are given by:

- $N = 2$: $\alpha = [0.138, 0.862]$
- $N = 3$: $\alpha = [2.30 \times 10^{-2}, 0.156, 0.821]$
- $N = 4$: $\alpha = [4.40 \times 10^{-3}, 2.96 \times 10^{-2}, 0.165, 0.801]$

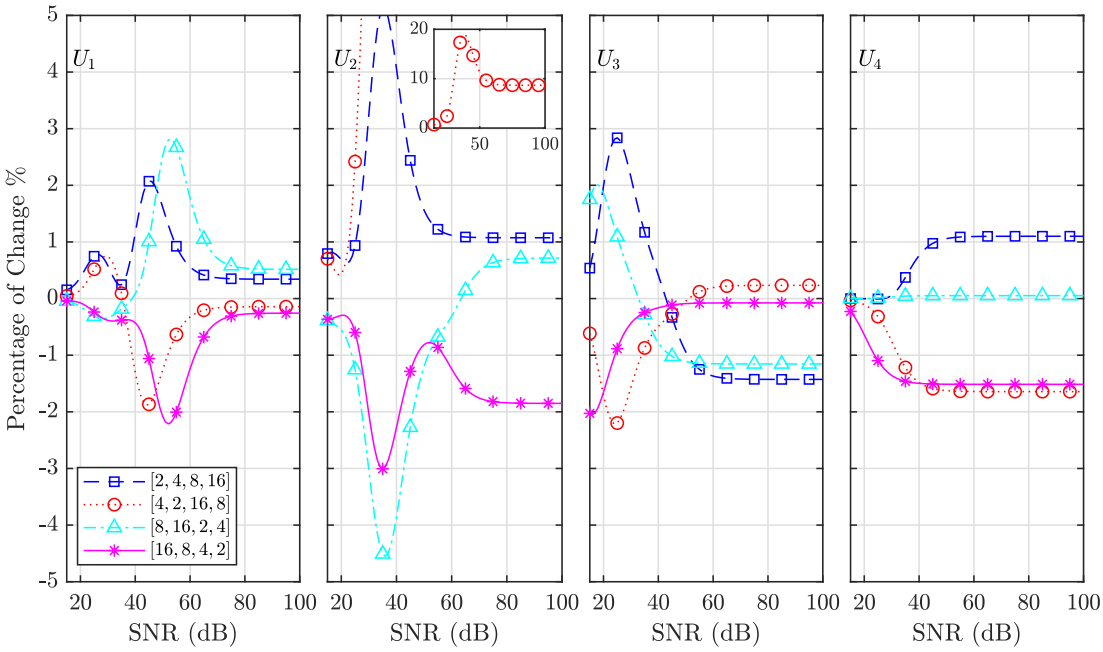


Figure 3.17: BER percentage of change with respect to the identical modulation case for $N = 4$.

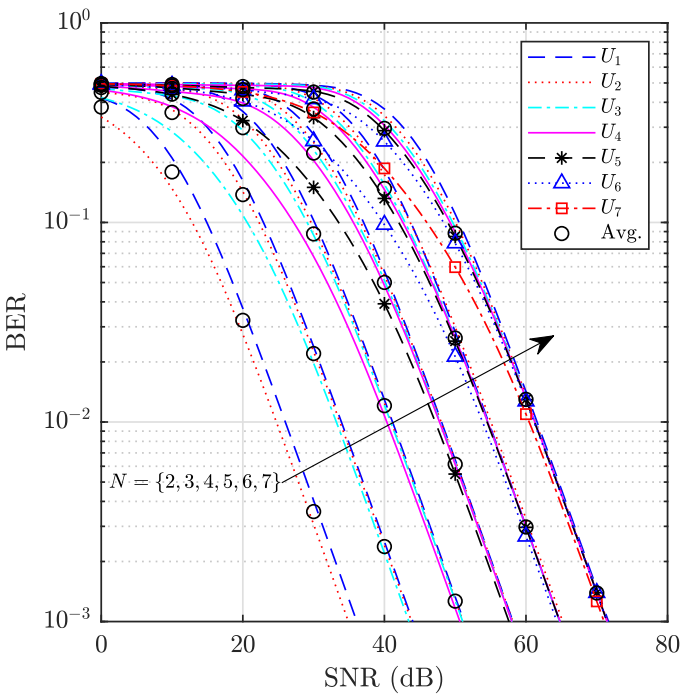


Figure 3.18: BER using optimal power coefficients for various number of users, where $M_n = 4\ \forall n$.

- $N = 5$: $\alpha = [9.00 \times 10^{-4}, 5.90 \times 10^{-3}, 3.33 \times 10^{-2}, 0.170, 0.790]$
- $N = 6$: $\alpha = [2.00 \times 10^{-4}, 1.20 \times 10^{-3}, 7.00 \times 10^{-3}, 3.57 \times 10^{-2}, 0.173, 0.783]$
- $N = 7$: $\alpha = [3.99 \times 10^{-5}, 2.69 \times 10^{-4}, 1.50 \times 10^{-3}, 7.70 \times 10^{-3}, 3.75 \times 10^{-2},$

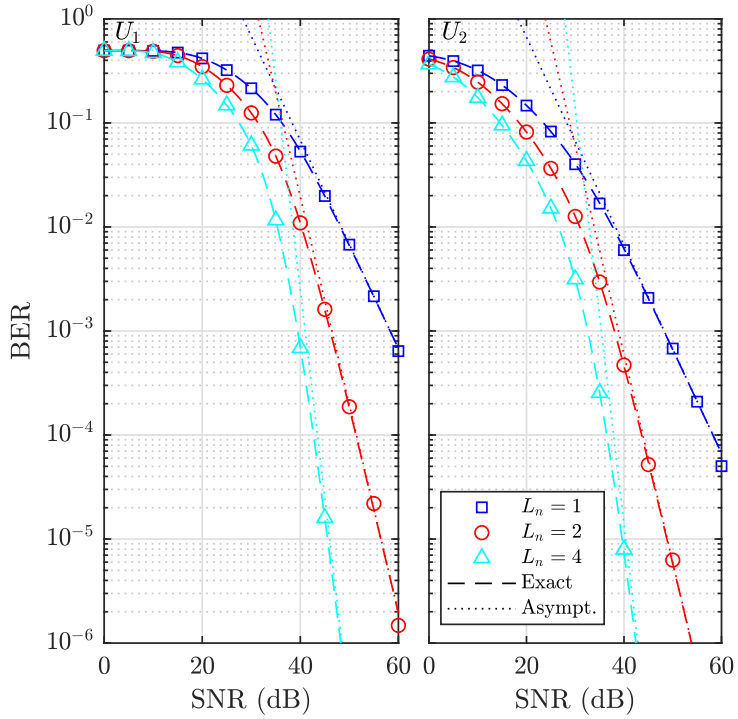


Figure 3.19: Analytical (dashed lines), simulated (markers) and asymptotic BER (dotted lines) results for different L_n, where $N = 2$ and $M_1 = M_2 = 64$.

$$0.175, 0.778].$$

The power coefficients are selected to minimize the system's average BER given that SNR $= 80$ dB. Therefore, the system's average BER can not be considered minimum at low SNR values. Nevertheless, by noting that the power coefficients remain approximately unchanged for a wide range of SNR values, then the presented average BER can be considered near-optimum. Tables A.1 and A.2 present the optimal power coefficients that minimize the system's average BER for various modulation orders and SNR conditions for $N = 2$ and 3, respectively. It can be noted from the selected power coefficients that the difference between α_1 and α_N becomes significant as N increases. For example, the power difference between α_1 and α_7 is about 42.9 dB. Such a high power difference is due to the necessity of compensating the high attenuation caused by the large scale fading. Moreover, it is noted from Figure 3.18 that the BER for all users and values of N approaches the system's average BER at high SNR values. It is also worth noting the trade-off between N and the BER, where adding one additional user to the network causes degradation in the system's average BER between 6.65 and 8.24 dB at BER of 10^{-3}.

Figure 3.19 considers the BER with receiver diversity. The results are obtained for $N = 2$ considering $L_n = 1, 2$ and 4 receiving antennas, and $\mathbf{m} = [64, 64]$. As noted from

the figure, the analytical and simulation results match very well for all the considered scenarios. Moreover, the figure presents the asymptotic BER, which approaches the exact BER high SNR values. As compared to SISO, the diversity gain for $L_n = 2$ is about 12.3 and 10.2 dB for U_1 and U_2, respectively, at BER of 10^{-3}. The gain increases to about 18.9 and 15.6 dB, respectively for U_1 and U_2, for $L_n = 4$.

3.8 Conclusions

This chapter has derived asymptotic and exact analytical BER expressions for NOMA over i.i.d Rayleigh flat fading channels. The derived expressions are applicable for any number of users, where each user has an arbitrary modulation order including BPSK and rectangular/square QAM. The results were corroborated via Monte Carlo simulation results. The derived expressions were used to provide insights about the BER performance in various conditions and system configurations, including some extreme scenarios in terms of the number of users and modulation orders. For example, the BER results revealed that the BER of all users converge to the system's average BER at high SNR values when the optimal power coefficients are adopted. Moreover, when the power coefficients are roughly in the middle of the PCBs, the BER of each user becomes almost independent of the modulation orders of other users, which might be necessary for adaptive modulation. The closed-form PCBs were derived for the $N = 2$ and 3 cases with arbitrary modulation orders. These expressions were utilised as linear and non-linear constraints to compute the optimal power assignment that minimizes the system's average BER. Interestingly, using high modulation orders and large number of users make the power control process very critical, where extremely fine power tuning is required.

Chapter 4

Power Tolerant NOMA Using Data-Aware Power Assignment

4.1 Chapter Introduction

Non-orthogonal multiple access (NOMA) is a promising candidate for future wireless networks due to its ability to improve the spectral-efficiency and network connectivity. Nevertheless, the error rate performance of NOMA depends significantly on the power assignment for each user, which requires accurate knowledge of the channel state information (CSI) at the transmitter, which can be challenging for several applications such as wireless sensor networks (WSNs) and Internet of Things (IoT). Therefore, this chapter proposes a power-tolerant NOMA by adaptively changing the signal power of each user to reduce the system sensitivity to inaccurate power assignments. The power adaptation in the power adaptive non-orthogonal multiple access (PANOMA) is performed based on the transmitted data, and it does not require accurate CSI. To quantify its potential, the bit error rate (BER) and the lower bound capacity performance, over Rayleigh fading channels, are derived in exact closed-forms for two and three users scenarios. The results demonstrate that PANOMA provides a tangible BER performance improvement over conventional power-domain (PD) NOMA when both schemes use optimal and sub-optimal power assignment. The sub-optimal power assignment is typically experienced in practical scenarios involving channel time variation and CSI estimation errors. Hence, robustness to imperfect power assignment is achieved with PANOMA. Specifically, it will be shown that the PANOMA offers BER reduction by a factor of 10 for certain scenarios when sub-optimal power values are assigned for both schemes. Also, it will be shown that for other scenarios PANOMA

provides a BER performance gain besides the robustness when both schemes use optimal power assignment. The integrity of the analytical results is verified via matching extensive Monte Carlo simulation experiments.

4.1.1 Chapter Organization

The rest of the chapter is organised as follows. In Section 4.2, the related work, and motivations and contributions are highlighted. In Section 4.3, the system and the channel models are introduced. Section 4.4 presents the performance analysis of PANOMA where the BER expressions are derived and the constrained capacity analysis is demonstrated. Section 4.5 presents the analytical and Monte Carlo simulation results as well as the optimal power assignments for NOMA and PANOMA systems. Finally, Section 4.6 concludes the chapter with a summary of the main findings.

4.2 Related Work

It is anticipated that the number of devices connected to mobile networks will reach to 25 billion devices by the end of 2020 [84]. The reason for this substantial increase is the unfolding evolution of traditional cities and homes toward becoming smart. This includes applications under the umbrella of the IoT and machine-to-machine communications, which play a central role in many applications such as waste management, smart transportation systems, and smart buildings [83,99–101]. To support these applications, the fifth generation (5G) of wireless networks are designed to provide an agile data transport infrastructure with capabilities that include massive connectivity, larger spectral-efficiency, and low latency [82]. For instance, according to the 5G standard, narrowband-IoT will support the massive connectivity of low-cost-low-capability devices, [102], which are known to have low data-rate requirements akin to WSNs, [103]. The responsibility to satisfy such requirements lies mostly within the multiple access functionality of the communication system, [9]. In the current 5G and previous mobile communication standards, orthogonal multiple access (OMA) techniques are mostly used where the resources are allocated to the users of one cell in such a way to prevent mutual interference.

Meanwhile, recent advances in NOMA, particularly interference cancellation (IC) to separate the users' signals at the receiver. IC techniques include parallel interference cancellation (PIC) and successive interference cancellation (SIC), with the usual

benchmark being the optimum maximum likelihood detector (MLD) [83]. It is shown in [83] that SIC outperforms PIC when each user is allocated a unique power value. Furthermore, PIC can be computationally inefficient as it requires a large number of iterations to provide an acceptable BER, [87]. In [104], it is claimed that SIC can overcome propagation errors using an iterative approach.

As a key performance metric, the BER of NOMA has been considered by several researchers [31, 34, 35, 43, 48, 95, 96]. In [43], an exact closed-form solution is derived for the downlink BER of two-user NOMA in a single-input-single-output (SISO) Rayleigh fading broadcast channel with imperfect SIC. The exact BER for two and three-user downlink glsnoma is derived in [48] for Nakagami-m fading channels using quadrature phase-shift keying (QPSK). In [95], the BER for downlink NOMA with an arbitrary number of users is derived for the Rayleigh fading channel where all users adopt binary phase-shift keying (BPSK). The analysis is extended for the quadrature amplitude modulation (QAM) in [24]. The BER results presented in [24, 31, 34, 35, 43, 48, 95, 96] show that the BER of NOMA systems degrades by inter-user interference (IUI), which can be significant if the users' power allocation is not performed diligently.

Power assignment in NOMA can be generally designed to be fixed or dynamic. Fixed power assignment has low complexity and does not require prior knowledge of the CSI at the transmitter [34,35,95,96,105]. However, it cannot satisfy the users' BER requirements or minimise the average BER. Alternatively, dynamic power assignment has the potential to satisfy quality of service (QoS) requirements and minimise the BER. Dynamic power assignment typically uses the instantaneous or statistical CSI to allocate the power for each user [31,43,48,106]. The assigned power can be considered optimum if it is used to minimise certain performance metrics such as the BER. Because the power assignment is performed at the transmitter side, accurate CSI should be priorly available at the transmitter, which is highly challenging for mobile channels and requires substantial feedback overhead. Adaptive power assignment with in accurate or outdated CSI will not be optimal and may cause severe BER degradation. In the literature, very little work has considered reducing the BER sensitivity to non-optimal power assignment. For example, interference alignment is proposed in [106] to improve the symbol error rate (SER) and support high-reliability low latency IoT systems. Dynamic power allocation is applied using statistical CSI to minimise the average asymptotic SER. Although the system offers SER improvement and resilience

to non-optimal power assignment, the system requires significant computational complexity, the detection process requires pilot symbols to avoid detection ambiguity, and the system is limited only to a two-user NOMA. Robust NOMA design with respect to sum rate and energy efficiency is considered in [107–109]. However, such schemes are not generally applicable to the BER scenario, and they require substantial computational power.

Other approaches for improving the BER of NOMA include applying a particular phase rotation for each user while keeping the power for each user fixed, as well as incorporating signal space diversity through inphase/quadrature interleaving, [97]. The presented results show that a gain of about 1.3 dB can be obtained, but only for one of the users. Furthermore, the achieved gain depends on the time/frequency selectivity of the channel and becomes less significant in flat fading channels. It is worth noting that this approach increases the transmitter and receiver complexity due to the additional operations required to modulate and recover the data symbols. Qiu *et al.* [110] proposed a class of NOMA schemes where all users' signals are mapped into n-dimensional constellations that correspond to the same algebraic lattices to enable every user to achieve full diversity. The presented results show a significant BER improvement at the expense of substantial additional complexity. Such design is based on the assumption that each symbol experiences mutually independent and identically distributed (i.i.d) channels, which is a limiting assumption as the channel coefficients are highly correlated in slow fading channels, unless interleavers with very high depth are used.

4.2.1 Motivation and Contribution

As can be noted from aforementioned discussions, the BER of NOMA is highly sensitive to the users' power assignment, particularly for IoT systems, where accurate CSI is typically difficult to obtain [103]. Moreover, most IoT applications are based on short packets, and thus, frequent CSI feedback may degrade the system efficiency significantly. For example in fourth generation (4G) and 5G networks, the CSI feedback typically requires 22 bits [111]. Consequently, link adaptation in IoT applications should be designed to be robust to inaccurate CSI conditions. Therefore, this paper proposes a power-tolerant data-aware adaptive power assignment technique to improve the BER performance of PD NOMA in scenarios where optimal power allocation cannot be maintained. The overall idea is motivated by the fact that prior knowledge of IUI is available at the transmitter, hence it can be utilised to adapt the power per NOMA

symbol depending on whether IUI is constructive or destructive. It is shown that the proposed data-aware PANOMA scheme maintains the signal to noise ratio (SNR) at the receivers almost unchanged, and hence, the BER and capacity experience negligible degradation as compared to NOMA.

The contributions of this chapter can be summarized as follows:

1. Propose a data-aware power assignment for PD NOMA systems called PANOMA to improve the system robustness to imperfect power assignments.

2. Evaluate the impact of non-optimal power assignment on the NOMA and PANOMA systems.

3. Extend the proposed system to QPSK modulation and determine the BER performance gain. Interestingly, it is shown that the inphase and quadrature components for each user should be assigned different power to minimise the BER.

4. Derive and evaluate the overall lower bound capacity of both systems and quantifying the PANOMA gain for BPSK and QPSK.

5. Find the optimal power assignment that minimises the system's average BER and maximises the system's constrained capacity for both systems.

4.3 System and Channel Models

4.3.1 Conventional Power-Domain NOMA

As depicted in Figure 4.1, this work considers the downlink of an IoT network where a control center communicates with a large number of IoT devices, also denoted as users, which are mostly actuators. To increase the system spectral efficiency, NOMA is deployed at the base station where each group of N IoT devices are configured to share the same transmission resources simultaneously. Therefore, the base station multiplexes the information symbols of N devices by assigning each user a distinct power coefficient based on the channel conditions. Without loss of generality, we assume that the N users are ordered in ascending order based on their average channel gain, i.e. $\mathbb{E}[|h_1|^2] > \mathbb{E}[|h_2|^2] > \cdots > \mathbb{E}[|h_N|^2]$, where h_n, $1 \leq n \leq N$, is the complex channel gain of the link between the base station and the nth user, which is denoted as U_n. Therefore, the power allocation is performed such that a user with severe fading conditions is assigned higher power than a user with good channel conditions [43, 48].

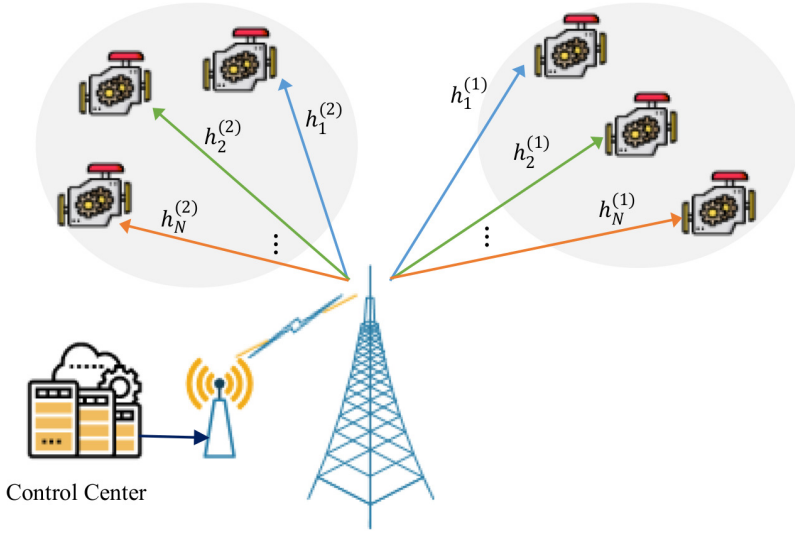


Figure 4.1: Illustrative diagram of a NOMA-IoT system where users are clustered in groups of three users and served by one NOMA signal.

In such scenarios, the power coefficients $\{\alpha_1, \alpha_2, \ldots, \alpha_N\}$ are assigned such that $\alpha_1 < \alpha_2 < \ldots < \alpha_N$, where $\sum_{n=1}^{N} \alpha_n = 1$. The superposition coding (SC) process at the base station is described by

$$x_{SC} = \sum_{n=1}^{N} \sqrt{\alpha_n} x_n \tag{4.1}$$

where x_n is the information symbol of the nth user drawn uniformly from an M-ary phase shift keying (M-PSK) constellation χ_n. The average power of the transmitted NOMA symbols is normalized to unity such that $\mathbb{E}[|x_{SC}|^2] = 1$, where x_{SC} is the transmitted NOMA symbol. Figure 4.2 shows the constellation diagram of the resultant superposition NOMA symbol for $N = 2$ and 3 using BPSK modulation. For $N = 2$, the NOMA constellation comprises 4 constellation points, with four different amplitudes. Moreover, it can be noted that the leftmost bit belongs to U_1, while the rightmost bit belongs to U_2. The same argument applies for cases where $N > 2$, and when the modulation orders are more than 2.

At the receiver front-end, the received baseband signal in flat fading channels is written as

$$y_n = h_n x_{SC} + w_n \tag{4.2}$$

where $w_n \sim \mathcal{CN}(0, N_0)$ is the additive white Gaussian noise (AWGN). In channels with small-scale Rayleigh fading and large scale pathloss, the channel gain $h_n = \sqrt{\beta_n} \times \hbar_n$,

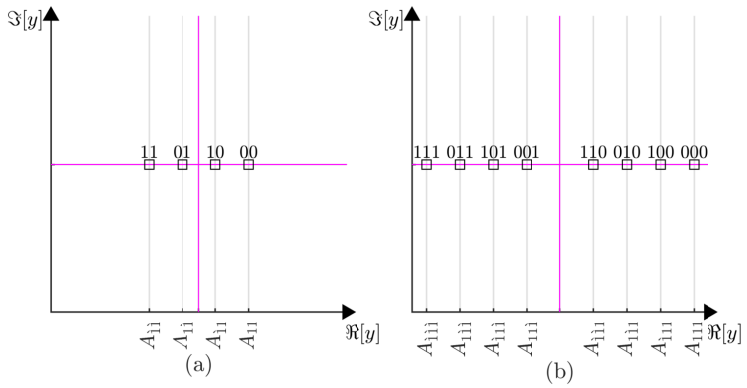


Figure 4.2: Constellation diagram of the transmitted signal when BPSK is used by all users: (a) $N = 2$. (b) $N = 3$.

$\hbar_n \sim \mathcal{CN}(0, \sigma^2_{\hbar_n})$, $\beta_n = \Upsilon_n^{-\lambda}$, Υ_n is the distance between the base station and U_n, λ is the pathloss exponent. The small-scale fading of different users is considered i.i.d [97, 110]. Given that CSI is known perfectly at the receiver, the information symbols can be recovered using the joint-multiuser maximum likelihood detector (JMLD),

$$\{\widehat{x}_1, \widehat{x}_2, ..., \widehat{x}_N\} = \arg\min_{x_i \in \chi_i} \left| y_n - h_n \sum_{i=1}^{N} \sqrt{\alpha_i} x_i \right|^2 \tag{4.3}$$

where $\{\widehat{x}_1, \widehat{x}_2, ..., \widehat{x}_N\}$ are the jointly detected N users' symbols. Alternatively, the SIC detector can be applied where the symbols of the user with maximum power U_N is detected first using a single-user MLD while considering all other users' signals as unknown noise. For the user with the second highest power, U_{N-1}, it has to detect and subtract the symbols of U_N, and then apply a single-user MLD to detect its own symbols. The minimum power user U_1 has to detect and subtract the symbols of $N-1$ users before it can detect its own symbols. It is worth noting that the SIC and JMLD detectors have identical BER performance, but the SIC has lower complexity [35].

The power assignment for each user may have significant effect on the performance of all users, and inappropriate power assignment can therefore result in noticeable performance degradation. Therefore, extensive research has been focused on optimal power assignment as reported in [31, 43, 48], and the references listed therein. For the BER case, the power optimization problem is nonlinear hence it is typically solved using certain searching methods, which achieve near-optimal solutions. Moreover, optimal power assignment requires perfect knowledge of the channel statistical information [43, 48], which is typically very challenging to obtain, particularly in time-varying channels.

4.3.2 The PANOMA Scheme

4.3.2.1 PANOMA for BPSK

To simplify the presentation of the proposed PANOMA scheme, we consider the two users scenario using BPSK modulation. As can be noted from (4.1) and Figure 4.2a, the constellation points of the two rightmost NOMA symbols are $A_{11} = (\sqrt{\alpha_1} + \sqrt{\alpha_2})$ and $A_{\grave{1}1} = (-\sqrt{\alpha_1} + \sqrt{\alpha_2})$. Therefore, the error event with respect to U_1 is largely determined by the Euclidean distance $D_1^C = \left|A_{11} - A_{\grave{1}1}\right| = 2\sqrt{\alpha_1}$. Given that the value of α_1 is selected to minimise the BER, the value of D_1^C will be maximised. However, a small deviation in α_1 from the optimal value will reduce D_1^C from its maximum value and cause BER degradation.

Inspired by the interference exploitation principle, [112, 113], the power assignment is performed adaptively based on the values of the transmitted symbols of both users. Therefore, the power is assigned based on the user index and values of the information symbols of the two users. Consequently, the transmitted PANOMA symbol is written in a conditional form to capture the dependency of the power on the information symbols. Therefore,

$$x_{SC} = \sum_{n=1}^{N} \sqrt{(\alpha_n | [x_1, x_2, \ldots, x_N])} x_n. \tag{4.4}$$

To reduce the sensitivity of the BER to power assignment process, the power for certain symbols is fixed regardless of the channel conditions, while the power for the remaining symbols can be selected to minimise the average BER. For example, given that $N = 2$ and BPSK is used for both users, the power assignment for the identical symbols cases, i.e. constructive interference, can be performed such that $\alpha_1|[1,1] = \alpha_1|[-1,-1] = 0.5$ which is always fixed. On the other hand, when the symbols are with opposite polarities, i.e. destructive interference, the power assignment for $\alpha_1|[1,-1]$ and $\alpha_1|[-1,1]$ can be selected to minimise the average BER using the formula derived in Section 3.4. Therefore, the transmitted NOMA symbol can be expressed as,

$$x_{SC} = \begin{cases} \sqrt{\frac{1}{2}} x_1 + \sqrt{\frac{1}{2}} x_2, & x_1 = x_2 \\ \sqrt{\alpha_1} x_1 + \sqrt{\alpha_2} x_2, & x_1 = -x_2 \end{cases}. \tag{4.5}$$

It is worth noting that the equal power selection $\alpha_1 = \alpha_2 = 1/2$ for the case of $x_1 = x_2$ maximises the NOMA symbol power for the identical symbols case, which can be

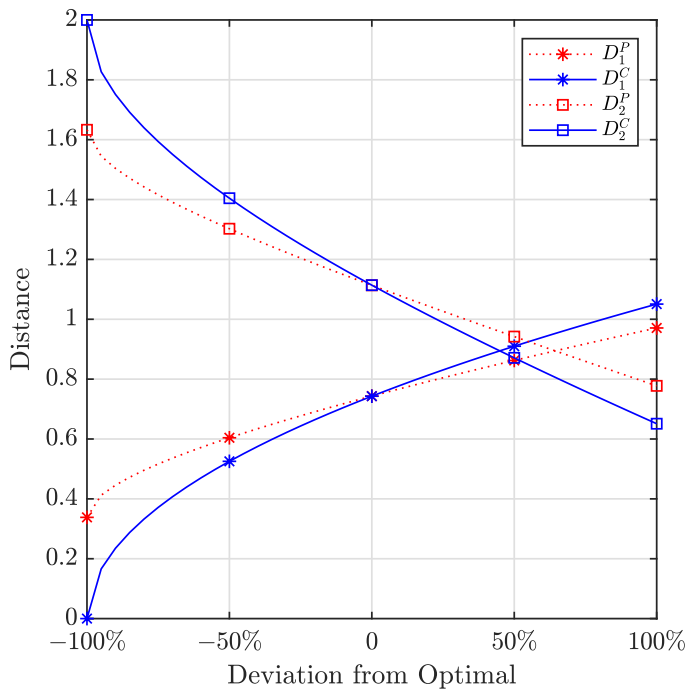


Figure 4.3: Euclidean distance versus deviation from optimal power at $E_b/N_0 = 35$ dB.

verified by computing

$$\frac{\partial}{\partial \alpha_1} x_{sc}^2 \big| [x_1 = x_2] = \frac{1 - 2\alpha_1}{\sqrt{\alpha_1 (1 - \alpha_1)}}, \tag{4.6}$$

which is maximum when $\alpha_1 = 1/2$. Similar to the conventional NOMA case, the average power of the PANOMA symbols should be normalized to unity. By comparing the Euclidean distance of the rightmost PANOMA symbols $D_1^P = \left|A_{11} - A_{\grave{1}1}\right| = \sqrt{2} + \sqrt{\alpha_1} - \sqrt{\alpha_2}$ to $D_1^C = 2\sqrt{\alpha_1}$, it can be noted that the constant scalar in D_1^P limits its variation with α_1. Figure 4.3 shows the distance D_1 and D_2 for NOMA and PANOMA versus the deviation from the optimal power, where $D_2 = \left|A_{\grave{1}1} - A_{1\grave{1}}\right|$. As can be noted from the figure, D_1 and D_2 for PANOMA change at a lower rate than NOMA, implying that they remain close to the optimum values.

Following the same approach for $N = 2$, PANOMA can be applied to $N = 3$ where the NOMA symbol can be expressed as

$$x_{SC} = \begin{cases} \sqrt{\frac{1}{3}}x_1 + \sqrt{\frac{1}{3}}x_2 + \sqrt{\frac{1}{3}}x_3, & x_1 = x_2 = x_3 \\ \sqrt{0.05x_1} + \sqrt{0.475}x_2 + \sqrt{0.475}x_3, & x_1 \neq x_2 = x_3 \\ \sqrt{\alpha_1}x_1 + \sqrt{\alpha_2}x_2 + \sqrt{\alpha_3}x_3, & x_1 = x_3 \neq x_2 \\ \sqrt{0.075x_1} + \sqrt{0.075}x_2 + \sqrt{0.85}x_3, & x_1 = x_2 \neq x_3 \end{cases}. \tag{4.7}$$

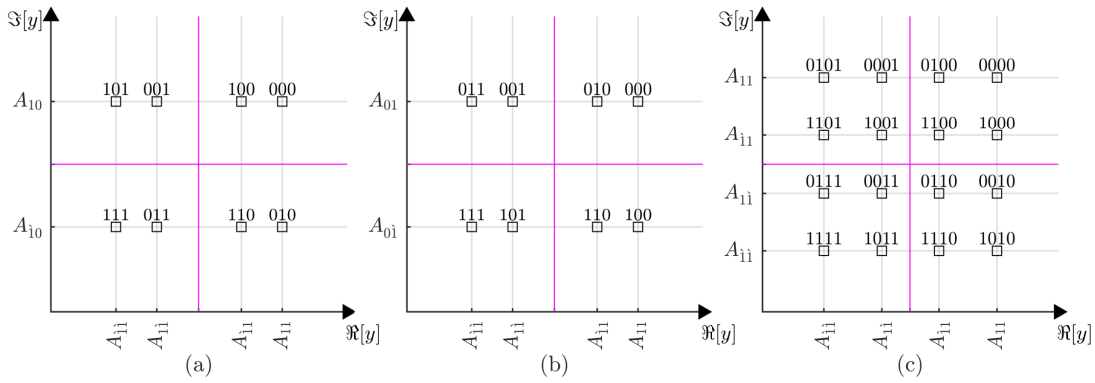


Figure 4.4: Constellation diagram of the transmitted signal when $N = 2$ and the modulation orders are: (a) $\mathbf{m} = [4,2]$. (b) $\mathbf{m} = [2,4]$. (c) $\mathbf{m} = [4,4]$.

Note that $1/3$ maximises the symbol power for the case where $x_1 = x_2 = x_3$ whereas the other fixed power coefficients are found via brute force.

As can be noted from (4.4) and (4.5), the data symbols of U_2 can be detected similar to conventional NOMA, i.e., using single-user MLD while considering all other users' symbols as unknown noise. For U_1, the SIC detector cannot be applied directly because the power assigned to U_1 might be either 0.5 or α_1. Therefore, the JMLD shown in (4.3) can be applied by considering the SC process in (4.4), or equivalently, a modified SIC detector can be applied. For example, when BPSK with $N = 2$ is used, as shown in Figure 4.2a, the interference can be cancelled by computing $\tilde{y} = y_1 - \frac{1}{2}h_1(\sqrt{2}+\sqrt{\alpha_2}-\sqrt{\alpha_1})\hat{x}_2$. Then, a single-user MLD can be used to extract the data symbols of U_1. The same approach can be applied for the case where $N > 2$ and using higher order modulations. Therefore, the proposed system complexity is identical to conventional NOMA. It is worth noting that the systems proposed in [97, 106], require pilot sequences to perform the detection process, which affects the spectral efficiency. Moreover, the receiver complexity is higher than conventional NOMA.

4.3.2.2 PANOMA for QPSK

The PANOMA principle can be extended to higher modulation orders such as QPSK where the inphase and quadrature components are assigned different power factors by treating each dimension independently. Therefore,

$$x_{SC} = \sum_{n=1}^{N} \sqrt{\alpha_n^{\mathrm{Re}}|\mathrm{Re}\,[x_1, x_2, \ldots, x_N]}\mathrm{Re}[x_n] + 1j\sqrt{\alpha_n^{\mathrm{Im}}|\mathrm{Im}\,[x_1, x_2, \ldots, x_N]}\mathrm{Im}[x_n] \quad (4.8)$$

where α_n^{Re} and α_n^{Im} are the power coefficients for the inphase and quadrature components, $\mathrm{Re}[\cdot]$ and $\mathrm{Im}[\cdot]$ represent the real and imaginary parts, respectively. Figure 4.4 shows the constellation diagrams for different modulation orders when $N = 2$.

To exemplify (4.8) and distinguish between the PANOMA and the conventional NOMA systems, the case of $N = 2$, $\mathbf{m} = [2,4]$ is considered where the constellation diagram is shown in Figure 4.4a. This example is interesting as the near user's symbol has a real part only whereas the far user's symbol is complex. Therefore, the NOMA symbol in such case can take the following general form

$$x_{SC} = \sqrt{(\alpha_1 | [x_1, x_2])} x_1 + \sqrt{\left(\alpha_2^{\mathrm{Re}} | [x_1, x_2]\right)} \mathrm{Re}[x_2] + 1j\sqrt{\left(\alpha_2^{\mathrm{Im}} | [x_1, x_2]\right)} \mathrm{Im}[x_2] \quad (4.9)$$

where for the conventional NOMA system, $\alpha_2^{\mathrm{Re}} = \alpha_2^{\mathrm{Im}}$ and are fixed for all information symbols. However, in PANOMA system the assignment is not necessarily equal. To find the power coefficients for the constructive interference case, the following proposition is made.

Proposition 1 *The power coefficients that maximise the constructive interference are* $\alpha_1 = 2/3$, $\alpha_2^{\mathrm{Re}} = 1/3$.

Proof 1 *Considering the constructive interference case when the transmitted NOMA word is* $\mathcal{X} = 111$, *the real part of* x_{SC}, $\mathrm{Re}[x_{SC}]$, *is*

$$\mathrm{Re}[x_{SC}] = \sqrt{\alpha_1} + \sqrt{\alpha_2^{\mathrm{Re}}} \frac{1}{\sqrt{2}} \quad (4.10)$$

where $\alpha_2^{\mathrm{Re}} = 1 - \alpha_1$. *The maxima of* (4.10) *can be found by considering the first derivative,*

$$\frac{\partial}{\partial \alpha_1} \mathrm{Re}[x_{SC}] = \frac{1}{2} \frac{1}{\sqrt{\alpha_1}} - \frac{1}{2\sqrt{2}} \frac{1}{\sqrt{1-\alpha_1}} \quad (4.11)$$

where the maxima solution is $\alpha_1 = 2/3$ *and* $\alpha_2^{\mathrm{Re}} = 1/3$. ■

Therefore, the NOMA symbol in PANOMA system can be expressed as

$$x_{SC} = \begin{cases} \sqrt{\frac{2}{3}} x_1 + \sqrt{\frac{1}{3}} \mathrm{Re}[x_2] + 1j\sqrt{\alpha_2^{\mathrm{Im}}} \mathrm{Im}[x_2], & \mathrm{sign}(x_1) = \mathrm{sign}(\mathrm{Re}[x_2]) \\ \sqrt{\alpha_1} x_1 + \sqrt{\alpha_2^{\mathrm{Re}}} \mathrm{Re}[x_2] + 1j\sqrt{\alpha_2^{\mathrm{Im}}} \mathrm{Im}[x_2], & \mathrm{sign}(x_1) \neq \mathrm{sign}(\mathrm{Re}[x_2]) \end{cases} \quad (4.12)$$

where α_2^{Re} and α_2^{Im} can be assumed equal for the case of $\mathrm{sign}(x_1) \neq \mathrm{sign}(\mathrm{Re}[x_2])$ even

though optimizing them independently might improve the overall performance. However, this introduces an extra overhead. In addition, $2/3$ factor is found that it maximises the real part of the NOMA symbol in the case of $\text{sign}(x_1) = \text{sign}(\text{Re}[x_2])$ as seen in the previous proof. In addition, the power assignment is different from one symbol to another. Therefore, the expression of $\alpha_{1,\max}^{(N=2)}$ does not apply here.

It is important to mention that the expression in (4.12) can be directly translated to the $\mathbf{m} = [4,2]$ case where its constellation diagram is shown in Figure 4.4b . Generally, the NOMA symbol for $\mathbf{m} = [4,2]$ can be written as

$$x_{SC} = \sqrt{\left(\alpha_1^{\text{Re}}|\,[x_1,x_2]\right)}\text{Re}[x_1] + 1j\sqrt{\left(\alpha_1^{\text{Im}}|\,[x_1,x_2]\right)}\text{Im}[x_1] + \sqrt{(\alpha_2|\,[x_1,x_2])}x_2. \quad (4.13)$$

By following the same approach as in $\mathbf{m} = [2,4]$ case, the NOMA symbol for the PANOMA system is given as follows,

$$x_{SC} = \begin{cases} \sqrt{\frac{1}{3}}\text{Re}[x_1] + \sqrt{\frac{2}{3}}x_2 + 1j\sqrt{\alpha_1^{\text{Im}}}\text{Im}[x_1], & \text{sign}(\text{Re}[x_1]) = \text{sign}(x_2) \\ \sqrt{\alpha_1}x_1 + \sqrt{\alpha_2^{\text{Re}}}\text{Re}[x_2] + 1j\sqrt{\alpha_2^{\text{Im}}}\text{Im}[x_2], & \text{sign}(\text{Re}[x_1]) \neq \text{sign}(x_2) \end{cases}. \quad (4.14)$$

Similarly, the NOMA symbol for $\mathbf{m} = [4,4]$ case can be written as

$$\begin{aligned} x_{SC} = &\sqrt{\left(\alpha_1^{\text{Re}}|\,[x_1,x_2]\right)}\text{Re}[x_1] + \sqrt{\left(\alpha_2^{\text{Re}}|\,[x_1,x_2]\right)}\text{Re}[x_2] \\ &+ 1j\sqrt{\left(\alpha_1^{\text{Im}}|\,[x_1,x_2]\right)}\text{Im}[x_1] + 1j\sqrt{\left(\alpha_2^{\text{Im}}|\,[x_1,x_2]\right)}\text{Im}[x_2] \end{aligned} \quad (4.15)$$

where its constellation diagram is shown in Figure 4.4c. By following the same approach used for $\mathbf{m} = [2,2]$ case, the PANOMA symbol can be written as

$$x_{SC} = \begin{cases} \sqrt{\frac{1}{2}}x_1 + \sqrt{\frac{1}{2}}x_2, & x_1 = x_2 \\ \sqrt{\frac{1}{2}}\text{Re}[x_1+x_2] + 1j\sqrt{\alpha_1^{\text{Im}}}\text{Im}[x_1] + 1j\sqrt{\alpha_2^{\text{Im}}}\text{Im}[x_2], & \text{Re}[x_1] = \text{Re}[x_2] \\ \sqrt{\alpha_1^{\text{Re}}}\text{Re}[x_1] + \sqrt{\alpha_2^{\text{Re}}}\text{Re}[x_2] + 1j\sqrt{\frac{1}{2}}\text{Im}[x_1+x_2], & \text{Im}[x_1] = \text{Im}[x_2] \\ \sqrt{\alpha_1}x_1 + \sqrt{\alpha_2}x_2, & x_1 = -x_2 \end{cases} \quad (4.16)$$

It can be seen that the inphase and quadrature phase power coefficients are assigned independently and the selected power coefficient is identical to the $\mathbf{m} = [2,2]$ case. Therefore, the PANOMA symbol for $\mathbf{m} = [4,4,4]$ case can be found following the same approach.

4.4 Performance Analysis of PANOMA

4.4.1 BER Analysis

The BER expressions of the PANOMA system is different from the BER expressions of the conventional NOMA system derived in Chapter 3. This is due to the unique power assignment of the PANOMA. However, the BER expressions general form derived in Chapter 3 still hold because the constellation diagrams and the decision region boundaries are similar. Therefore, the conditional BER of the PANOMA system can have the following general form

$$P_{B_n}|\gamma_n = \sum_i c_i Q\left(\sqrt{\mathcal{A}_{i,n}}\right) \tag{4.17}$$

where c_i and $\mathcal{A}_{i,n}$ definitions can be found in Tables A.1 and A.2 in Appendix A. It is worth noting that the value of $\mathcal{A}_{i,n}$ depend mainly on the symbols' amplitudes in the NOMA constellation diagram. The actual value of $\mathcal{A}_{i,n}$ for PANOMA is different because of the unique power assignment. To illustrate this difference, the following example is considered. Let the adopted modulation orders be $\mathbf{m} = [2,2]$. Therefore, $A_{11} = \frac{\sqrt{2}}{\theta}$ and $A_{\grave{1}1} = \frac{-\sqrt{\alpha_1}+\sqrt{\alpha_2}}{\theta}$, where θ is a normalization factor to ensure that $\mathbb{E}[|x_{sc}|^2] = 1$. For this specific example it is given as $\theta = \frac{2+(-\sqrt{\alpha_1}+\sqrt{\alpha_2})^2}{2}$.

4.4.2 Constrained Capacity Analysis

The channel capacity for the nth user is defined as maximum achievable mutual information $I(x_n, y_n)$ between the channel input x_n and output y_n when computed over all possible input probability distributions $p(x_n)$ [114, pp. 183],

$$C_n = \max_{p(x_n)} I\left(x_n; y_n\right). \tag{4.18}$$

The mutual information can be expressed as

$$I\left(x_n; y_n\right) = H\left(x_n\right) - H\left(x_n \mid y_n\right) \tag{4.19}$$

where $H(x_n)$ is the entropy of the source and $H\left(x_n \mid y_n\right)$ is the conditional entropy. For BPSK modulation, the channel between the data source at the base station and detector output of the nth user can be considered as a binary symmetric channel (BSC). Therefore, (4.19) is maximised when the source binary data symbols are equally likely,

which implies that $H(x_n) = 1$ and

$$H(x_n \mid y_n) \triangleq \Omega(P_{B_n}) = \acute{P}_{B_n} \log_2\left(\frac{1}{\acute{P}_{B_n}}\right) + P_{B_n} \log_2\left(\frac{1}{P_{B_n}}\right) \tag{4.20}$$

where $\acute{P}_{B_n} = 1 - P_{B_n}$. Therefore,

$$C_n = 1 - \Omega(P_{B_n}) \tag{4.21}$$

and thus, the system capacity is given as $C = \sum_{n=1}^{N} C_n$, which is considered a lower bound to the channel symmetric capacity [115, 116]. For higher modulation orders the channel is no longer binary which makes it more challenging to obtain the conditional entropy. It is worth noting that at extremely high SNR values, the BER tends to be extremely low. Therefore, $\Omega(P_{B_n}) \approx 0$ and $C_n \approx 1$. Consequently, $C \approx N$.

4.5 Results and Discussions

This section presents the analytical and simulation results for the BER and capacity of the downlink conventional NOMA and proposed PANOMA systems. For more clarity, the conventional NOMA will be denoted as C-NOMA. In addition, this section demonstrates the optimal power assignment to: 1) Minimize the system's average BER. 2) Maximize the system's constrained capacity. The results are presented for $N = 2$ and $N = 3$ cases, using BPSK and gray-coded QPSK modulation. The performance of a single-user OMA with maximum transmit power and located at a distance Υ_1 from the base station is used as a benchmark and it is denoted as SU. The BER is computed analytically and validated by Monte Carlo simulation. The capacity is computed analytically for BPSK case while it is validated with Monte Carlo simulation for other cases. In each simulation run, 3×10^7 independent bits are generated with equal probability. The simulations were conducted using a work station equipped with 2.5 GHz 64 bits quad-core Intel Core i7 processor and 16 GB RAM. Moreover, the MATLAB Parallel Computing Toolbox™ was configured to perform parallel computing using four cores. The small-scale fading is considered to be flat and the channel fading coefficients for each realization are modeled as i.i.d Rayleigh random variables with $\sigma^2_{\hbar_n} = 1$. The large-scale fading is considered as fixed pathloss with an exponent of $\lambda = 2.7$. The users' distances from the base station are respectively given by Υ_1, $\Upsilon_2 = 1.67\Upsilon_1$ and $\Upsilon_3 = 2.78\Upsilon_1$. Consequently, the pathloss coefficients are $\beta_1 = 0$ dB,

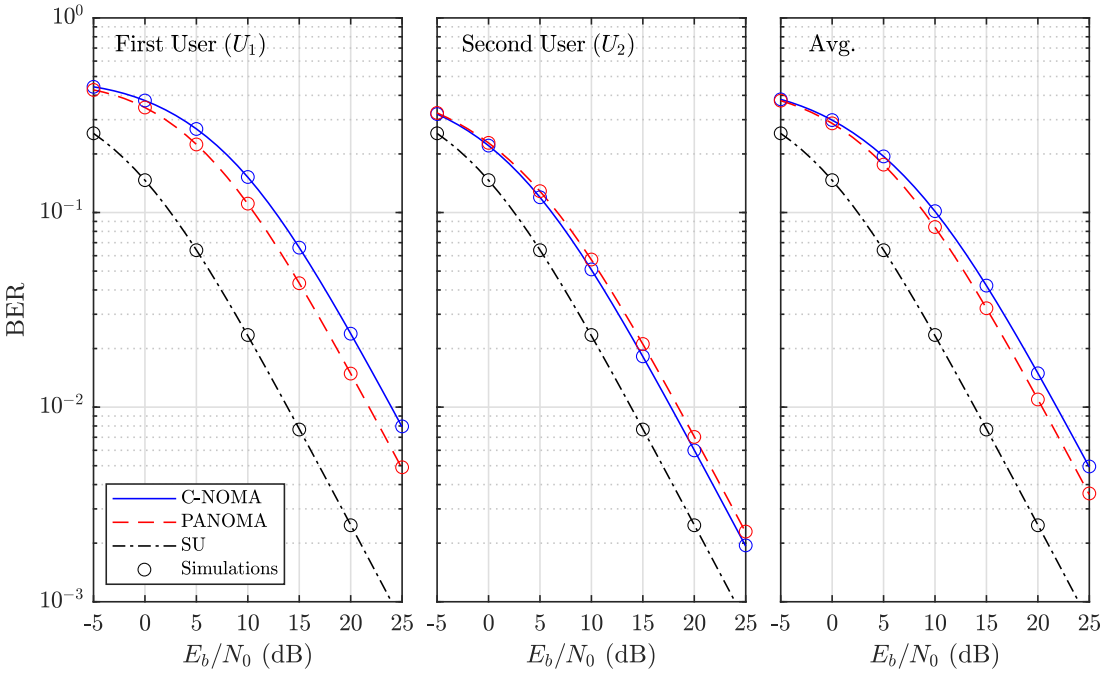


Figure 4.5: Analytical and simulation BER results of C-NOMA and PANOMA systems for $\mathbf{m} = [2,2]$ and $\mathbf{m} = [4,4]$.

$\beta_2 = 6$ dB and $\beta_3 = 12$ dB. Hence, U_1 can be considered as the near user, while U_2 and U_3 are the middle and far users for $N = 3$. For $N = 2$, U_2 is the far user. The base station and all users are assumed to be equipped with a single antenna. Unless it is mentioned otherwise, all the results for the $N = 2$ case are obtained using fixed power assignment with factors of $\alpha_1 = 0.05$ and $\alpha_2 = 0.95$, where the PANOMA power is assigned as described in (4.5). For the $N = 3$ case, two fixed power assignment patterns are adopted which are $\mathcal{P}_1 = [0.05, 0.15, 0.80]$ and $\mathcal{P}_2 = [0.1, 0.2, 0.7]$, where the first value in each pattern corresponds to α_1, while the second and third values correspond to α_2 and α_3. It is worth noting that the bit energy per noise spectral density is defined as $\frac{E_b}{N_0} \triangleq \frac{1}{2\sigma_w^2 \mathcal{M}}$ where $\mathcal{M} = \sum_n \log_2 M_n$.

4.5.1 BER Performance: Robustness

Figure 4.5 shows the BER of the C-NOMA and PANOMA systems using with $\mathbf{m} = [2,2]$ and $\mathbf{m} = [4,4]$. Although the power assignment for both systems is not optimal, it can be noted that PANOMA outperforms C-NOMA by 2.18 dB for U_1, while a degradation of about 0.7 dB is obtained for U_2. However, the average BER of both users, $(P_{B_1} + P_{B_2})/2$, which is typically considered as the key metric [43, 48, 117], improves by about 1.41 dB. Such degradation is justified by the fact that the power assignment process in the PANOMA is designed to minimise the average BER, not

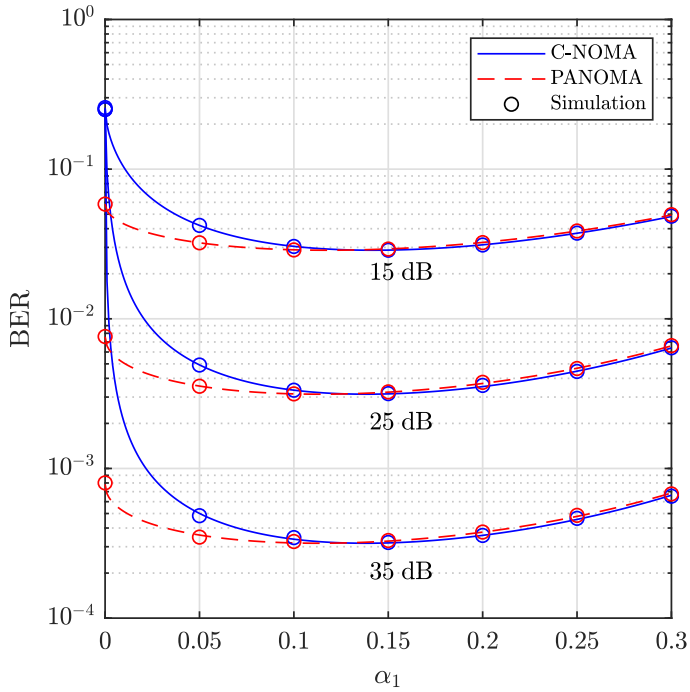


Figure 4.6: Analytical and simulated average BER for C-NOMA and PANOMA versus α_1 at different E_b/N_0 values where $\mathbf{m} = [2,2]$ and $\mathbf{m} = [4,4]$.

the BER of the individual users. Furthermore, NOMA system has an inferior BER performance compared to OMA scheme because of the IUI. Also, it is important to mention that the performance of $\mathbf{m} = [2,2]$ and $\mathbf{m} = [4,4]$ are identical in terms of E_b/N_0 where this follows the common finding for single-user in the literature. It can be noted from Figure 4.5 that the analytical and simulation results match very well for considered scenarios. The same observation is also applicable to the rest figures.

Figure 4.6 shows the average BER versus α_1 at E_b/N_0 of 15, 25 and 35 dB where $\mathbf{m} = [2,2]$ and $\mathbf{m} = [4,4]$. It can be noted from the figure that the average BER of PANOMA degrades slightly as compared to C-NOMA when α_1 deviates from the optimal power for both systems. When the value of α_1 is substantially different from the optimal, the average BER of both systems converge. Moreover, it can be noted that the C-NOMA BER is highly sensitive for $\alpha_1 < 0.1$, where the degradation becomes severe. Also, Figure 4.7 is identical to Figure 4.6 except that $\mathbf{m} = [4,2]$. The enhanced-PANOMA is denoted as E-PANOMA is considered where it assigns the power coefficient of the U_1 signal's imaginary part independently. It is found that $\alpha_1^{\text{Im}} = 0.4681$ is optimal at high E_b/N_0 values. The three demonstrated systems achieve almost identical system's average BER at the optimal power assignment and at $\alpha_1 > 0.2$. However, PANOMA shows a barely seen improvement compared to C-NOMA when $\alpha_1 < 0.2$. This is because of the fact that α_1^{Im} tends to go to zero for such region. Interestingly,

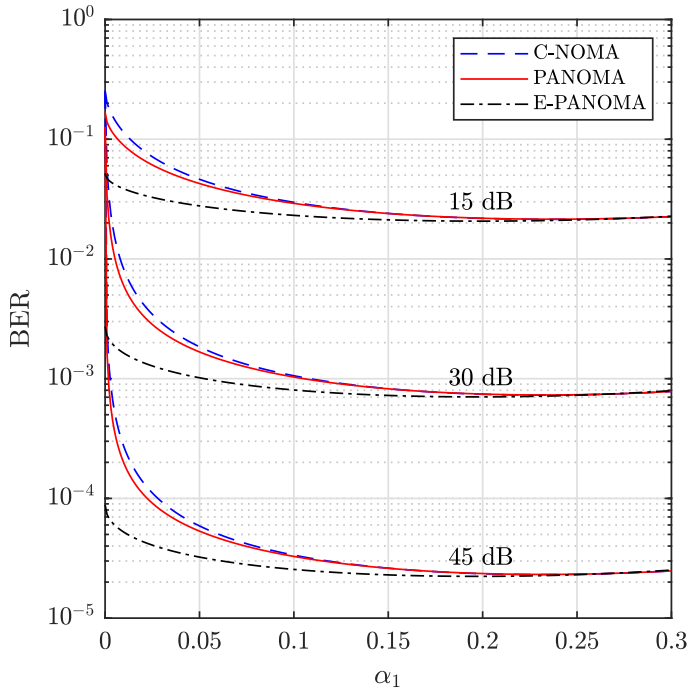


Figure 4.7: Analytical average BER of C-NOMA, PANOMA, E-PANOMA versus α_1 at different E_b/N_0 values, where $\mathbf{m} = [4,2]$.

E-PANOMA which fixes α_1^{Im} at 0.4681 achieves the best performance for $\alpha_1 < 0.2$ region where the PANOMA problem of α_1^{Im} tending to go to zero is solved.

Figure 4.8 is presented to demonstrate the effect of power allocation on the BER of the C-NOMA and PANOMA systems for $N = 3$ case. Therefore, two different power assignments are used which are $\mathcal{P}_1$ and $\mathcal{P}_2$. It is worth mentioning that $\mathcal{P}_1$ is considered optimal for the C-NOMA system at high E_b/N_0 values. As can be noted from the figure, for $\mathcal{P}_1$ both systems offer approximately the identical BER even though the PANOMA power assignment is not optimal. For $\mathcal{P}_2$, which is near optimal for PANOMA and non-optimal for the C-NOMA, the PANOMA BER remained almost unchanged, while the C-NOMA suffered a drastic degradation, which confirms that PANOMA is robust against the imperfect power allocation that might be caused by imperfect CSI at the transmitter. More specifically, the PANOMA have an advantage of 8.1 dB in the average BER as compared to C-NOMA when both systems use $\mathcal{P}_2$. Furthermore, when comparing the average BER of $N = 3$ and $N = 2$, the average BER performance is degraded for $N = 3$ because of the increased IUI.

4.5.2 BER Performance: Robustness and Preformance Gain

Furthermore, Figure 4.9 is similar to the previous figure except that $\mathbf{m} = [2,4]$. Also, E-PANOMA is considered which assigns the power coefficient for the imaginary part

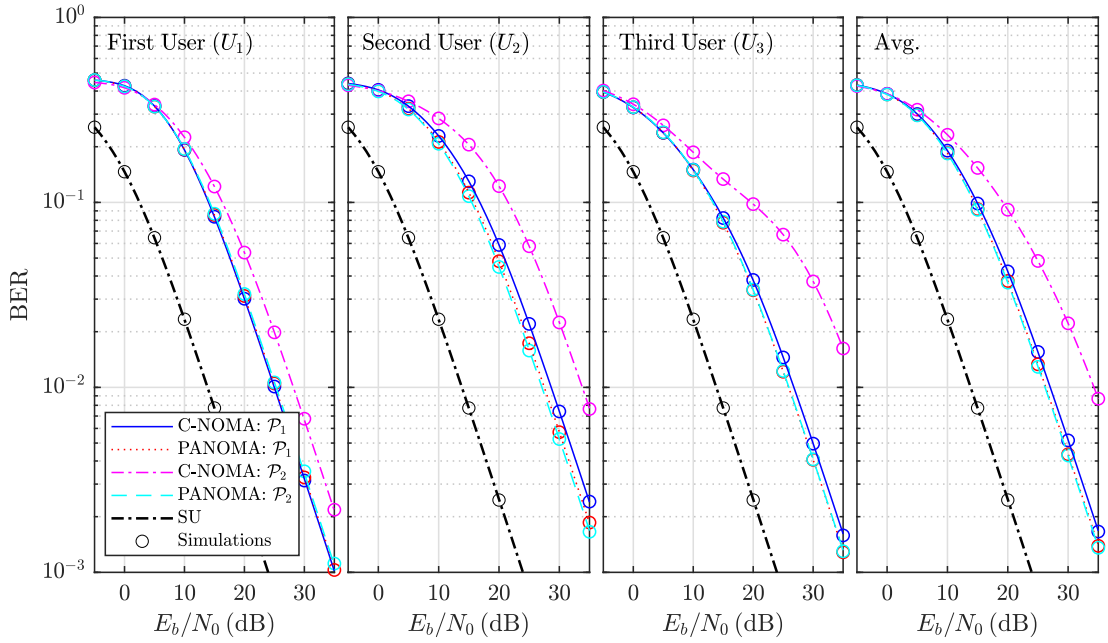


Figure 4.8: Analytical and simulation BER results of C-NOMA and PANOMA systems for $\mathbf{m} = [2,2,2]$ and $\mathbf{m} = [4,4,4]$.

of U_2 signal independently where $\alpha_2^{\text{Im}} = 0.6267$ is found to be optimal at high E_b/N_0 values. The BER performance can be checked at three main check points: 1) Optimal power, where PANOMA outperforms C-NOMA. 2) Sub-optimal power $<$ optimal power, where PANOMA and E-PANOMA are better and the gaps between both schemes and C-NOMA increase rapidly as $\alpha_1 \to 0$ which ensures PANOMA's robustness. This is justified by the fixed power assignment for the case of $x_1 = \text{Re}[x_2]$ and also because α_2^{Im} does not tend to be zero. 3) Sub-optimal power $>$ optimal power, where PANOMA performance is identical to C-NOMA for certain power values, but it faces a marginal loss before both systems' performances converge as $\alpha_1 \to 1/3$. It can be seen that E-PANOMA achieves a better system's average BER compared to PANOMA.

Figure 4.10 studies the PANOMA and C-NOMA sensitivity to the sub-optimal power assignments by considering the deviation from the optimal average BER. The figure shows the BER performance against the deviation from the optimal power assignment (Σ), which is given as $\Sigma = \frac{\alpha_1 - \alpha_1^*}{\alpha_1^*}$ where α_1^* is the optimal power coefficient. It can be seen that the PANOMA's BER performance remains almost unchanged even though the deviation from the optimal power assignment can be huge. Furthermore, it is clear that PANOMA is better when both systems use the optimal power assignment. Hence, the figure confirms that PANOMA is robust to imperfect power assignment, and it achieves better optimal BER performance.

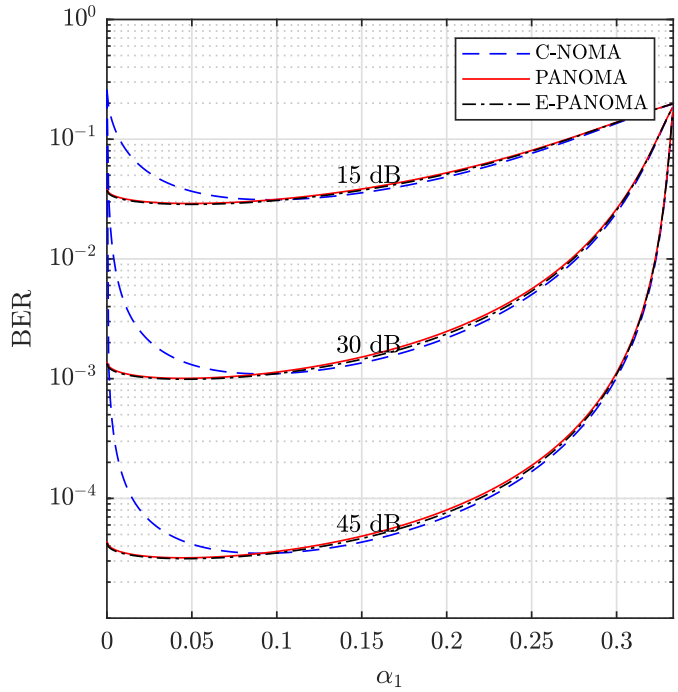


Figure 4.9: Analytical average BER of C-NOMA, PANOMA, E-PANOMA versus α_1 at different E_b/N_0 values, where $\mathbf{m} = [2,4]$.

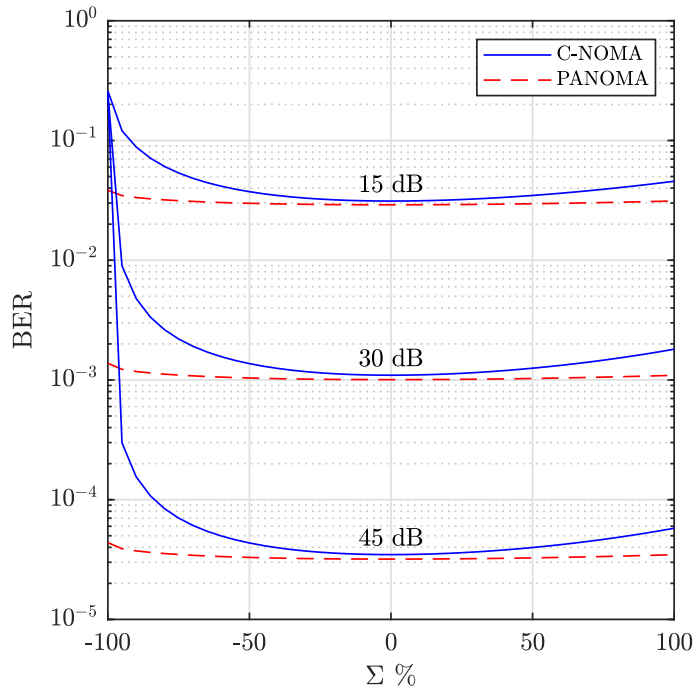


Figure 4.10: Analytical average BER of C-NOMA and PANOMA versus Σ at different E_b/N_0 values, where $\mathbf{m} = [2,4]$.

Figure 4.11 is similar to Figure 4.10 except that different pathloss coefficients are considered for U_2 which are due to either Υ_2 or λ change. It is noted that as $\Sigma \to -100\%$, C-NOMA faces drastic degradation. However, PANOMA's performance remains almost unchanged. In addition, the gap between PANOMA and C-NOMA

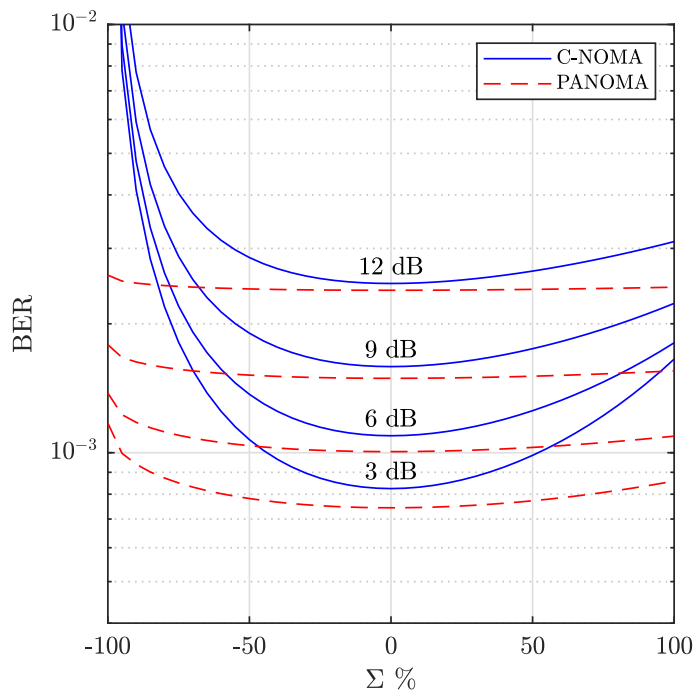


Figure 4.11: Analytical average BER of C-NOMA and PANOMA versus Σ at different pathloss coefficients for U_2 where $\mathbf{m} = [2,4]$ and $E_b/N_0 = 30$ dB.

increases as the channel disparity between the two users is less, which is expected because the performance is dominated by the AWGN when the channel disparity is high.

4.5.3 Capacity Performance

Figures 4.12–4.14 quantify the PANOMA gain with respect to the achieved system capacity for $N = 2$ and $N = 3$ cases, where the power allocation for $N = 3$ is $\mathcal{P}_2$. For example, Figure 4.12 demonstrates the capacity for identical modulation orders when $N = 2$. It can be seen that the achieved capacity at extremely high SNR values tend to saturate at $2\log_2 M_n$ which is due to the fact that the BER reaches extremely low levels, and hence, the conditional entropy drops to almost zero. In addition, maximum gain is 1.2 dB for BPSK and 1.45 dB for QPSK. Furthermore, when considering $N = 2$ and non-identical modulation orders as in Figure 4.13, it can be seen that E-PANOMA is the superior always for such power assignment. Also, $\mathbf{m} = [2,4]$ outperforms $\mathbf{m} = [4,2]$ by up to 2.5 dB. Moreover, Figure 4.14 shows that the maximum gain for $N = 3$ case is about 6.8 and 6.9 dB for BPSK and QPSK, respectively. At high E_b/N_0 values, NOMA achieves a capacity of $\mathcal{M}$ bit/symbol. Nevertheless, the single-user offers higher capacity at low E_b/N_0 values. As can be noted from these figures, the

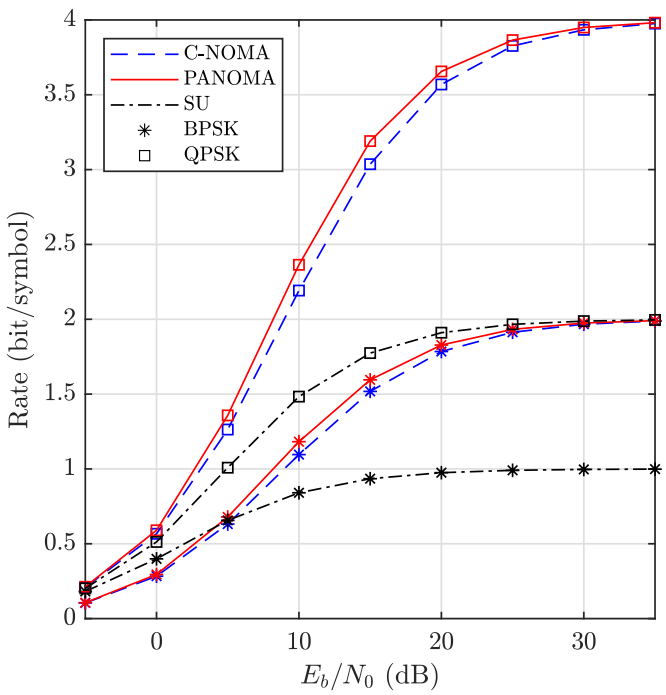


Figure 4.12: Constrained capacity for identical BPSK and QPSK cases, $N = 2$.

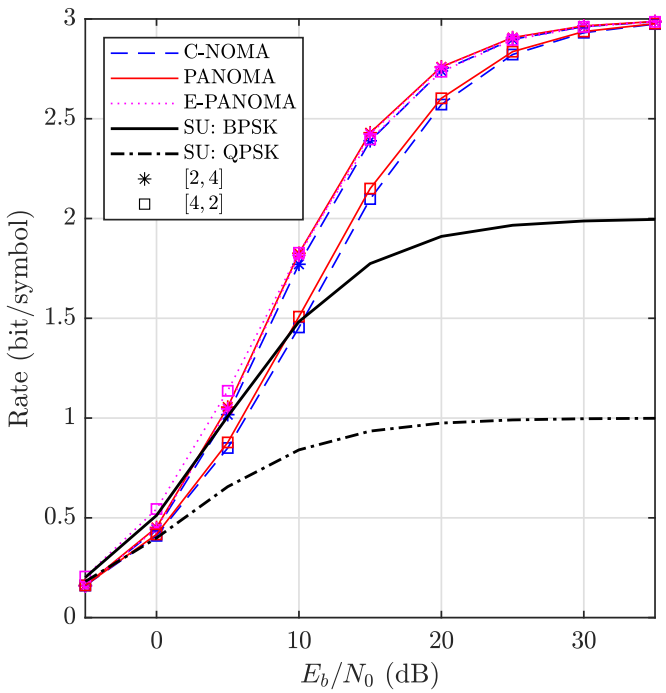


Figure 4.13: Constrained capacity of C-NOMA and PANOMA for $\mathbf{m} = [2,4]$ and $\mathbf{m} = [4,2]$.

capacity of PANOMA and C-NOMA decreases and approaches zero when E_b/N_0 is decreased. Such behaviour is obtained because the capacity at low E_b/N_0 is dominated by the AWGN, and thus, reducing the IUI or optimizing the power will have a very limited impact on the system performance. By increasing the E_b/N_0, the impact of the

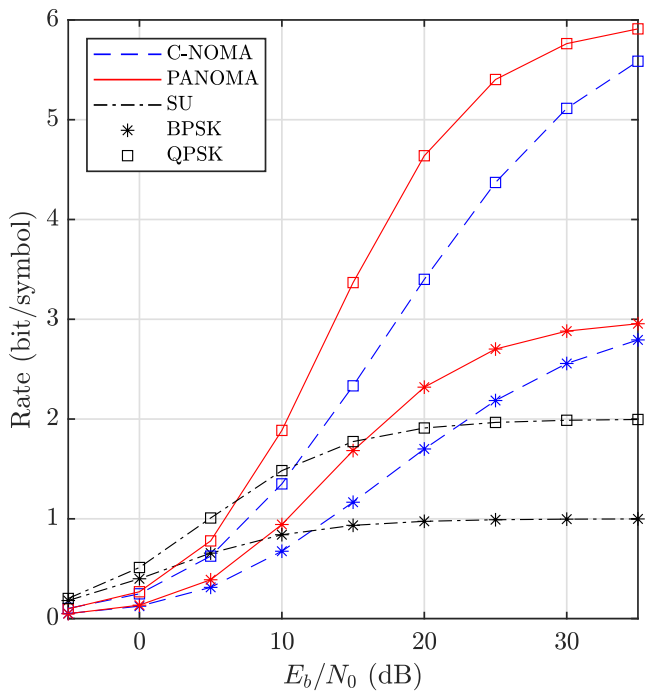


Figure 4.14: Constrained capacity of C-NOMA, PANOMA and E-PANOMA using identical BPSK and QPSK, where $N = 3$..

proposed power assignment becomes more apparent.

4.5.4 Optimal Power Assignments

While Tables B.1 and B.2 in Appendix B demonstrate the optimal power assignments to minimise the system's average BER for $N = 2$ and $N = 3$ cases under various E_b/N_0 conditions, Tables B.3 and B.4 in Appendix B show the optimal power assignments that maximise the capacity for the same scenarios. It is confirmed that identical QPSK and identical BPSK modulation orders for $N = 2$ system achieve identical performance with respect to E_b/N_0. Even though PANOMA has a different power assignment compared to C-NOMA the performance is identical. In addition, it is noted that E-PANOMA always achieves a better system's average BER. However, this enhancement is justified because of the extra degree of freedom given when the power coefficient for the imaginary part of the signal can be optimized independently. Moreover, PANOMA system's performance faces a slight degradation compared to C-NOMA for $N = 3$. The reason for this degradation is believed to be the power coefficients in (4.7) which are found by brute force. This implies that a better technique could be used to eliminate the degradation. Moreover, it is observed that the systems' constrained capacity reaches its maximum limit at the highest E_b/N_0 of 45 dB where the limit is $\mathcal{M}$.

4.6 Conclusions

This chapter proposed an efficient power assignment scheme to improve the BER performance of NOMA-IoT systems by increasing the system tolerance to non-optimal power assignment. Exact closed-form analytical BER and lower bound capacity expressions were derived for the JMLD receiver of both C-NOMA and PANOMA over Rayleigh fading channels, where the cases of two and three users were considered. These expressions were verified via Monte Carlo simulation results, which showed that the proposed PANOMA is superior in terms of BER and capacity. This demonstrated the ability of PANOMA to operate effectively under non-optimal power allocation, which is typically caused by the lack of accurate CSI at the transmitter. Also, PANOMA opened the door for independent in-phase and quadrature-phase power optimization which showed a performance enhancement with the cost of extra overhead. Moreover, it can be noted that the performance gain achieved using PANOMA is maintained even when the number of users increased from two to three. The same conclusion can be also made for the capacity results where the PANOMA outperformed the NOMA. Furthermore, when PANOMA was applied to QPSK, the performance improvement was identical to the BPSK case.

Chapter 5

NOMA with Joint-Multiuser Gray-Labeling

5.1 Chapter Introduction

This chapter considers the design of a generalized Gray-mapping process to non-orthogonal multiple access (NOMA) with arbitrary modulation orders and number of users. Unlike orthogonal multiple access, joint-multiuser Gray-mapping can provide significant bit error rate (BER) improvement, which can be used to alleviate the degradation caused by the multiuser interference inherent in non-orthogonal schemes. The obtained improvement is due to the increased Euclidean distance that Gray-mapping provides for certain users. To evaluate the impact of Gray-mapping, closed-form expressions are derived for the exact BER with imperfect successive interference cancellation. The obtained analytical and simulation results demonstrate that the proposed scheme can offer up to 10 dB gain over conventional NOMA in certain scenarios, where such a performance gain can be shared between the users by selecting appropriate power assignment. Moreover, feasibility maps are generated to demonstrate the additional flexibility that Gray-mapping can offer in terms of quality of service satisfaction for the various layers.

5.1.1 Chapter Organisation

The rest of the chapter is organized as follows. Section 5.2 is the related work, and the motivations and contributions, Section 5.3 presents the system and channel models. Section 5.4 is the BER analysis. Section 5.5 includes the numerical results and

discussions while Section 5.6 concludes the work and provides final remarks.

5.2 Related Work

The growing demand for wireless connectivity makes it challenging to accommodate the required requested by mobile broadband and terrestrial broadcast systems due to spectrum scarcity. However, several non-orthogonal schemes such as the non-orthogonal multiplexing (NOM) and NOMA have demonstrated significant ability to improve the spectral efficiency, support massive connectivity, serve users with diverse quality of service (QoS) requirements, and improve fairness amongst users [118–120]. Particularly, NOM schemes such as layer division multiplexing and hierarchical modulation are included in the digital video broadcasting-terrestrial (DVB-T) and advanced television systems committee 3.0 (ATSC 3.0) specifications [118, 121, 122], while the NOMA based multiuser superposition transmission (MUST) is included in the third generation partnership project (3GPP) specifications for Release 14 of the Long Term Evolution (LTE) [123]. The basic concept of such non-orthogonal schemes is to multiplex the streams/users in the power-domain with different power coefficients via superposition coding (SC), while successive interference cancellation (SIC) is used at the receiver to suppress undesired interference [9]. Nonetheless, the error rate performance degrades due to non-orthogonality and imperfect interference cancellation, [24]. Therefore, error rate analysis of NOM and NOMA has recently received much interest from the research community [24, 34, 120, 124–126]. For example, [125] derives the BER of uniform quadrature amplitude modulation (QAM)-based NOM system, where the power coefficients are optimized to minimize the sum BER. Moreover, the BER for downlink (DL) NOMA over Rayleigh fading channels is analyzed in [34] for the case of square QAM with two users, while [24] expands the work for the case of an arbitrary number of users. In addition, the authors of [126] derive BER expressions for amplitude-coherent detection NOMA using unipolar amplitude-shift keying, where two and three users are considered in ordered single-input-single-output (SISO) Rayleigh fading channels.

On the other hand, NOM is generalized to uplink (UL) transmission where automatic repeat requests to improve the spectral and energy efficiencies [124]. Furthermore, UL NOMA is analyzed in [66, 68, 69, 127]. For example, Liu and Beaulieu [66] derived closed-form union bounds on the symbol error rate (SER) and BER of a two-user NOMA with quadrature phase-shift keying (QPSK) for arbitrary relative phase

offset in additive white Gaussian noise (AWGN) channels. Furthermore, Liu and Beaulieu [127] considered asynchronous UL NOMA and derived the SER for a two-user system with arbitrary QAM over Rayleigh fading channels. Semira *et al.* [68] propose a joint-multiuser maximum likelihood detector (JMLD) to eliminate the error floor caused by the SIC. Furthermore, the work is extended in [69] to derive an upper bound expression for the BER over Rayleigh fading channels while considering an arbitrary number of users, square QAM and maximum ratio combining (MRC)-JMLD.

Several approaches and assistive techniques have been introduced to improve the BER performance of NOMA including intelligent reflecting surface (IRS)-assisted NOMA [88, 119, 128], constellation rotation [129, 130], data- and channel-aware interference exploitation [95, 106] and bit-to-symbol mapping [97, 98, 131–134]. Moreover, two main bit-to-symbol mapping schemes for DL NOMA are highlighted in the literature, which are the disjoint- and joint-multiuser Gray-mapping. The former results when the data streams of each user are Gray-mapped independently before the SC, while the latter is performed by jointly Gray-mapping all users' data streams before SC.

Yan *et al.* [131] propose joint-multiuser Gray-mapping for NOMA for the first time to enhance the system's performance. While a general mapper/demapper design for any modulation order or number of users is missing, the presented example is for the two-user case with QPSK. In [97], constellation rotation is proposed while considering joint-multiuser Gray-mapping. Similarly, two-user case with QPSK is considered and upper SER bounds are derived. Shieh *et al.* [98] study different mapping schemes for two-user NOMA in AWGN channels without SIC, where uniform constellation is considered such that the bits after SC are mapped to a composite higher modulation constellation that is equivalent to a single-user. The SER and block error rate (BLER) simulation results are presented for M-ary pulse amplitude modulation (M-PAM) with different modulation orders, whereas the only closed-form expression derived is the near-user SER because the far-user performance is identical in the joint- and disjoint-multiuser Gray-mapping schemes. Although the results show that the joint-multiuser Gray-mapping improves the SER and BLER performance of the near-user compared to other mapping schemes, all mapping schemes have the same asymptotic behavior. Furthermore, the authors of [132] consider joint-multiuser Gray-mapping for a single-input-multiple-output (SIMO) system with only two users per subcarrier, where the far-user is multiplexed with all near-users' subcarriers. BER expressions are derived over Rayleigh fading channels for QPSK and MRC using SIC.

Han *et al.* [133] derive SER and BER closed-form expressions for the joint- and disjoint-multiuser Gray-mapping schemes for the two-user case with QPSK in AWGN channels. The presented results highlight that the error propagation due to SIC imperfections can be eliminated at high signal to noise ratios (SNRs) when Gray-mapping is performed jointly. The work in [133] is extended in [134] by considering an arbitrary number of users where closed-form expressions are presented for specific cases such as the first, second and last user, and the expressions for other users can be implicitly inferred. In addition, approximate BER expressions were provided.

Furthermore, symmetrical coding is proposed in [135] for a two-user NOMA with binary phase-shift keying (BPSK) modulation. The proposed scheme manipulates the transmitted symbols such that the NOMA bits are Gray-mapped for the near-user as well. Closed-form BER expressions over Nakagami-m fading are derived. The power coefficients are optimized based on statistical channel state information (CSI) to improve the near-user BER performance.

5.2.1 Motivations and Contributions

As can be noted from the aforementioned discussion, joint-multiuser Gray-mapping has the potential to improve the BER performance of NOMA [97, 98, 131–134]. Based on the surveyed literature and to the best of the authors' knowledge, all mentioned papers lack a general mapper/demapper design for the joint-multiuser Gray-mapping. Furthermore, the presented analyses have constrained the number of users and modulation orders. For instance, the work in [98] proposes Gray-mapping based on a look-up table (LUT) rather than actual SC which limits the flexibility of applying different powers for different users. Additionally, the number of users is limited to 2, and increasing the number of users would require additional LUTs, and hence, higher computational complexity. Furthermore, the work in [133, 134] considers only QPSK, which limits the contribution of the scheme. This work considers generalizing the Gray-mapping process NOMA systems with arbitrary modulation orders. Therefore, the main contributions can be summarized as follows:

1. Design a novel joint-multiuser Gray-mapper/ demapper for NOMA with an arbitrary rectangular QAM and arbitrary number of users.

2. Derive exact conditional BER expressions while considering imperfect SIC.

3. Generalize the derived exact conditional BER expressions for the two-user scenario to square QAM with arbitrary modulation orders.

4. Quantify the performance gain of the joint-multiuser Gray-mapping over the disjoint-multiuser Gray-mapping.

5. Evaluate the impact of varying the channel quality, number layers, and modulation orders on the BER.

6. Generate feasibility maps to highlight the flexibility of joint-multiuser Gray-mapping in terms of QoS satisfaction for various layers.

5.3 System And Channel Models

In power-domain NOMA, N users extract their messages from the broadcast NOMA message transmitted by the base station (BS). The broadcast NOMA message contains the information symbols of N users $\{U_1, U_2, \ldots, U_N\}$ that are multiplexed by allocating a certain power level for each layer as described in [24,27]. Typically, the power for the users is allocated in the opposite order of the channel strength such that the user with the worst channel condition would get the highest power. This enables reliable SIC detection and ensures that the receivers are connected [119]. Consequently, the coefficients are assigned such that $\alpha_1 < \alpha_2 < \cdots < \alpha_N$, where $\sum_{n=1}^{N} \alpha_n = 1$. It is also worth noting that the other conditions required to enable reliable SIC detection are described in [24, 27]. Given that U_n is assigned $\mathcal{M}_n \in \{1, 2, \ldots\}$ bits/symbol, then the modulation order is $M_n = 2^{\mathcal{M}_n}$. In this work, we assume a coordinated scenario, where the BS informs the users about their allocated power coefficients and modulation orders. Considering BPSK and rectangular QAM, the information symbol of U_n is denoted by x_n, which is uniformly drawn from $\mathbb{S}_n = \left\{ s_0^{(n)}, s_1^{(n)}, \ldots, s_{M_n-1}^{(n)} \right\}$, where the constellation points are arranged over a $M_n^{(x)} \times M_n^{(y)} \triangleq M_n$ rectangle with $M_n^{(x)} = 2^{\lceil 0.5 \log_2(M_n) \rceil}$ being the length and $M_n^{(y)} = 2^{\lfloor 0.5 \log_2(M_n) \rfloor}$ being the width of the rectangle, where $\lceil \cdot \rceil$ is the ceiling function and $\lfloor \cdot \rfloor$ is the floor function. The power normalization factor is given as

$$\kappa_n = \frac{1}{3} \left[\left(M_n^{(x)} \right)^2 + \left(M_n^{(y)} \right)^2 - 2 \right] \tag{5.1}$$

which ensures unity average power for the nth user symbols, i.e. $\mathrm{E}[|x_n|^2] = 1$.

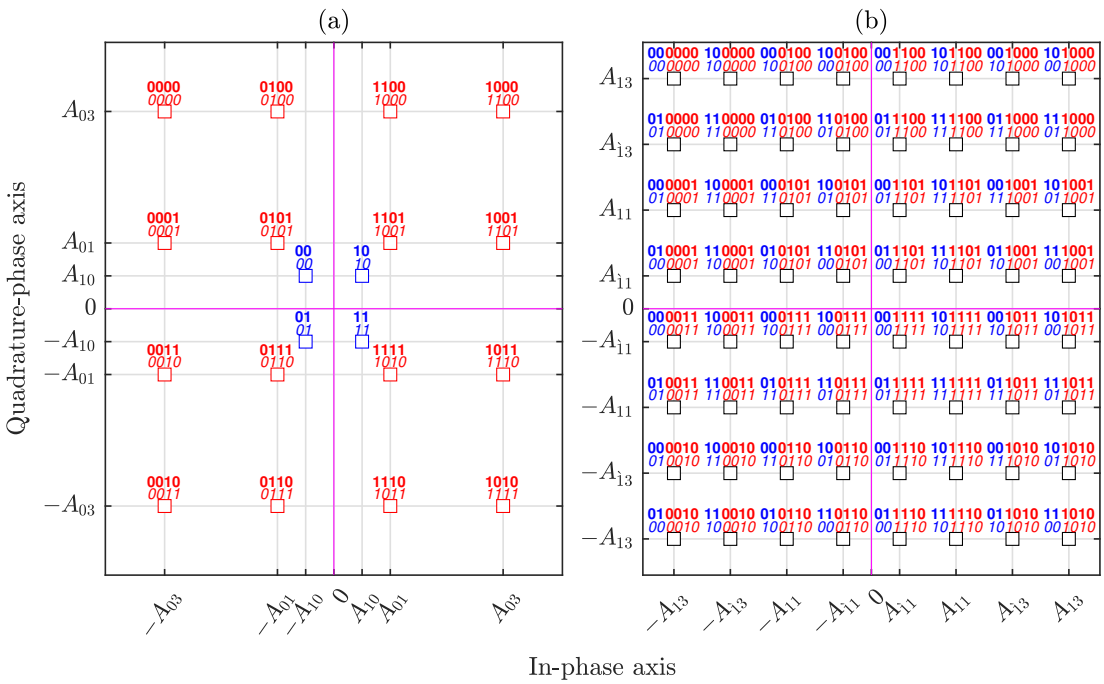


Figure 5.1: Constellation diagrams for a) Individual user constellation diagram for $M_1 = 4$ (blue) and $M_2 = 16$ (red) after scaling where the bits block is shown above each symbol: Gray-mapping (bold), Natural binary mapping (italic). b) NOMA constellation diagram for $[M_1, M_2] = [4, 16]$ where the bits block is shown above each symbol: Disjoint-multiuser Gray-mapping (bold), joint-multiuser Gray-mapping (italic).

5.3.1 Disjoint-Multiuser Gray-Mapping

In disjoint-multiuser Gray-mapping, each user's bits block $\mathbf{b}_n$ is independently mapped to a symbol x_n using a rectangular Gray-mapped M_n-QAM constellation mapper $\mu\{\cdot\}$, such that $x_n = \mu\{\mathbf{b}_n\}$ where $x_n \in \mathbb{S}_n = \left\{s_0^{(n)}, s_1^{(n)}, \ldots, s_{M_n-1}^{(n)}\right\}$. Given that NOMA multiplexes the users' symbols using distinct power coefficients α_n, the NOMA symbol can be expressed as

$$x = \sum_{n=1}^{N} \sqrt{\alpha_n} x_n, \quad x \in \mathbb{S} = \{s_0, s_1, \ldots, s_{M-1}\} \tag{5.2}$$

where $M = M_1 \times M_2 \times \cdots \times M_n$ and the NOMA symbols set $\mathbb{S}$ can be generated using the Minkowski sum such that $\mathbb{S} = \{\sum_n \sqrt{\alpha_n}\mathbf{s}^{(n)} | \mathbf{s}^{(n)} \in \mathbb{S}_n\}$ [136]. The constellation diagram for the individual users' in a two-user NOMA after power scaling is shown in Figure 5.1a where $[M_1, M_2] = [4, 16]$, and the superposed NOMA symbols are shown in Figure 5.1b for both the disjoint- and joint-multiuser Gray-mapping schemes. To compare the two mapping schemes, let the Hamming distance between any two adjacent symbols for a particular user's bits be denoted as δ_n. Therefore, it can be seen that the disjoint-multiuser Gray-mapping guarantees a maximum $\delta_n = 1$, where Figure

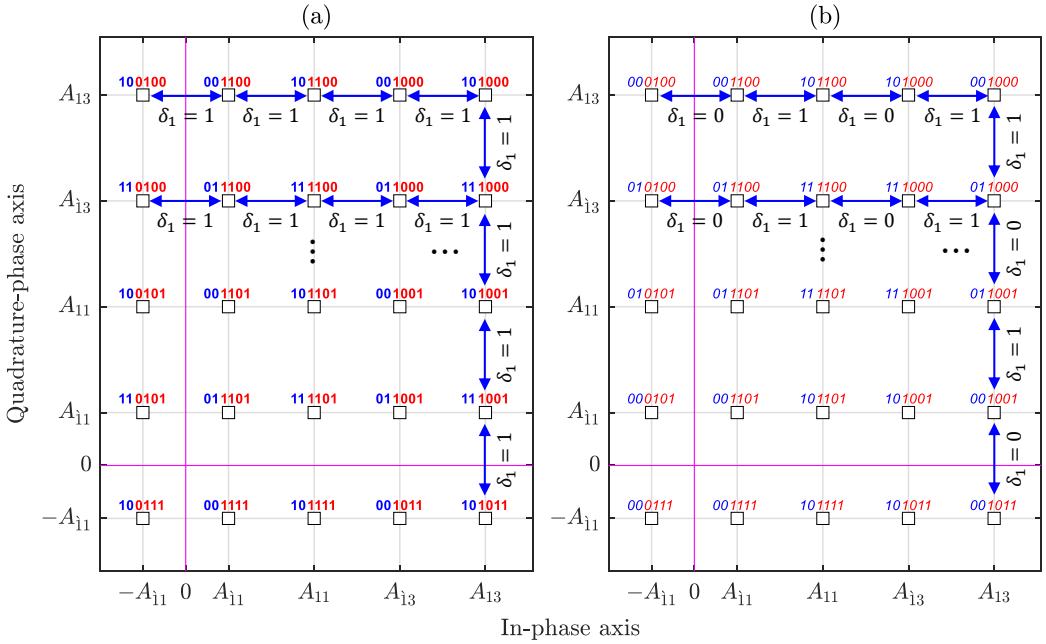


Figure 5.2: Constellation diagrams of NOMA for $[M_1, M_2] = [4, 16]$ showing δ_1, where the bits block is shown above each symbol: a) disjoint-multiuser Gray-mapping, b) joint-multiuser Gray-mapping.

5.2a shows part of the constellation diagram for the considered example highlighting δ_1. Nonetheless, it does not guarantee that the maximum $\delta_1 + \delta_2 = 1$. For example, we can see from Figure 5.2a that the adjacent symbols with bits 011001 and 001011 have a $\delta_1 + \delta_2 = 2$ while $\delta_1 = \delta_2 = 1$.

The real and imaginary components of $\mathbb{S}$ elements are denoted by $A_{\nu_1,\nu_2,\ldots,\nu_N}$ and $A_{\xi_1,\xi_2,\ldots,\xi_N}$, respectively. For the real components $\nu_n \in \{0, \pm 1, \pm 3, \ldots, \pm M_n^{(x)} - 1\}$, while for the imaginary components $\xi_n \in \{0, \pm 1, \pm 3, \ldots, \pm M_n^{(y)} - 1\}$, which are related to the individual users' symbols. For notational simplicity, we define $\grave{\nu}_n \triangleq -\nu_n$ and $\grave{\xi}_n \triangleq -\xi_n$. Therefore, the real and imaginary components of $\mathbb{S}$ can be expressed as

$$A_{\nu_1,\nu_2,\ldots,\nu_N} = \sum_{n=1}^{N} \nu_n \sqrt{\frac{\alpha_n}{\kappa_n}} \tag{5.3}$$

and

$$A_{\xi_1,\xi_2,\ldots,\xi_N} = \sum_{n=1}^{N} \xi_n \sqrt{\frac{\alpha_n}{\kappa_n}}. \tag{5.4}$$

At the receivers sides, the received signal at the nth user receiver can be expressed as

$$y_n = h_n x + w_n \tag{5.5}$$

where the complex channel coefficient is denoted as h_n, the AWGN is $w_n \sim \mathcal{CN}(0, N_0)$ and $N_0/2 \triangleq \sigma_w^2$ is the complex noise power. Assuming Rician fading with large-scale

fading, the channel coefficient can be expressed as

$$h_n = \sqrt{\frac{d_n^{-\rho_n} K_n}{K_n + 1}} \left(1 + \check{h}_n\right) \tag{5.6}$$

where $\check{h}_n \sim \mathcal{CN}(0,1)$, d_n is the distance between the BS and the receiver, ρ_n is the pathloss exponent and K_n is the Rician factor which represents the ratio of the average power of the line-of-sight (LOS) component to the non-line-of-sight (NLOS) components. By defining the instantaneous SNR as $\gamma_n \triangleq \frac{2|h_n|^2}{N_0}$, then its probability density function (PDF) can be expressed as [22, Eq. (2.16)]

$$f(\gamma_n) = \frac{\left(1+\Omega^2\right) e^{-\Omega^2}}{\bar{\gamma}_n} \exp\left(-\frac{\left(1+\Omega^2\right)\gamma_n}{\bar{\gamma}_n}\right) \times I_0\left(2\Omega\sqrt{\frac{(1+\Omega^2)\gamma_n}{\bar{\gamma}_n}}\right), \ \gamma_n \geq 0 \tag{5.7}$$

where $\Omega^2 = d_n^{-\rho_n} K_n$, $I_0(\cdot)$ is the modified Bessel function and $\bar{\gamma}_n = \frac{2\mathrm{E}\left[|h_n|^2\right]}{N_0}$.

For coherent detection, U_N has the maximum power coefficient, and thus, it can be detected directly using a single-layer maximum likelihood detector (MLD) while considering all other users' signals as unknown additive noise. Therefore,

$$\hat{x}_N = \arg\min_{x_N \in \mathbb{S}_N} \left|y_n - h_n\sqrt{\alpha_N} x_N\right|^2. \tag{5.8}$$

On the other hand, for $1 < n < N$, the nth user will have interference from high and low power users. Therefore, SIC is used to cancel the interference from the higher power signals while lower power signals will be treated as noise. This can be expressed recursively as

$$\hat{x}_n = \arg\min_{x_n \in \mathbb{S}_n} \left|y_n - h_n \sum_{i=n+1}^{N} \sqrt{\alpha_i}\hat{x}_i - h_n\sqrt{\alpha_n} x_n\right|^2. \tag{5.9}$$

Furthermore After $\hat{x}_n$ detection, a rectangular Gray-mapped M_n-QAM constellation demapper $\mu^{-1}\{\cdot\}$ is used to find the nth user bits block such that $\hat{\mathbf{b}}_n = \mu^{-1}\{\hat{x}_n\}$. While SIC is a low-complexity detector, its introduced delay increases as the number of users increases. Therefore, for delay-sensitive applications JMLD should be used at the expense of increased computational complexity [23]. In addition, NOMA is well situated

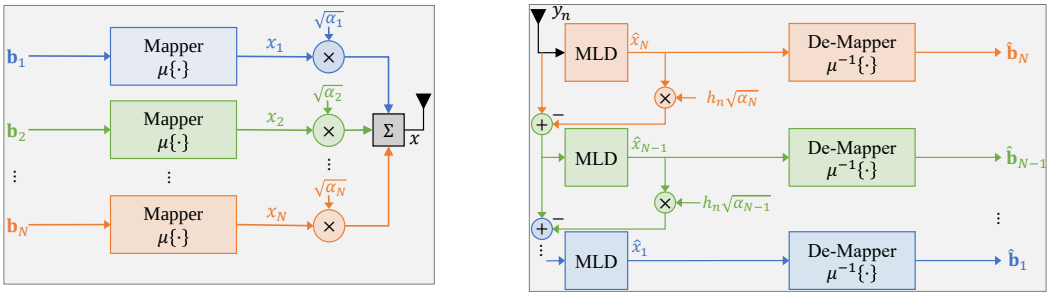


Figure 5.3: NOMA system block diagram for disjoint-multiuser Gray-mapping.

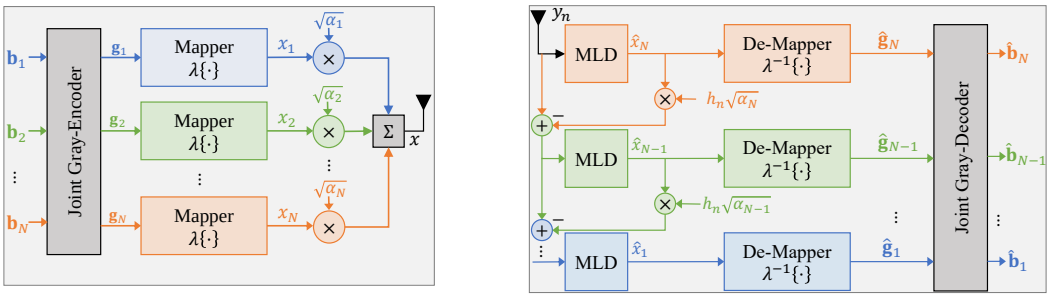


Figure 5.4: NOMA system block diagram for joint-multiuser Gray-mapping.

for scenarios including two and three users being multiplexed in the same communication resource. Therefore, either the complexity of JMLD or the delay of SIC for these scenarios is acceptable. Figure 5.3 illustrates the NOMA system with disjoint-multiuser Gray-mapping at the transmitter and the receivers.

5.3.2 Joint-multiuser Gray-Mapping

To guarantee joint-multiuser Gray-mapping and the maximum Hamming distance between adjacent symbols to be one regardless of the number of users, the bits block $\mathbf{b}_n \forall n$ should be jointly Gray-encoded to generate the block $\mathbf{g}_n$, which is then mapped to a symbol x_n using a rectangular natural binary mapped M_n-QAM constellation mapper $\lambda\{\cdot\}$, such that $x_n = \lambda\{\mathbf{g}_n\}$. The NOMA multiplexing and the symbol detection are identical to the previous section. However, $\hat{x}_n$ is demapped using a rectangular natural binary mapped M_n-QAM constellation demapper $\lambda^{-1}\{\cdot\}$ such that $\hat{\mathbf{g}}_n = \lambda^{-1}\{\hat{x}_n\}$. Finally, $\hat{\mathbf{g}}_n$ is jointly Gray-decoded and the user bits block $\hat{\mathbf{b}}_n$ is generated. Figure 5.4 illustrates the NOMA system with joint-multiuser Gray-mapping at the transmitter and the receivers. Meanwhile channel coding would lead to a better interference cancellation if SIC is applied on the codeword-level rather than the symbol-level, the codeword-level SIC lacks optimality and optimal detector needs to be designed [137,

Section VI]. Nonetheless, it is outside the scope of this work as the insights obtained with or without off-the-shelf channel coding are generally the same and the impact of channel coding can be interpreted in terms of SNR improvement or coding gain. In the following, the mapping and demapping processes of NOMA with joint-multiuser Gray-mapping will be described in detail.

5.3.2.1 Mapping

The nth user bits block can be divided into in-phase and quadrature bits. In other words, $\mathbf{b}_n = \left[b_n^{(1)}, b_n^{(2)}, \ldots, b_n^{(\mathcal{M}_n)}\right] = \left[\mathbf{b}_n^{(x)}, \mathbf{b}_n^{(y)}\right]$, and this can be written as $\left[\mathbf{b}_n^{(x)}, \mathbf{b}_n^{(y)}\right] = \left[b_n^{(1)}, \ldots, b_n^{\left(\mathcal{M}_n^{(x)}\right)}, b_n^{\left(\mathcal{M}_n^{(x)}+1\right)}, \ldots, b_n^{\left(\mathcal{M}_n^{(x)}+\mathcal{M}_n^{(y)}\right)}\right]$, where $\mathcal{M}_n = \log_2(M_n)$ and $\mathcal{M}_n^{(i)} = \log_2\left(M_n^{(i)}\right)$, $i \in \{x, y\}$. Hence, the in-phase and quadrature NOMA bits can be arranged based on the users' priorities, i.e. based on the allocated power coefficients, $\tilde{\mathbf{b}}^{(x)} = \left[\mathbf{b}_N^{(x)}, \mathbf{b}_{N-1}^{(x)}, \ldots, \mathbf{b}_1^{(x)}\right]$ and $\tilde{\mathbf{b}}^{(y)} = \left[\mathbf{b}_N^{(y)}, \mathbf{b}_{N-1}^{(y)}, \ldots, \mathbf{b}_1^{(y)}\right]$. Therefore, the arranged jointly Gray-encoded in-phase and quadrature NOMA bits, $\tilde{\mathbf{g}}^{(x)}$ and $\tilde{\mathbf{g}}^{(y)}$, can be computed by

$$\tilde{g}_i^{(x)} = \begin{cases} \tilde{b}_i^{(x)}, & i = 1 \\ \tilde{b}_i^{(x)} \oplus \tilde{g}_{i-1}^{(x)} & i > 1 \end{cases} \tag{5.10}$$

$$\tilde{g}_i^{(y)} = \begin{cases} \tilde{b}_i^{(y)}, & i = 1 \\ \tilde{b}_i^{(y)} \oplus \tilde{g}_{i-1}^{(y)} & i > 1 \end{cases}. \tag{5.11}$$

Since $\tilde{\mathbf{g}}^{(x)} = \left[\mathbf{g}_N^{(x)}, \mathbf{g}_{N-1}^{(x)}, \ldots, \mathbf{g}_1^{(x)}\right]$ and $\tilde{\mathbf{g}}^{(y)} = \left[\mathbf{g}_N^{(y)}, \mathbf{g}_{N-1}^{(y)}, \ldots, \mathbf{g}_1^{(y)}\right]$, the jointly Gray-encoded bits for each layer can be expressed as $\mathbf{g}_n = \left[\mathbf{g}_n^{(x)}, \mathbf{g}_n^{(y)}\right]$. The joint-multiuser Gray-mapped NOMA symbols are shown in Figures 5.1b and 5.2b. As can be noted from Figure 5.2b, the maximum $\delta_1 = 1$ implies that selecting a symbol adjacent to the transmitted symbol affects only one of the users, while the other user remains unaffected. Therefore, we obtain $\delta_1 = 0$ for several cases. Consequently, the joint-multiuser Gray-mapped always ensures that the sum of δ_n across two symbols is 1.

5.3.2.2 Demapping

At the nth user, after symbol detection and acquisition of the jointly Gray-encoded NOMA bits block, $\mathbf{g} = [\mathbf{g}_n, \mathbf{g}_{n+1}, \ldots, \mathbf{g}_N]$, the in-phase and quadrature jointly Gray-encoded NOMA bits can be rearranged such that $\tilde{\mathbf{g}}^{(x)} = \left[\mathbf{g}_N^{(x)}, \mathbf{g}_{N-1}^{(x)}, \ldots, \mathbf{g}_n^{(x)}\right]$ and $\tilde{\mathbf{g}}^{(y)} =$

$\left[\mathbf{g}_N^{(y)}, \mathbf{g}_{N-1}^{(y)}, \ldots, \mathbf{g}_n^{(y)}\right]$. Hence, the jointly Gray-encoded NOMA bits can be decoded as follows,

$$\tilde{b}_i^{(x)} = \begin{cases} \tilde{g}_i^{(x)}, & i = 1 \\ \tilde{g}_i^{(x)} \oplus \tilde{g}_{i-1}^{(x)} & i > 1 \end{cases} \tag{5.12}$$

$$\tilde{b}_i^{(y)} = \begin{cases} \tilde{g}_i^{(y)}, & i = 1 \\ \tilde{g}_i^{(y)} \oplus \tilde{g}_{i-1}^{(y)} & i > 1 \end{cases}. \tag{5.13}$$

Thus, $\tilde{\mathbf{b}}^{(x)} = \left[\mathbf{b}_N^{(x)}, \mathbf{b}_{N-1}^{(x)}, \ldots, \mathbf{b}_n^{(x)}\right]$ and $\tilde{\mathbf{b}}^{(y)} = \left[\mathbf{b}_N^{(y)}, \mathbf{b}_{N-1}^{(y)}, \ldots, \mathbf{b}_n^{(y)}\right]$. Consequently, $\mathbf{b}_n = \left[\mathbf{b}_n^{(x)}, \mathbf{b}_n^{(y)}\right]$.

5.3.3 Disjoint- and Joint-multiuser Gray-Mapping Examples

For more clarity, this subsection presents an example of the joint-multiuser Gray-mapping process of the NOMA bits shown in Figure 5.1b. Consider that $\mathbf{b}_1 = [0,1]$ and $\mathbf{b}_2 = [1,0,1,0]$. For the disjoint-multiuser Gray-mapping NOMA, the symbols for each user can be found using the rectangular Gray-mapped M_n-QAM mapper $\mu\{\cdot\}$ such that $x_n = \mu\{\mathbf{b}_n\}$. Hence, $x_1 = \mu\{\mathbf{b}_1\} = \frac{1}{\sqrt{2}}(-1-1j)$ and $x_2 = \mu\{\mathbf{b}_2\} = \frac{3}{\sqrt{10}}(1-1j)$. Therefore, the NOMA symbol is

$$\begin{aligned} x &= \sqrt{\alpha_1}x_1 + \sqrt{\alpha_2}x_2 \\ &= -A_{10} - 1jA_{10} + A_{03} - 1jA_{03} \\ &= A_{\grave{1}3} - 1jA_{13} = s_{26} \end{aligned} \tag{5.14}$$

with the NOMA bits block being $\mathbf{b} = [0,1,1,0,1,0]$ that is shown above the NOMA symbol in Figure 5.1b with the bold text.

On the other hand, for the joint-multiuser Gray-mapping NOMA, the bits in the block are jointly Gray-encoded following (5.10)–(5.11). Therefore, $\mathbf{b}_1^{(x)} = [0]$, $\mathbf{b}_1^{(y)} = [1]$, $\mathbf{b}_2^{(x)} = [1,0]$, and $\mathbf{b}_2^{(y)} = [1,0]$. Hence, $\tilde{\mathbf{b}}^{(x)} = [1,0,0]$, $\tilde{\mathbf{b}}^{(y)} = [1,0,1]$, $\tilde{\mathbf{g}}^{(x)} = [1,1,1]$, and $\tilde{\mathbf{g}}^{(y)} = [1,1,0]$. It can be also shown that the jointly Gray-encoded bits block for U_1 and U_2 are $\mathbf{g}_1 = [1,0]$ and $\mathbf{g}_2 = [1,1,1,1]$. Furthermore, the symbols for each user can be found using the rectangular natural binary mapped M_n-QAM mapper $\lambda\{\cdot\}$ such that $x_n = \lambda\{\mathbf{g}_n\}$. Hence, $x_1 = \lambda\{\mathbf{g}_1\} = \frac{1}{\sqrt{2}}(1+1j)$ and $x_2 = \lambda\{\mathbf{g}_2\} = \frac{3}{\sqrt{10}}(1-1j)$.

Therefore, the NOMA symbol is

$$\begin{aligned} x &= \sqrt{\alpha_1}x_1 + \sqrt{\alpha_2}x_2 \\ &= A_{10} + 1jA_{10} + A_{03} - 1jA_{03} \\ &= A_{13} - 1jA_{\grave{1}3} = s_{26} \end{aligned} \tag{5.15}$$

with the NOMA bits block being $\mathbf{b} = [0,1,1,0,1,0]$ that is shown above the NOMA symbol in Figure 5.1b with the italic text. Therefore, the disjoint- and joint-multiuser Gray-mapping map the information bits to two different points on the constellation diagram.

5.4 BER Analysis of Joint-multiuser Gray-Mapping

The BER analysis of NOMA with disjoint-multiuser Gray-mapping is illustrated in [24], where the BER is derived for an arbitrary number of users and modulation orders. The analysis relies on the NOMA constellation diagram where the decision regions' boundaries for each bit in the NOMA bits block are found. Then for a given transmitted NOMA symbol, the probability that this symbol falls in the incorrect decision region is found. Due to the constellation symmetry, BER is computed by considering the first quadrant symbols only.

Following the same framework in [24], the conditional BER of NOMA with joint-multiuser Gray-mapping can be derived. It is worth mentioning that such a framework is general, and is even applicable to single-user systems. However, deriving the exact BER is unique for each particular system. To illustrate the conditional BER analysis, selected examples for the two-user case are considered.

5.4.1 $[M_1, M_2] = [2,2]$

The NOMA constellation for this toy example consists of 4 constellation points which are $s_0 = A_{\grave{1}\grave{1}}$, $s_1 = A_{11}$, $s_2 = A_{1\grave{1}}$ and $s_3 = A_{\grave{1}1}$. The index in s represents the decimal representation of the NOMA block bits. Due to symmetry, the first quadrant is sufficient to compute the BER. Hence, the transmission of s_1 and s_3 is considered. Consequently, the BER of U_n can be found by averaging the conditional BERs given

s_1 or s_3 are transmitted, i.e.,

$$P_n = \frac{1}{2}\left(P_n|s_1 + P_n|s_3\right). \tag{5.16}$$

5.4.1.1 U_1 analysis

Since U_1 is assigned BPSK, then it has one bit denoted as $b_1^{(1)}$, then $P_1^{(1)}$ can be found by considering the following cases:

Case 1: The BS transmits s_1. The received signal is erroneously detected if it falls in the incorrect decision region. For this case, there is one decision region that is bounded by A_{01} and $-A_{01}$. Consequently, the BER for this case can be written as

$$\begin{aligned} P_1|s_1 &= \Pr\left(-A_{12} < w_n < -A_{10}\right) \\ &= Q\left(A_{10}\sqrt{\gamma_n}\right) - Q\left(A_{12}\sqrt{\gamma_n}\right) \end{aligned} \tag{5.17}$$

where for notational simplicity w_n refers to either the real or imaginary part of the AWGN, which can be recognized from the context.

Case 2: The BS transmits s_3. It is noted that there are two incorrect decision regions for this case. The first is bounded by A_{01} and ∞, while the second is bounded by $-A_{01}$ and $-\infty$. Consequently, the BER for this case can be written as

$$\begin{aligned} P_1|s_3 &= \Pr\left(w_n > A_{10}\right) + \Pr\left(w_n < -A_{\grave{1}2}\right) \\ &= Q\left(A_{10}\sqrt{\gamma_n}\right) + Q\left(A_{\grave{1}2}\sqrt{\gamma_n}\right). \end{aligned} \tag{5.18}$$

Therefore, the conditional BER of U_1 can be expressed as

$$P_1 = Q\left(A_{10}\sqrt{\gamma_n}\right) + \frac{1}{2}Q\left(A_{\grave{1}2}\sqrt{\gamma_n}\right) - \frac{1}{2}Q\left(A_{12}\sqrt{\gamma_n}\right). \tag{5.19}$$

5.4.1.2 U_2 analysis

Since U_2 is assigned BPSK, then it has one bit denoted as $b_2^{(1)}$, then $P_2^{(1)}$ can be found by considering the following cases:

Case 1: The BS transmits s_1. It is noted that there is one incorrect decision region for this case which is bounded by 0 and $-\infty$. Consequently, the BER for this case can be written as

$$P_2|s_1 = \Pr\left(w_n < -A_{11}\right) = Q\left(A_{11}\sqrt{\gamma_n}\right). \tag{5.20}$$

Case 2: The BS transmits s_3. It is noted that there is one incorrect decision region

for the case which is found to be identical to the previous case. Hence, the BER for this case is

$$P_2|s_3 = \Pr\left(w_n < -A_{\grave{1}1}\right) = Q\left(A_{\grave{1}1}\sqrt{\gamma_n}\right). \tag{5.21}$$

Therefore, the conditional BER of U_2 can be expressed as

$$P_2 = \frac{1}{2}Q\left(A_{11}\sqrt{\gamma_n}\right) + \frac{1}{2}Q\left(A_{\grave{1}1}\sqrt{\gamma_n}\right). \tag{5.22}$$

5.4.2 $[M_1, M_2] = [4, 16]$

The NOMA constellation for this example is shown in in Figure 5.1b. Following the same approach, the BER for both users can be analyzed as follows.

5.4.2.1 U_1 analysis

Since U_1 is assigned QPSK, then it has two bits. Consequently, the two bits have to be analyzed independently. Starting off with $b_1^{(1)}$, then $P_1^{(1)}$ can be found by considering the following cases:

Case 1: The BS transmits s_8. The received signal is erroneously detected if it falls in the incorrect decision regions. For this case, there are two decision regions, the first is bounded by A_{01} and A_{03}, while the second is bounded by $-A_{01}$ and $-A_{03}$. Consequently, the BER for this case is expressed as

$$\begin{aligned} P_1^{(1)}|s_8 &= \Pr\left(-A_{12} < w_n < -A_{10}\right) + \Pr\left(-A_{16} < w_n < -A_{14}\right) \\ &= Q\left(A_{10}\sqrt{\gamma_n}\right) - Q\left(A_{12}\sqrt{\gamma_n}\right) + Q\left(A_{14}\sqrt{\gamma_n}\right) - Q\left(A_{16}\sqrt{\gamma_n}\right). \end{aligned} \tag{5.23}$$

It is noted that $P_1^{(1)}|s_8 = P_1^{(1)}|s_{24} = P_1^{(1)}|s_{25} = P_1^{(1)}|s_9$.

Case 2: The BS transmits s_{40}. The received signal is erroneously detected if it falls in the incorrect decision regions. For this case, there are three decision regions, the first is bounded by A_{03} and ∞, the second is bounded by A_{01} and $-A_{01}$, while the third is bounded by $-A_{03}$ and $-\infty$. Consequently, the BER for this case is expressed as

$$\begin{aligned} P_1^{(1)}|s_{40} &= \Pr\left(w_n > A_{10}\right) + \Pr\left(-A_{\acute{1}4} < w_n < -A_{\acute{1}2}\right) + \Pr\left(w_n < -A_{\acute{1}6}\right) \\ &= Q\left(A_{10}\sqrt{\gamma_n}\right) + Q\left(A_{\acute{1}2}\sqrt{\gamma_n}\right) - Q\left(A_{\acute{1}4}\sqrt{\gamma_n}\right) + Q\left(A_{\acute{1}6}\sqrt{\gamma_n}\right). \end{aligned} \tag{5.24}$$

It is noted that $P_1^{(1)}|s_{40} = P_1^{(1)}|s_{56} = P_1^{(1)}|s_{57} = P_1^{(1)}|s_{41}$.

Case 3: The BS transmits s_{44}. The received signal is erroneously detected if it falls

in the incorrect decision regions. For this case, there are three decision regions which are identical to Case 3. Consequently, the BER for this case is expressed as

$$\begin{aligned} P_1^{(1)}|s_{44} &= \Pr\left(w_n > A_{\acute{1}2}\right) + \Pr\left(-A_{12} < w_n < -A_{10}\right) + \Pr\left(w_n < -A_{14}\right) \\ &= Q\left(A_{\acute{1}2}\sqrt{\gamma_n}\right) + Q\left(A_{10}\sqrt{\gamma_n}\right) - Q\left(A_{12}\sqrt{\gamma_n}\right) + Q\left(A_{14}\sqrt{\gamma_n}\right). \end{aligned} \tag{5.25}$$

It is noted that $P_1^{(1)}|s_{44} = P_1^{(1)}|s_{60} = P_1^{(1)}|s_{61} = P_1^{(1)}|s_{45}$.

Case 4: The BS transmits s_{12}. The received signal is erroneously detected if it falls in the incorrect decision regions. For this case, there are two decision regions which are identical to Case 1. Consequently, the BER for this case is expressed as

$$\begin{aligned} P_1^{(1)}|s_{12} &= \Pr\left(A_{10} < w_n < A_{12}\right) + \Pr\left(-A_{\acute{1}4} < w_n < -A_{\acute{1}2}\right) \\ &= Q\left(A_{10}\sqrt{\gamma_n}\right) - Q\left(A_{12}\sqrt{\gamma_n}\right) + Q\left(A_{\acute{1}2}\sqrt{\gamma_n}\right) + Q\left(A_{\acute{1}4}\sqrt{\gamma_n}\right). \end{aligned} \tag{5.26}$$

It is noted that $P_1^{(1)}|s_{12} = P_1^{(1)}|s_{28} = P_1^{(1)}|s_{29} = P_1^{(1)}|s_{13}$.

Consequently, the BER for $b_1^{(1)}$ can be written as

$$P_1^{(1)} = \frac{1}{4}\left(P_1^{(1)}|s_8 + P_1^{(1)}|s_{40} + P_1^{(1)}|s_{44} + P_1^{(1)}|s_{12}\right) \tag{5.27}$$

which can be written as

$$\begin{aligned} P_1^{(1)} = {} & Q\left(A_{10}\sqrt{\gamma_n}\right) + \frac{3}{4}Q\left(A_{\acute{1}2}\sqrt{\gamma_n}\right) - \frac{3}{4}Q\left(A_{12}\sqrt{\gamma_n}\right) \\ & + \frac{2}{4}Q\left(A_{14}\sqrt{\gamma_n}\right) - \frac{2}{4}Q\left(A_{\acute{1}4}\sqrt{\gamma_n}\right) + \frac{1}{4}Q\left(A_{\acute{1}6}\sqrt{\gamma_n}\right) - \frac{1}{4}Q\left(A_{16}\sqrt{\gamma_n}\right). \end{aligned} \tag{5.28}$$

Due to symmetry, it is found that $P_1^{(1)} = P_1^{(2)}$. Therefore, the BER of U_1 is found to be identical to $P_1^{(1)}$, i.e., $P_1 = P_1^{(1)}$.

5.4.2.2 U_2 analysis

Since U_2 is assigned 16-QAM, then it has four bits. Consequently, the four bits must be analyzed independently. Starting off with $b_2^{(1)}$, then $P_2^{(1)}$ can be found by considering the following cases:

Case 1: The BS transmits s_8. The received signal is erroneously detected if it falls in the incorrect decision regions. For this case, there is one decision region, which is

bounded by 0 and $-\infty$. Consequently, the BER for this case is expressed as

$$P_2^{(1)}|s_8 = \Pr\left(w_n < -A_{13}\right) = Q\left(A_{13}\sqrt{\gamma_n}\right). \tag{5.29}$$

It is noted that $P_2^{(1)}|s_8 = P_2^{(1)}|s_{24} = P_2^{(1)}|s_{25} = P_2^{(1)}|s_9$.

Case 2: The BS transmits s_{40}. The received signal is erroneously detected if it falls in the incorrect decision regions. The decision region for this case is identical to Case 1. Consequently, the BER for this case is expressed as

$$P_2^{(1)}|s_{40} = \Pr\left(w_n < -A_{\acute{1}3}\right) = Q\left(A_{\acute{1}3}\sqrt{\gamma_n}\right). \tag{5.30}$$

It is noted that $P_2^{(1)}|s_{40} = P_2^{(1)}|s_{56} = P_2^{(1)}|s_{57} = P_2^{(1)}|s_{41}$.

Case 3: The BS transmits s_{44}. The received signal is erroneously detected if it falls in the incorrect decision regions. The decision region for this case is identical to Case 1. Consequently, the BER for this case is expressed as

$$P_2^{(1)}|s_{44} = \Pr\left(w_n < -A_{11}\right) = Q\left(A_{11}\sqrt{\gamma_n}\right). \tag{5.31}$$

It is noted that $P_2^{(1)}|s_{44} = P_2^{(1)}|s_{60} = P_2^{(1)}|s_{61} = P_2^{(1)}|s_{45}$.

Case 4: The BS transmits s_{12}. The received signal is erroneously detected if it falls in the incorrect decision regions. The decision region for this case is identical to Case 1. Consequently, the BER for this case is expressed as

$$P_2^{(1)}|s_{12} = \Pr\left(w_n < -A_{11}\right) = Q\left(A_{\acute{1}1}\sqrt{\gamma_n}\right). \tag{5.32}$$

It is noted that $P_2^{(1)}|s_{12} = P_2^{(1)}|s_{28} = P_2^{(1)}|s_{29} = P_2^{(1)}|s_{13}$.

Consequently, the BER for $b_2^{(1)}$ can be written as

$$P_2^{(1)} = \frac{1}{4}\left(P_2^{(1)}|s_8 + P_2^{(1)}|s_{40} + P_2^{(1)}|s_{44} + P_2^{(1)}|s_{12}\right) \tag{5.33}$$

which can be written as

$$P_2^{(1)} = \frac{1}{4}\left(Q\left(A_{\acute{1}1}\sqrt{\gamma_n}\right) + Q\left(A_{11}\sqrt{\gamma_n}\right) + Q\left(A_{13}\sqrt{\gamma_2}\right) + Q\left(A_{\acute{1}3}\sqrt{\gamma_n}\right)\right). \tag{5.34}$$

Due to symmetry, it is found that $P_2^{(1)} = P_2^{(3)}$. When considering $b_2^{(2)}$, then $P_2^{(2)}$ can be found by considering the following cases:

Case 1: The BS transmits s_8. The received signal is erroneously detected if it falls

in the incorrect decision regions. For this case, there is one decision region, which is bounded by A_{02} and $-A_{02}$. Consequently, the BER for this case is expressed as

$$\begin{aligned} P_2^{(2)}|s_8 &= \Pr\left(-A_{15} < w_n < -A_{11}\right) \\ &= Q\left(A_{11}\sqrt{\gamma_n}\right) - Q\left(A_{15}\sqrt{\gamma_n}\right). \end{aligned} \tag{5.35}$$

It is noted that $P_2^{(2)}|s_8 = P_2^{(2)}|s_{24} = P_2^{(2)}|s_{25} = P_2^{(2)}|s_9$.

Case 2: The BS transmits s_{40}. The received signal is erroneously detected if it falls in the incorrect decision regions. The decision region for this case is identical to Case 1. Consequently, the BER for this case is expressed as

$$\begin{aligned} P_2^{(1)}|s_{40} &= \Pr\left(-A_{\acute{1}5} < w_n < -A_{\acute{1}1}\right) \\ &= Q\left(A_{\acute{1}1}\sqrt{\gamma_n}\right) - Q\left(A_{\acute{1}5}\sqrt{\gamma_n}\right). \end{aligned} \tag{5.36}$$

It is noted that $P_2^{(2)}|s_{40} = P_2^{(2)}|s_{56} = P_2^{(2)}|s_{57} = P_2^{(2)}|s_{41}$.

Case 3: The BS transmits s_{44}. The received signal is erroneously detected if it falls in the incorrect decision regions. For this case, there are two decision regions, the first is bounded by A_{02} and ∞, while the second is bounded by $-A_{02}$ and $-\infty$ Consequently, the BER for this case is expressed as

$$\begin{aligned} P_2^{(2)}|s_{44} &= \Pr\left(w_n > A_{\acute{1}1}\right) + \Pr\left(w_n < -A_{13}\right) \\ &= Q\left(A_{\acute{1}1}\sqrt{\gamma_n}\right) + Q\left(A_{13}\sqrt{\gamma_n}\right). \end{aligned} \tag{5.37}$$

It is noted that $P_2^{(2)}|s_{44} = P_2^{(2)}|s_{60} = P_2^{(2)}|s_{61} = P_2^{(2)}|s_{45}$.

Case 4: The BS transmits s_{12}. The received signal is erroneously detected if it falls in the incorrect decision regions. The decision region for this case is identical to Case 3. Consequently, the BER for this case is expressed as

$$\begin{aligned} P_2^{(2)}|s_{12} &= \Pr\left(w_n > A_{11}\right) + \Pr\left(w_n < -A_{\acute{1}3}\right) \\ &= Q\left(A_{11}\sqrt{\gamma_n}\right) + Q\left(A_{\acute{1}3}\sqrt{\gamma_n}\right). \end{aligned} \tag{5.38}$$

It is noted that $P_2^{(2)}|s_{12} = P_2^{(2)}|s_{28} = P_2^{(2)}|s_{29} = P_2^{(2)}|s_{13}$. Consequently, the BER for $b_2^{(2)}$ can be written as

$$P_2^{(2)} = \frac{1}{4}\left(P_2^{(2)}|s_8 + P_2^{(2)}|s_{40} + P_2^{(2)}|s_{44} + P_2^{(2)}|s_{12}\right) \tag{5.39}$$

Table 5.1: $\mathbf{g}_n$ per bit for U_1, $N = 2$.

$[M_1, M_2]$	$b_1^{(k)}$	n	$\mathbf{g}_n$
$[4,4]$	$b_1^{(1)}$	1	$1,\grave{1},1$
		2	$0,2,2$
$[4,16]$	$b_1^{(1)}$	1	$1,\grave{1},1,\grave{1},1,\grave{1},1$
		2	$0,2,2,4,4,6,6$
$[4,64]$	$b_1^{(1)}$	1	$1,\grave{1},1,\grave{1},1,\grave{1},1,\grave{1},1,\grave{1},1,\grave{1},1,\grave{1},1$
		2	$0,2,2,4,4,6,6,8,8,10,10,12,12,14,14$
$[16,4]$	$b_1^{(1)}$	1	$1,3,\grave{3},\grave{1},1,3$
		2	$0,0,2,2,2,2$
$[16,4]$	$b_1^{(2)}$	1	$1,3,5,\grave{5},\grave{3},\grave{1},1,3,5$
		2	$0,0,0,2,2,2,2,2,2$
$[16,16]$	$b_1^{(1)}$	1	$1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3$
		2	$0,0,2,2,2,2,4,4,4,4,6,6,6,6$
$[16,16]$	$b_1^{(2)}$	1	$1,3,5,\grave{5},\grave{3},\grave{1},1,3,5,\grave{5},\grave{3},\grave{1},1,3,5,\grave{5},\grave{3},\grave{1},1,3,5$
		2	$0,0,0,2,2,2,2,2,2,4,4,4,4,4,4,6,6,6,6,6,6$
$[16,64]$	$b_1^{(1)}$	1	$1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3$
		2	$0,0,2,2,2,2,4,4,4,4,6,6,6,6,8,8,8,8,10,10,10,10,12,12,12,12,14,14,14,14$
$[16,64]$	$b_1^{(2)}$	1	$1,3,5,\grave{5},\grave{3},\grave{1},1,3,5,\grave{5},\grave{3},\grave{1},1,3,5,\grave{5},\grave{3},\grave{1},1,3,5,\grave{5},\grave{3},\grave{1},1,3,5,\grave{5},\grave{3},\grave{1},1,3,5,\grave{5},\grave{3},\grave{1},1,3,5$
		2	$0,0,0,2,2,2,2,2,2,4,4,4,4,4,4,6,6,6,6,6,6,8,8,8,8,8,8,10,10,10,10,10,10,12,12,12,12,12,12,14,14,14,14,14,14$
$[64,4]$	$b_1^{(1)}$	1	$1,3,5,7,\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7$
		2	$0,0,0,0,2,2,2,2,2,2,2,2$
$[64,4]$	$b_1^{(2)}$	1	$1,3,5,7,9,11,\grave{11},\grave{9},\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7,9,11$
		2	$0,0,0,0,0,0,2,2,2,2,2,2,2,2,2,2,2,2$
$[64,4]$	$b_1^{(3)}$	1	$1,3,5,7,9,11,13,\grave{13},\grave{11},\grave{9},\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7,9,11,13$
		2	$0,0,0,0,0,0,0,2,2,2,2,2,2,2,2,2,2,2,2,2,2$
$[64,16]$	$b_1^{(1)}$	1	$1,3,5,7,\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7,\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7,\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7$
		2	$0,0,0,0,2,2,2,2,2,2,2,2,4,4,4,4,4,4,4,4,6,6,6,6,6,6,6,6$
$[64,16]$	$b_1^{(2)}$	1	$1,3,5,7,9,11,\grave{11},\grave{9},\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7,9,11,\grave{11},\grave{9},\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7,9,11,\grave{11},\grave{9},\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7,9,11$
		2	$0,0,0,0,0,0,2,2,2,2,2,2,2,2,2,2,2,2,4,4,4,4,4,4,4,4,4,4,4,4,6,6,6,6,6,6,6,6,6,6,6,6$
$[64,16]$	$b_1^{(3)}$	1	$1,3,5,7,9,11,13,\grave{13},\grave{11},\grave{9},\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7,9,11,13,\grave{13},\grave{11},\grave{9},\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7,9,11,13,\grave{13},\grave{11},\grave{9},\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7,9,11,13$
		2	$0,0,0,0,0,0,0,2,2,2,2,2,2,2,2,2,2,2,2,2,2,4,4,4,4,4,4,4,4,4,4,4,4,4,6,6,6,6,6,6,6,6,6,6,6,6,6,6$

which can be written as

$$P_2^{(2)} = \frac{2}{4}Q\left(A_{11}\sqrt{\gamma_n}\right) + \frac{2}{4}Q\left(A_{\grave{1}1}\sqrt{\gamma_n}\right) + \frac{1}{4}Q\left(A_{13}\sqrt{\gamma_n}\right) + \frac{1}{4}Q\left(A_{\grave{1}3}\sqrt{\gamma_n}\right) - \frac{1}{4}Q\left(A_{15}\sqrt{\gamma_n}\right) - \frac{1}{4}Q\left(A_{\grave{1}5}\sqrt{\gamma_n}\right). \quad (5.40)$$

Due to symmetry, it is found that $P_2^{(2)} = P_2^{(4)}$. Therefore, the BER for U_2 can be written as

$$P_2 = \frac{1}{2}\left(P_2^{(1)} + P_2^{(2)}\right) \quad (5.41)$$

which can be written as

$$P_2 = \frac{3}{8}Q\left(A_{11}\sqrt{\gamma_n}\right) + \frac{3}{8}Q\left(A_{\grave{1}1}\sqrt{\gamma_n}\right) + \frac{2}{8}Q\left(A_{13}\sqrt{\gamma_n}\right) + \frac{2}{8}Q\left(A_{\grave{1}3}\sqrt{\gamma_n}\right) - \frac{1}{8}Q\left(A_{15}\sqrt{\gamma_n}\right) - \frac{1}{8}Q\left(A_{\grave{1}5}\sqrt{\gamma_n}\right). \quad (5.42)$$

Table 5.2: $\mathbf{c}$ per bit for U_1, $N = 2$.

$[M_1, M_2]$	$b_1^{(k)}$	β	$\mathbf{c}$
$[4,4]$	$b_1^{(1)}$	2	$2,1,\grave{1}$
$[4,16]$	$b_1^{(1)}$	4	$4,3,\grave{3},\grave{2},2,1,\grave{1}$
$[4,64]$	$b_1^{(1)}$	8	$8,7,\grave{7},\grave{6},6,5,\grave{5},\grave{4},4,3,\grave{3},\grave{2},2,1,\grave{1}$
$[16,4]$	$b_1^{(1)}$	4	$2,2,1,1,\grave{1},\grave{1}$
$[16,4]$	$b_1^{(2)}$	4	$4,2,\grave{2},1,\grave{1},\grave{2},2,1,\grave{1}$
$[16,16]$	$b_1^{(1)}$	8	$4,4,3,3,\grave{3},\grave{3},\grave{2},\grave{2},2,2,1,1,\grave{1},\grave{1}$
$[16,16]$	$b_1^{(2)}$	8	$8,4,\grave{4},3,\grave{3},\grave{6},6,3,\grave{3},2,\grave{2},\grave{4},4,2,\grave{2},1,\grave{1},\grave{2},2,1,\grave{1}$
$[16,64]$	$b_1^{(1)}$	16	$8,8,7,7,\grave{7},\grave{7},\grave{6},\grave{6},6,6,5,5,\grave{5},\grave{5},\grave{4},\grave{4},4,4,3,3,\grave{3},\grave{3},\grave{2},\grave{2},2,2,1,1,\grave{1},\grave{1}$
$[16,64]$	$b_1^{(2)}$	16	$16,8,\grave{8},7,\grave{7},\grave{14},14,7,\grave{7},6,\grave{6},\grave{12},12,6,\grave{6},5,\grave{5},\grave{10},10,5,\grave{5},4,\grave{4},\grave{8},8,4,\grave{4},3,\grave{3},\grave{6},6,3,\grave{3},2,\grave{2},\grave{4},4,2,\grave{2},1,\grave{1},\grave{2},2,1,\grave{1}$
$[64,4]$	$b_1^{(1)}$	8	$2,2,2,2,1,1,1,1,\grave{1},\grave{1},\grave{1},\grave{1}$
$[64,4]$	$b_1^{(2)}$	8	$4,4,2,2,\grave{2},\grave{2},1,1,\grave{1},\grave{1},\grave{2},\grave{2},2,2,1,1,\grave{1},\grave{1}$
$[64,4]$	$b_1^{(3)}$	8	$8,6,\grave{6},\grave{4},4,2,\grave{2},1,\grave{1},\grave{2},2,3,\grave{3},\grave{4},4,3,\grave{3},\grave{2},2,1,\grave{1}$
$[64,16]$	$b_1^{(1)}$	16	$4,4,4,4,3,3,3,3,\grave{3},\grave{3},\grave{3},\grave{3},\grave{2},\grave{2},\grave{2},\grave{2},2,2,2,2,1,1,1,1,\grave{1},\grave{1},\grave{1},\grave{1}$
$[64,16]$	$b_1^{(2)}$	16	$8,8,4,4,\grave{4},\grave{4},3,3,\grave{3},\grave{3},\grave{3},\grave{3},3,3,3,3,\grave{3},\grave{3},2,2,\grave{2},\grave{2},\grave{2},\grave{2},2,2,2,2,\grave{2},\grave{2},1,1,\grave{1},\grave{1},\grave{1},\grave{1},1,1,1,1,\grave{1},\grave{1}$
$[64,16]$	$b_1^{(3)}$	16	$16,12,\grave{12},\grave{8},8,4,\grave{4},3,\grave{3},\grave{6},6,9,\grave{9},\grave{12},12,9,\grave{9},\grave{6},6,3,\grave{3},2,\grave{2},\grave{4},4,6,\grave{6},\grave{8},8,6,\grave{6},\grave{4},4,2,\grave{2},1,\grave{1},\grave{2},2,3,\grave{3},\grave{4},4,3,\grave{3},\grave{2},2,1,\grave{1}$

5.4.3 BER Generalization

By inference, the conditional BER expression of U_n for the kth bit can be written in general as [124, Eq. (9)],

$$P_n^{(k)} = \frac{1}{\beta} \sum_j c_j Q\left(\Delta_j \sqrt{\gamma_n}\right) \tag{5.43}$$

where c_j, β and $\Delta_j = g_{1,j}\sqrt{\frac{\alpha_1}{\kappa_1}} + g_{2,j}\sqrt{\frac{\alpha_2}{\kappa_2}}$ are constants with a certain pattern. Tables 5.1 and 5.2 show the pattern of $\mathbf{g}_n$, $\mathbf{c}_n$ and β for U_1 considering the square M-ary quadrature amplitude modulation (M-QAM). Similarly, Tables 5.3 and 5.4 illustrates the same for U_2. By observing the regular pattern of β, it can be written in a closed-form as $\beta = \frac{\sqrt{M_1 M_2}}{2}$. In the following, the regular pattern of $\mathbf{g}_n$ and $\mathbf{c}_n$ will describe the BER expressions for both users.

5.4.3.1 U_1 general expression

As can be noted from Tables 5.1 and 5.2, the per bit BER expressions involve two parts which are: 1) The single-user BER which is interference-free, more specifically the terms where $\mathbf{g}_2$ goes to zero. 2) The SIC error propagation due to the interference. Consequently, it can be expressed as

$$P_1^{(k)} = P_{1,\text{SU}}^{(k)} + P_{1,\text{SIC}}^{(k)}. \tag{5.44}$$

While focusing on the interference-free terms, it can be seen that $\mathbf{g}_1$ always starts with 1 and increases by 2 depending on M_1 and the bit being observed. Therefore, its pattern

Table 5.3: $\mathbf{g}_n$ per bit for U_2, $N = 2$.

$[M_1, M_2]$	b_n	n	$\mathbf{g}_n$
$[4,4]$	$b_2^{(1)}$	1	$\grave{1},1$
		2	$1,1$
$[4,16]$	$b_2^{(1)}$	1	$\grave{1},1,\grave{1},1$
		2	$1,1,3,3$
$[4,16]$	$b_2^{(2)}$	1	$\grave{1},1,\grave{1},1,\grave{1},1$
		2	$1,1,3,3,5,5$
$[4,64]$	$b_2^{(1)}$	1	$\grave{1},1,\grave{1},1,\grave{1},1,\grave{1},1$
		2	$1,1,3,3,5,5,7,7$
$[4,64]$	$b_2^{(2)}$	1	$\grave{1},1,\grave{1},1,\grave{1},1,\grave{1},1,\grave{1},1,\grave{1},1$
		2	$1,1,3,3,5,5,7,7,9,9,11,11$
$[4,64]$	$b_2^{(3)}$	1	$\grave{1},1,\grave{1},1,\grave{1},1,\grave{1},1,\grave{1},1,\grave{1},1,\grave{1},1$
		2	$1,1,3,3,5,5,7,7,9,9,11,11,13,13$
$[16,4]$	$b_2^{(1)}$	1	$\grave{3},\grave{1},1,3$
		2	$1,1,1,1$
$[16,16]$	$b_2^{(1)}$	1	$\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3$
		2	$1,1,1,1,3,3,3,3$
$[16,16]$	$b_2^{(2)}$	1	$\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3$
		2	$1,1,1,1,3,3,3,3,5,5,5,5$
$[16,64]$	$b_2^{(1)}$	1	$\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3$
		2	$1,1,1,1,3,3,3,3,5,5,5,5,7,7,7,7$
$[16,64]$	$b_2^{(2)}$	1	$\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3$
		2	$1,1,1,1,3,3,3,3,5,5,5,5,7,7,7,7,9,9,9,9,11,11,11,11$
$[16,64]$	$b_2^{(3)}$	1	$\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3,\grave{3},\grave{1},1,3$
		2	$1,1,1,1,3,3,3,3,5,5,5,5,7,7,7,7,9,9,9,9,11,11,11,11,13,13,13,13$
$[64,4]$	$b_2^{(1)}$	1	$\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7$
		2	$1,1,1,1,1,1,1,1$
$[64,16]$	$b_2^{(1)}$	1	$\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7,\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7$
		2	$1,1,1,1,1,1,1,1,3,3,3,3,3,3,3,3$
$[64,16]$	$b_2^{(2)}$	1	$\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7,\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7,\grave{7},\grave{5},\grave{3},\grave{1},1,3,5,7$
		2	$1,1,1,1,1,1,1,1,3,3,3,3,3,3,3,3,5,5,5,5,5,5,5,5$

can materialize the interference-free BER. Similarly, the pattern of $\mathbf{c}$ is identical to the single-user case reported in [138, Table I]. While noting [138, Eq. (14)], the interference-free BER term can be written as

$$P_{1,\mathrm{SU}}^{(k)} = \frac{2}{\sqrt{M_1}} \sum_{i=0}^{\left(1-2^{-k}\right)\sqrt{M_1}-1} \left(2^{k-1} - \left\lfloor \frac{i \times 2^{k-1}}{\sqrt{M_1}} + \frac{1}{2} \right\rfloor\right) (-1)^{\left\lfloor \frac{i\times 2^{k-1}}{\sqrt{M_1}} \right\rfloor} \times Q\left(A_{\delta_i,0}\sqrt{\gamma_1}\right) \quad (5.45)$$

where $\delta_i = 2i+1$. On the other hand, to express the BER term related to the error propagation from the imperfect SIC, the regular pattern of $\mathbf{g}_1$ is visualized to start from $1-\left(2-2^{1-k}\right)\sqrt{M_1}$ up to $\left(2-2^{1-k}\right)\sqrt{M_1}-1$ with an increment of 2 and it keeps repeating for $\sqrt{M_2}-1$ times. Furthermore, the regular pattern of $\mathbf{g}_2$ is visualized to

Table 5.4: $\mathbf{c}_n$ per bit for U_2, $N = 2$.

$[M_1, M_2]$	b_n	β	$\mathbf{c}$
$[4,4]$	$b_2^{(1)}$	2	$1,1$
$[4,16]$	$b_2^{(1)}$	4	$1,1,1,1$
$[4,16]$	$b_2^{(2)}$	4	$2,2,1,1,\grave{1},\grave{1}$
$[4,64]$	$b_2^{(1)}$	8	$1,1,1,1,1,1,1,1$
$[4,64]$	$b_2^{(2)}$	8	$2,2,2,2,1,1,1,1,\grave{1},\grave{1},\grave{1},\grave{1}$
$[4,64]$	$b_2^{(3)}$	8	$4,4,3,3,\grave{3},\grave{3},\grave{2},\grave{2},2,2,1,1,\grave{1},\grave{1}$
$[16,4]$	$b_2^{(1)}$	4	$1,1,1,1$
$[16,16]$	$b_2^{(1)}$	8	$1,1,1,1,1,1,1,1$
$[16,16]$	$b_2^{(2)}$	8	$2,2,2,2,1,1,1,1,\grave{1},\grave{1},\grave{1},\grave{1}$
$[16,64]$	$b_2^{(1)}$	16	$1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1$
$[16,64]$	$b_2^{(2)}$	16	$2,2,2,2,2,2,2,2,1,1,1,1,1,1,1,1,\grave{1},\grave{1},\grave{1},\grave{1},\grave{1},\grave{1},\grave{1},\grave{1}$
$[16,64]$	$b_2^{(3)}$	16	$4,4,4,4,3,3,3,3,\grave{3},\grave{3},\grave{3},\grave{3},\grave{2},\grave{2},\grave{2},\grave{2},2,2,2,2,2,1,1,1,1,\grave{1},\grave{1},\grave{1},\grave{1}$
$[64,4]$	$b_2^{(1)}$	8	$1,1,1,1,1,1,1,1$
$[64,16]$	$b_2^{(1)}$	16	$1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1$
$[64,16]$	$b_2^{(2)}$	16	$2,2,2,2,2,2,2,2,1,1,1,1,1,1,1,1,\grave{1},\grave{1},\grave{1},\grave{1},\grave{1},\grave{1},\grave{1},\grave{1}$

start by 2 and repeats for $\left(2-2^{1-k}\right)\sqrt{M_1}$ times, then it increases by 2 up to $2\sqrt{M_2}-2$. The $\mathbf{c}$ pattern is slightly complicated. Nonetheless, it is cyclic with a certain frequency, where the main component is a scaled coefficient from an arithmetic triangle with an order of 2^{k-1}. Consequently, the BER term related to the SIC error can be expressed as

$$P_{1,\mathrm{SIC}}^{(k)} = \frac{2}{\sqrt{M_1 M_2}} \sum_{m=1}^{\sqrt{M_2}-1} \sum_{i=0}^{\left(2-2^{1-k}\right)\sqrt{M_1}-1} (-1)^{k(m+1)} \times (-1)^{\left\lfloor \frac{2^k}{\sqrt{M_1}} - \log_{\max(2,k)}(k)\left[\frac{2^{k-1}}{\sqrt{M_1}}+\frac{1}{2}\right]\right\rfloor} \left(\sqrt{M_2}-m\right) \times \binom{2}{\left\lfloor \frac{i\times 2^{k-1}}{\sqrt{M_1}} \right\rfloor}_{2^{k-1}} Q\left(A_{\delta_i,\delta_m}\sqrt{\gamma_1}\right) \quad (5.46)$$

where $\binom{\cdot}{\cdot}_s$ is the coefficient from the generalized Pascal's triangle with order s, $\delta_i = 2i - \sqrt{M_1}\left(2-2^{1-k}\right) - 1$ and $\delta_m = 2m$.

5.4.3.2 U_2 general expression

It can be noted from Tables 5.3 and 5.4 that the pattern of per bit BER expressions can be described as follows. $\mathbf{g}_1$ starts from $1-\left(2-2^{1-k}\right)\sqrt{M_1}$ up to $\left(2-2^{1-k}\right)\sqrt{M_1}-1$, where it repeats for $\sqrt{M_2}\left(1-2^{-k}\right)$ times. In addition, $\mathbf{g}_2$ starts from 1 that repeats for $\left(2-2^{1-k}\right)\sqrt{M_1}$ times, then it increases by 2 up to $2\sqrt{M_2}\left(1-2^{-k}\right)-1$. The regular pattern of $\mathbf{c}$ is similar to that of the single-user case, nonetheless, it has to account for

the repetitions. Consequently, the per bit BER of U_2 can be expressed as

$$P_2^{(k)} = \frac{2}{\sqrt{M_1 M_2}} \sum_{m=1}^{\left(1-2^{-k}\right)\sqrt{M_2}} \sum_{i=0}^{\sqrt{M_1}-1} (-1)^{\left\lfloor \frac{(m-1)\times 2^{k-1}}{\sqrt{M_1}} \right\rfloor} \times \left(2^{k-1} - \left\lfloor \frac{(m-1)\times 2^{k-1}}{\sqrt{M_1}} + \frac{1}{2} \right\rfloor\right) Q\left(A_{\delta_i,\delta_m}\sqrt{\gamma_2}\right) \quad (5.47)$$

where $\delta_i = 2i - \sqrt{M_1}\left(2 - 2^{1-k}\right) - 1$ and $\delta_m = 2m - 1$.

5.4.3.3 Conditional BER

Finally, the NOMA conditional BER, P_n, can be computed by averaging the conditional BER for all bits such that

$$P_n = \frac{2}{\log_2(M_n)} \sum_{i=1}^{\log_2\left(\sqrt{M_n}\right)} P_n^{(k)}. \quad (5.48)$$

The average BER expression for the Rician fading channel is described in Appendix C.1. Furthermore, the performance gain of joint-multiuser Gray-mapping is analysed in Appendix C.2, where the proof is shown. While the presented analysis is only applicable for square QAM, interested readers can refer to the MATLAB code [139] to compute the symbolic instantaneous BER expressions for any number of layers and modulation orders.

5.5 Numerical Results and Discussions

This section presents the analytical and simulation results of the average BER for the disjoint-multiuser and joint-multiuser Gray-mapping NOMA systems considering $N \in \{2, 3, \cdots, 5\}$ for selected modulation orders. Furthermore, the PA falls within the bounds derived in [24, 27] to allow reliable SIC detection. We assume that U_n is located at a normalized distance of $d_n = 10^{\frac{3}{5\rho_n}(n-1)}$ from the BS. The large-scale fading is considered as fixed pathloss with an exponent of $\rho_n = 2.7$. The AWGN variance N_0 is assumed to be common for all receivers, and hence, the transmit SNR $\triangleq 1/N_0$ [24]. Unless otherwise stated, the Rician factor is fixed to $K_n = 5$ dB. Also, two PA criteria are considered in this work, 1) The equally spaced constellation PA [1] where the selected power, in this case, is denoted by $\mathcal{P}_1$. 2) The other PA aims at minimizing the average BER of U_1 while satisfying a certain threshold τ_n for the rest of receivers at

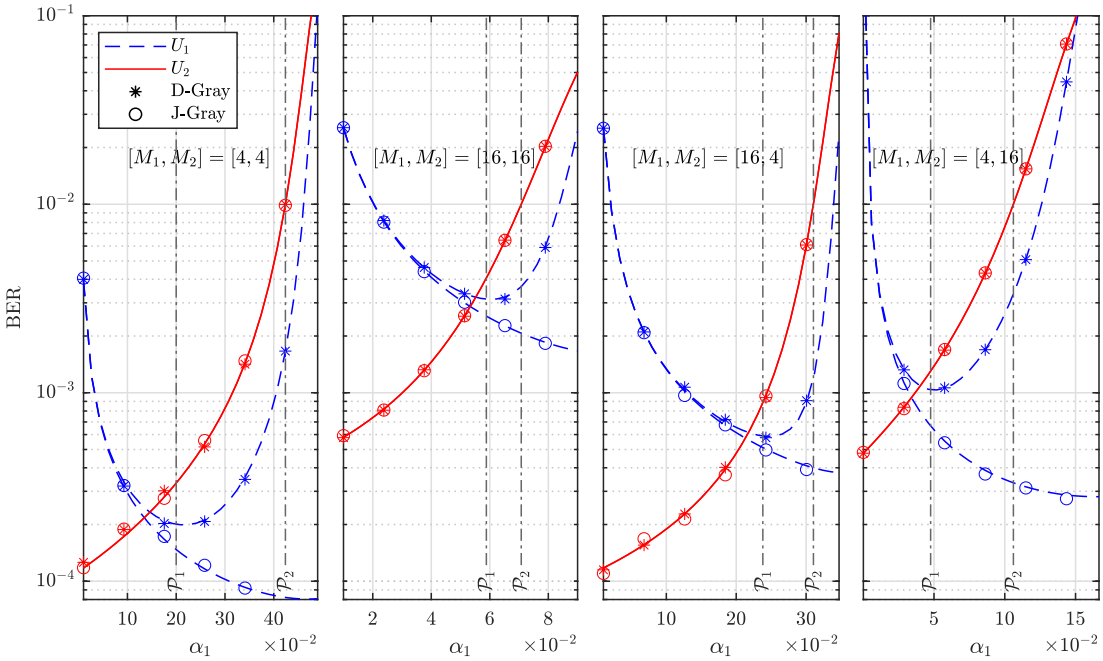


Figure 5.5: Average BER vs. α_1 for $N = 2$ considering $[M_1, M_2] = [4,4]$, $[16,16]$, $[16,4]$, $[4,16]$ and highlighting two power assignments (PAs), where SNR = 35 dB.

a given SNR, where the selected power is denoted by $\mathcal{P}_2$. Furthermore, the markers in the figures represent the simulation results, while the lines represent the analytical solution. Also, D-Gray and J-Gray stand for the disjoint-multiuser Gray-mapping and joint-multiuser Gray-mapping, respectively.

Figure 5.5 presents the average BER vs. α_1 for the two-user case considering SNR = 35 dB and the following modulation orders: $[M_1, M_2] = [4,4]$, $[16,16]$, $[16,4]$, $[4,16]$. The figure confirms that there is a match between the analytical and simulated results which will be seen consistently in the following figures. The x-axis represents α_1 such that $0 < \alpha_1 < \alpha_{1,\max}$ to ensure reliable SIC detection [24, 27], where $\alpha_{1,\max} = \frac{\kappa_1}{\kappa_1 + \kappa_2 \Lambda_1}$ and $\sqrt{\Lambda_n} = M_n^{(x)}$. Two PAs are highlighted in the figure, which are $\mathcal{P}_1$ and $\mathcal{P}_2$, where the former ensures equally spaced constellation points, and the latter minimizes P_1 while ensuring that $P_2 \leq 10^{-2}$. In general, $\mathcal{P}_1 < \mathcal{P}_2$ and U_1 BER performance of J-Gray is better than that of the D-Gray at $\mathcal{P}_1$ and $\mathcal{P}_2$. In addition, P_1 degrades whenever either user's modulation order increases. While D-Gray is convex, the J-Gray is monotonically decreasing with α_1. Consequently, resulting in a performance gain for the J-Gray with respect to the D-Gray, where the performance gain is seen to increase as α_1 increases indicating that the error propagation due to imperfect SIC is reduced with J-Gray. The reason for such a phenomenon can be explained by Figure 5.2, when α_1 increases, the Euclidean distance between x and

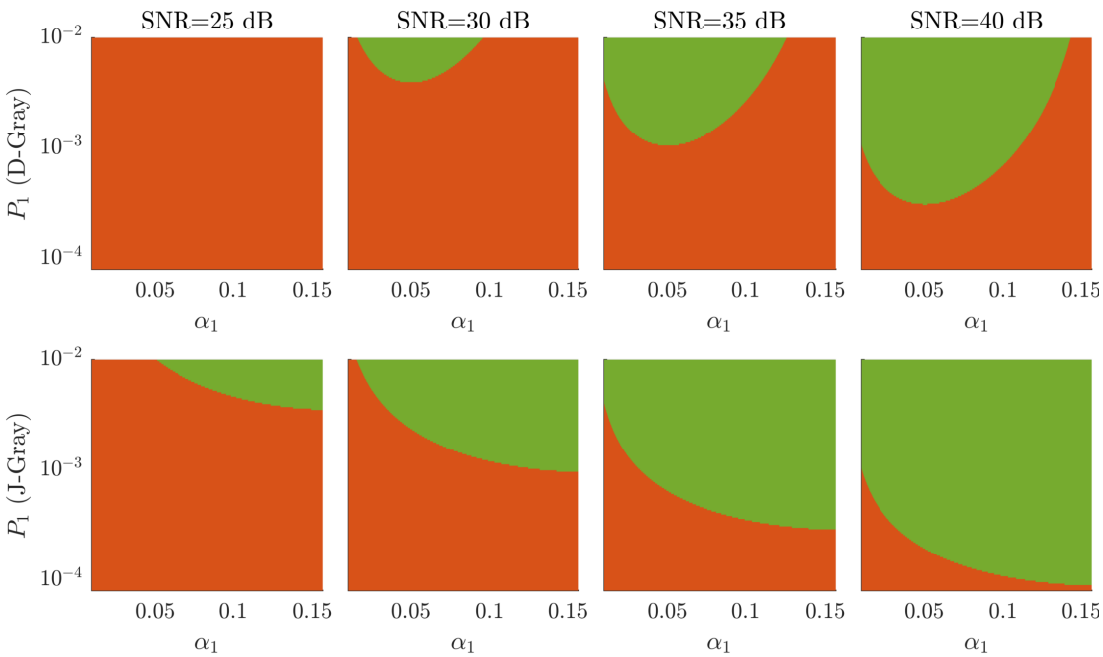


Figure 5.6: The feasibility region for the system over α_1- P_1 plane for different SNRs, where $N = 2$, $[M_1, M_2] = [4, 16]$ and U_2 receives on best-effort basis.

the y-axis becomes less. Hence, the error considering a cross-over for a D-Gray system becomes higher. Whereas, for the J-Gray system, such an error is mitigated as the cross-over with respect to the first point near the y-axis will not cause any error. On the other hand, U_2 performance is identical for both mapping schemes because the J-Gray scheme does not change the bit-to-symbol mapping with respect to U_2. Nonetheless, U_2 may gain from J-Gray by selecting appropriate PA that will allow sharing of the performance gain between both layers. For instance, when considering $[M_1, M_2] = [4, 4]$, $\mathcal{P}_1$ achieves $P_1 = 9 \times 10^{-3}$ for the D-Gray, while such a BER can be achieved by the J-Gray if we back-off slightly with the PA. Such a back-off will improve the BER of U_2 as α_2 has increased. Moreover, U_2 performance is almost independent from the U_1 modulation order which is understood as U_2 is given a higher power coefficient. Additionally, P_2 degrades as M_2 increases, which is aligned with the single-user case.

Figures 5.6 and 5.7 show the system feasibility region over the α_1–P_1 plane for the two-user case considering different SNRs, where $[M_1, M_2] = [4, 16]$ and P_1 spans from 10^{-2} to 10^{-4} which represents the practical BER range for the uncoded systems. It is worth mentioning that the figures use binary color coding to distinguish between the feasible and infeasible regions. Figure 5.6 considers that U_2 is received on the best-effort basis which is reasonable in the cases where its receiver is a delay-tolerant device and can trade re-transmissions for reliability by combining the re-transmission

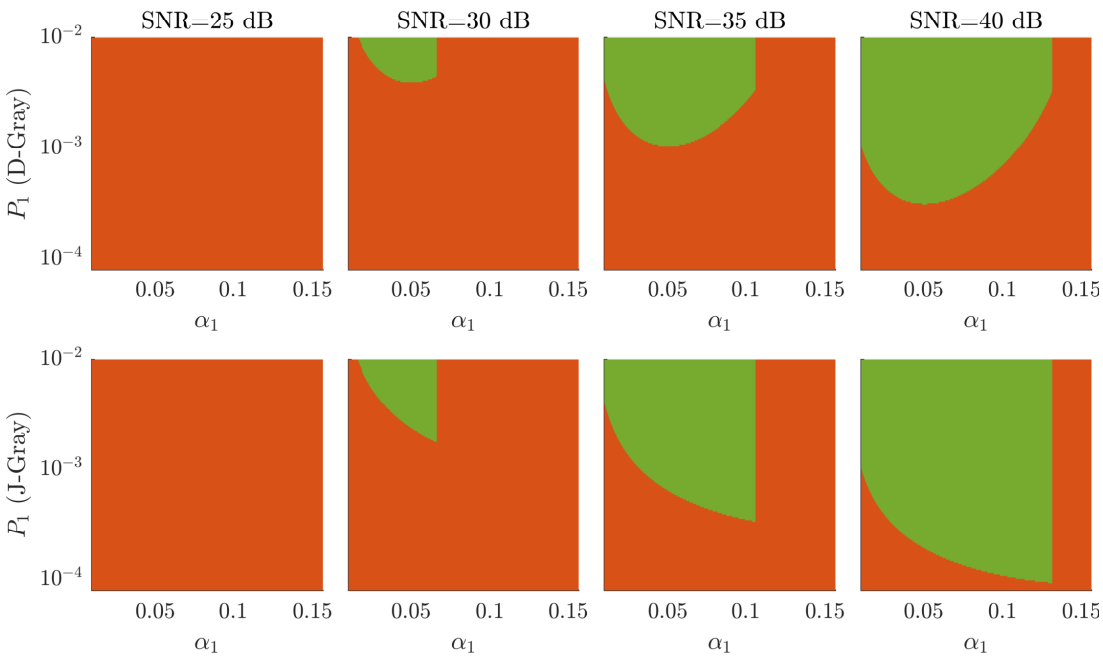


Figure 5.7: The feasibility region for the system over α_1- P_1 plane for different SNRs, where $N = 2$, $[M_1, M_2] = [4, 16]$ and P_2 threshold is 10^{-2}.

samples. The figures show that the feasibility area of the D-Gray system is convex with a global minimum, while the J-Gray system's feasibility area is monotonically decreasing as α_1 increases. It is seen that the feasibility region's area for both systems increases as the SNR grows, where stricter P_1 values can be satisfied. Also, it is noted that at an SNR of 25 dB the D-Gray is infeasible, unlike the J-Gray which is feasible. The relative improvement in the feasibility area of the J-Gray with respect to the D-Gray is around 400% at 30 dB. Additionally, the achieved relative improvement reduces to 127% and 72% for 35 dB and 40 dB.

Unlike Figure 5.6, Figure 5.7 considers that U_2 has a BER threshold of 10^{-2} to be satisfied. Similar conclusions are drawn here as well. Nonetheless, both mapping schemes are infeasible at 25 dB as $P_2 > 10^{-2}$ for the whole α_1–P_1 plane. The feasibility areas for both systems are reduced compared to the best-effort cases. Additionally, the relative improvement in the feasibility area of the J-Gray with respect to the D-Gray is around 53%, 44%, and 42% for 30 dB, 35 dB, and 40 dB.

Figure 5.8 presents the average BER of U_1 versus SNR for the two-user case considering the following modulation orders: $[M_1, M_2] = [4, 4]$, $[16, 4]$, $[4, 16]$, $[16, 16]$. Two PAs, $\mathcal{P}_1$ and $\mathcal{P}_2$, are considered to highlight the performance gain of the J-Gray. The results agree with the previous finding that the performance of U_1 is impacted

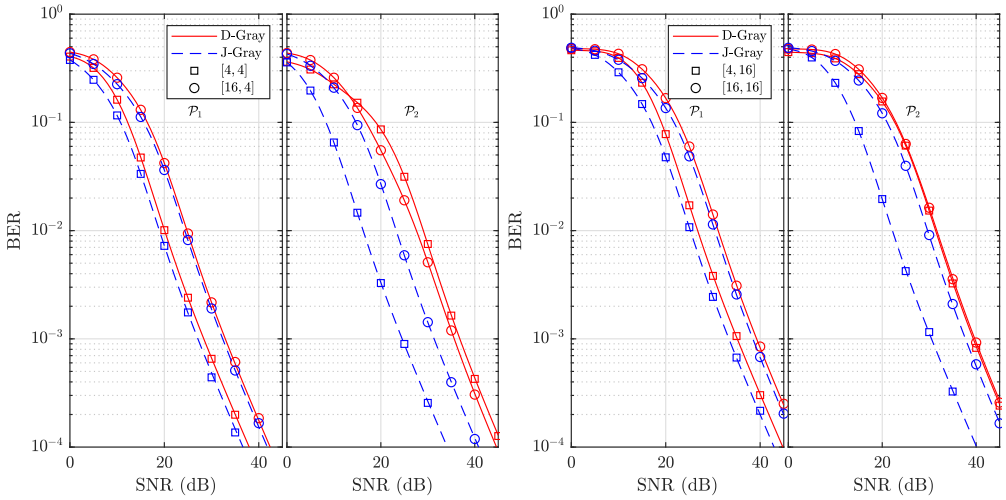


Figure 5.8: Average BER vs. SNR for U_1 considering $\mathcal{P}_1$, $\mathcal{P}_2$, where $N = 2$ and $[M_1, M_2] = [4,4]$, $[16,4]$, $[4,16]$, $[16,16]$.

by both users' modulation orders especially the modulation order of U_2 as the latter is given the highest power coefficient. This can be seen clearly at PA of $\mathcal{P}_1$ as the rank of P_1 performance from the best to the worst based on the modulation orders is: $[M_1, M_2] = [4,4]$, $[M_1, M_2] = [16,4]$, $[M_1, M_2] = [4,16]$, $[M_1, M_2] = [16,16]$. When considering $\mathcal{P}_1$, the performance gain of the J-Gray at a BER threshold of 10^{-2} varies between ~ 0.5 dB and ~ 1.6 dB, where the minimum and maximum gains are achieved by $[M_1, M_2] = [16,4]$ and $[M_1, M_2] = [4,16]$, respectively. On the other hand, when considering $\mathcal{P}_2$, the performance gain of the J-Gray at a BER threshold of 10^{-2} grows massively. The ordered gain from maximum to minimum is $\sim 12.8, 9.2, 4.3, 2.0$ dB for $[M_1, M_2] = [4,4]$, $[M_1, M_2] = [4,16]$, $[M_1, M_2] = [16,4]$ and $[M_1, M_2] = [16,16]$, respectively.

To illustrate the effect of the Rician factor on the BER performance and the J-Gray performance gain, Figure 5.9 presents the average BER of U_1 vs. SNR for the two-user case considering two PAs, $\mathcal{P}_1$ and $\mathcal{P}_2$, where $\mathcal{P}_2$ minimizes the average BER of U_1 while satisfying $P_2 \leq 10^{-2}$ at SNR $= 35$, $K_n = \{0, 5, 10, 20\}$ dB and $[M_1, M_2] = [4,4]$. A general rule of thumb is that the average BER performance at moderate and high SNRs improves as the Rician factor increases. For instance, the worst-case performance is when $K_n = 0$ dB. On the contrary, the best-case performance is when $K_n = 20$ dB which can be approximately seen as the special case of the AWGN channel. When considering $\mathcal{P}_1$, the performance gain of the J-Gray at a BER threshold of 10^{-2} drops from 1.3 dB to 0.6 dB as K_n increases from 0 dB to 20 dB. On the other hand, when considering $\mathcal{P}_2$, the performance gain of the J-Gray at a BER threshold of 10^{-2} grows significantly for most of the cases. The ordered gain from maximum to minimum is

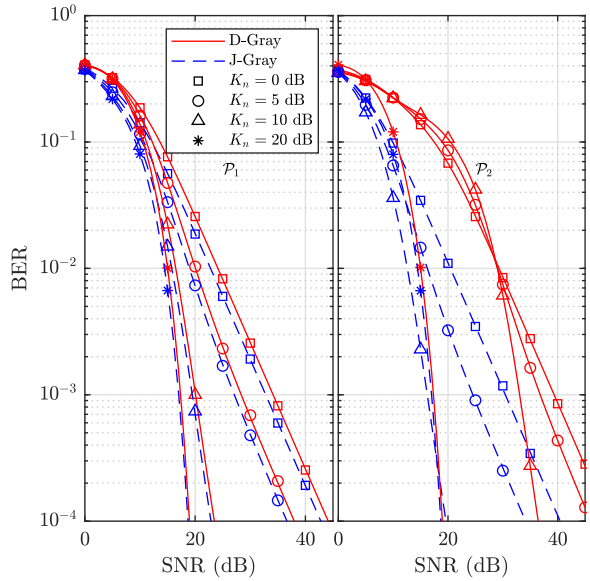


Figure 5.9: Average BER vs. SNR for U_1 considering $\mathcal{P}_1$, $\mathcal{P}_2$ and $K_n = \{0,5,10,20\}$ dB, where $N = 2$ and $[M_1, M_2] = [4,4]$.

$\sim$ 16.4, 12.8, 8.9, 0.6 dB for the Rician factor of 10, 5, 0 and 20 dB, respectively.

Figures 5.10–5.12 present the average BER vs. α_1 for the three-user case considering SNR $= 40$ dB and the following modulation orders: $[M_1, M_2, M_3] = [4,4,4]$, $[2,4,16]$, $[16,4,2]$. To ensure reliable SIC detection for the three-user case, two conditions should be satisfied which are $0 < \alpha_1 < \alpha_{1,\max}$ and $\rho(\alpha_1) < \alpha_2 < \varepsilon(\alpha_1)$ [24,27], where

$$\alpha_{1,\max} = \frac{\kappa_1}{\sqrt{\Lambda_2}\Lambda_1\kappa_3\left(\sqrt{\Lambda_2}+2\right)+\Lambda_1\left(\kappa_2+\kappa_3\right)+\kappa_1} \tag{5.49}$$

$$\varepsilon(\alpha_1) = \frac{\kappa_2}{\kappa_1\varpi_1^2} \times \left(\overline{\varpi}_1\varpi_2 + \varpi_1\varpi_3 - 2\kappa_3\sqrt{\Lambda_1\Lambda_2\alpha_1\left(\varpi_1\varpi_3 - \kappa_2\varpi_2\right)}\right) \tag{5.50}$$

$\rho(\alpha_1) = \frac{\kappa_2}{\kappa_1}\Lambda_1\alpha_1$, $\varpi_1 = \Lambda_2\kappa_3 + \kappa_2$, $\bar{\varpi}_1 = \Lambda_2\kappa_3 - \kappa_2$, $\varpi_2 = \Lambda_1\alpha_1\kappa_3$ and $\varpi_3 = (1-\alpha_1)\kappa_1$. Four different settings of α_2 are considered in the figures such that $\mathcal{A}_a = \rho(\alpha_1) + \omega$, $\mathcal{A}_b = \rho(\alpha_1) + 2\omega$, $\mathcal{A}_c = \rho(\alpha_1) + 3\omega$ and $\mathcal{A}_d = \rho(\alpha_1) + 4\omega$, where $\omega = \frac{\varepsilon(\alpha_1)-\rho(\alpha_1)}{5}$. Similar findings to the two-layer case are observed here as well. For instance, U_1 performance with D-Gray has a convex shape, whereas the J-Gray results in a monotonically decreasing behavior with α_1. Consequently, resulting in a J-Gray performance gain for U_1, which increases as α_1 increases. Another interesting finding is that U_1 performance with J-Gray is almost independent of α_2 value which is not the case with D-Gray. Moreover, while fixing α_1, U_2 performance with the D-Gray is not monotonically decreasing with α_2, unlike the J-Gray which is monotonically decreasing with

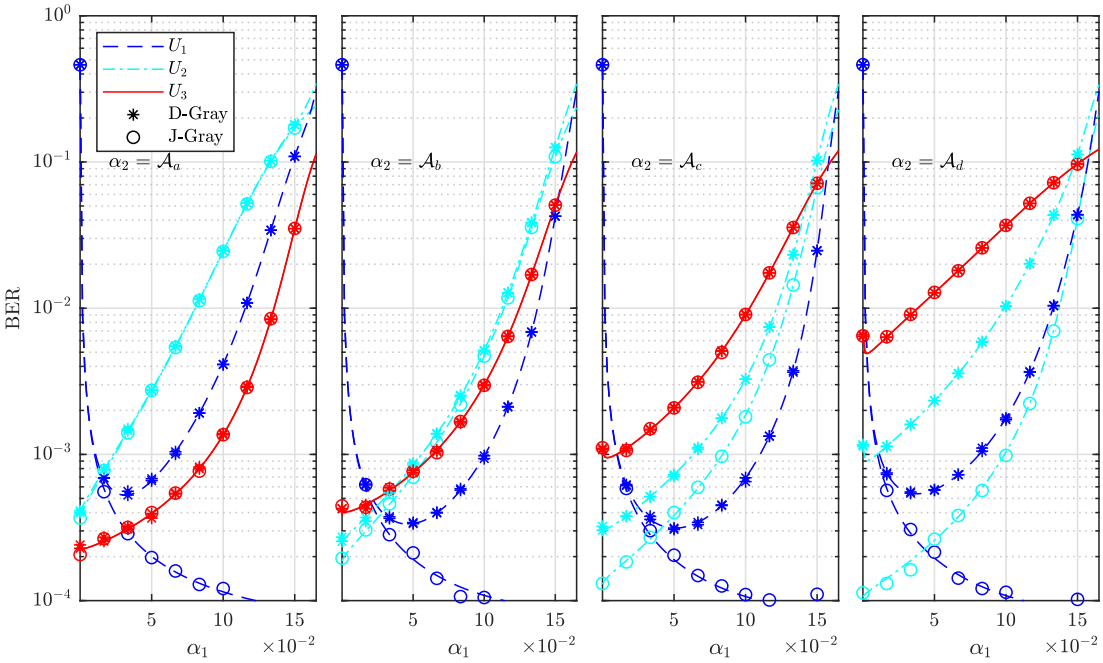


Figure 5.10: Average BER vs. α_1 considering various α_3 values, where $N = 3$, $M_n = 4 \forall n$ and SNR $= 40$ dB.

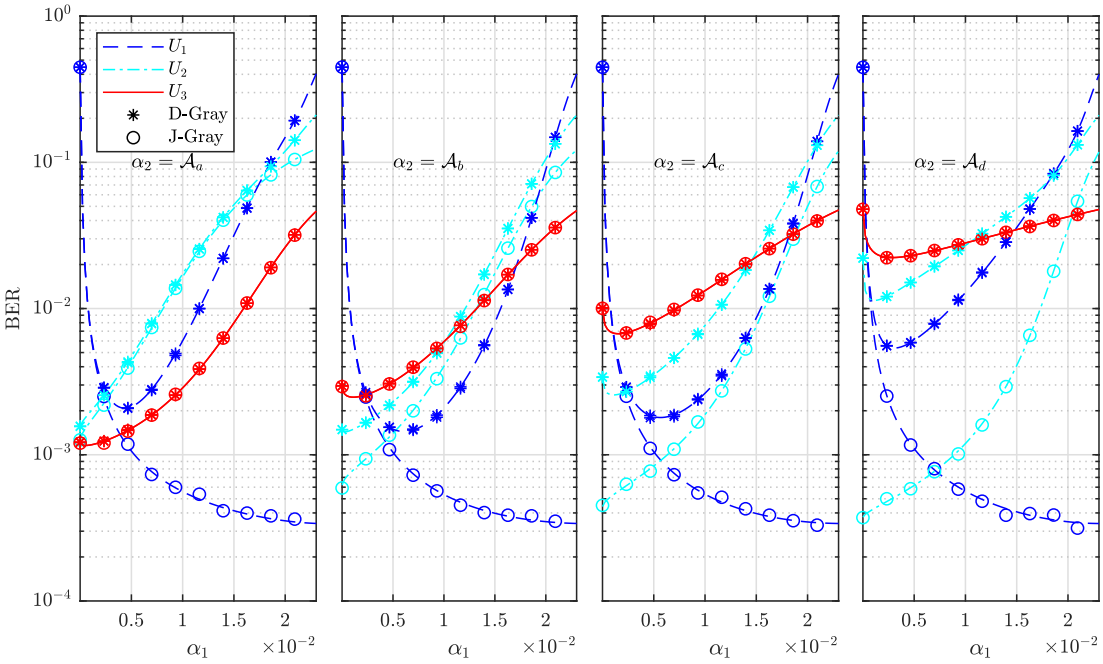


Figure 5.11: Average BER vs. α_1 considering various α_3 values, where $N = 3$, $[M_1, M_2, M_3] = [2, 4, 16]$ and SNR $= 40$ dB.

α_2. This lends the J-Gray performance gain for U_2 which increases significantly as α_2 increases. Additionally, the average BER performance of U_3 is independent of the mapping scheme which is justified as the mapping scheme does not change the bit-to-symbol mapping with respect to U_3. Furthermore, it is noted that U_1 performance depends on all users' modulation orders, where the U_1 performance from the

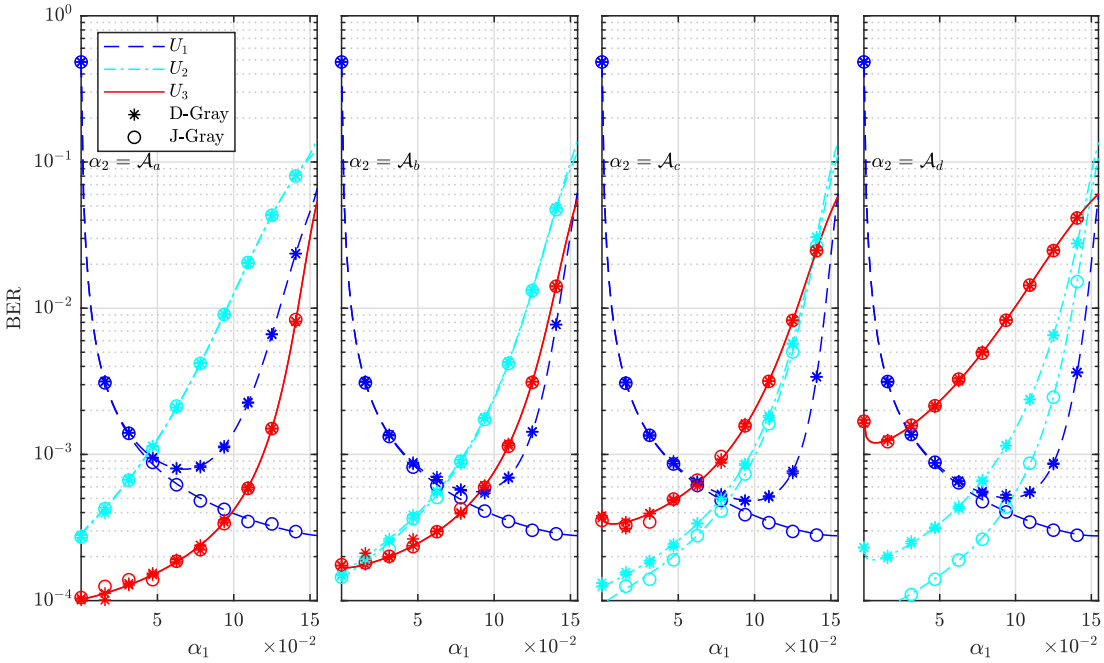


Figure 5.12: Average BER vs. α_1 considering various α_3 values, where $N = 3$, $[M_1, M_2, M_3] = [16, 4, 2]$ and SNR $= 40$ dB.

best to the worst based on the modulation orders is as follows: $[M_1, M_2, M_3] = [4, 4, 4]$, $[M_1, M_2, M_3] = [16, 4, 2]$ then $[M_1, M_2, M_3] = [2, 4, 16]$. In addition, U_2 performance is independent of M_1 and it depends on M_2 and M_3, where the performance degrades whenever either M_2 or M_3 increases. Finally, U_3 performance depends only on its modulation order, where its performance degrades as its modulation order increases.

Furthermore, Figure 5.13 presents the average BER of U_1 vs. SNR for the three-user case considering the following modulation orders: $[M_1, M_2, M_3] = [4, 4, 4]$, $[2, 4, 16]$, $[16, 4, 2]$. Two PAs, $\mathcal{P}_1$ and $\mathcal{P}_2$, are considered to highlight the performance gain of the J-Gray. $\mathcal{P}_1$ ensures equally spaced constellation points, whereas $\mathcal{P}_2$ minimizes the average BER of U_1 while satisfying $P_n \leq 10^{-2} \forall n \in \{2, 3\}$ at SNR $= 40$ dB. The results agree with the previous finding that the performance of U_1 is impacted by all users' modulation orders especially the modulation order of U_3 as the latter is given the highest power coefficient. This can be seen clearly at PA of $\mathcal{P}_1$ as the rank of P_1 performance from the best to the worst based on the modulation orders is: $[M_1, M_2, M_3] = [4, 4, 4]$, $[M_1, M_2, M_3] = [16, 4, 2]$, $[M_1, M_2, M_3] = [2, 4, 16]$. Additionally, the performance gain of the J-Gray at a BER threshold of 10^{-2} varies between ~ 0.6 dB and ~ 1.7 dB. On the other hand, when considering $\mathcal{P}_2$, the performance gain of the J-Gray at a BER threshold of 10^{-2} grows significantly, where it varies between ~ 4.9 dB and ~ 1.8 dB.

To show the impact of increasing the number of users on the BER performance and the J-Gray performance gain, Figure 5.14 presents the average BER of all users

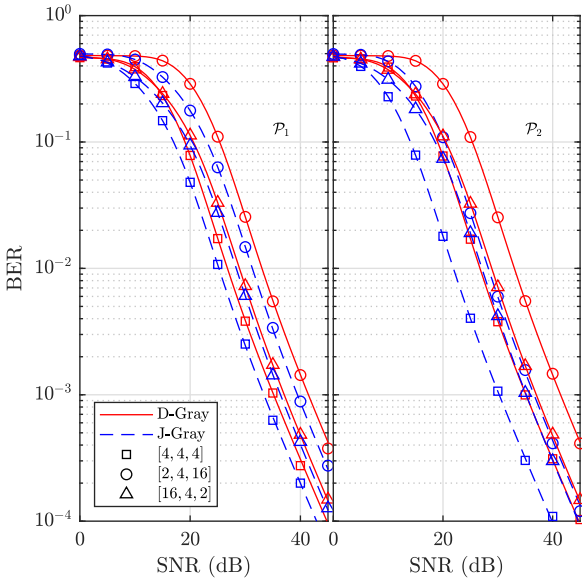


Figure 5.13: Average BER vs. SNR for U_1, where $N = 3$ and $[M_1, M_2, M_3] = [4,4,4]$, $[2,4,16]$, $[16,4,2]$.

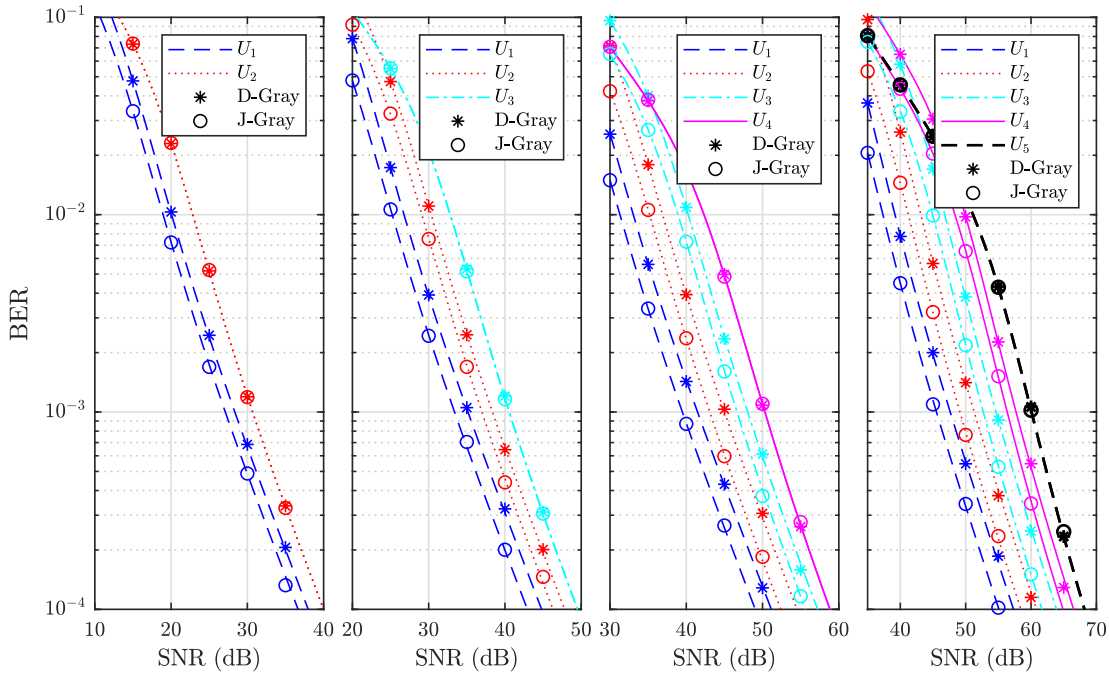


Figure 5.14: Average BER vs. SNR considering $N = \{2,3,4,5\}$ and $\mathcal{P}_1$, where $M_n = 4, \forall n$.

vs. SNR for $N \in \{2,3,4,5\}$ cases considering $\mathcal{P}_1$, where $M_n = 4 \forall n$. A general rule of thumb is that the average BER performance degrades as the number of NOMA users increases. For instance, a BER threshold of 10^{-2} is achieved by U_N at ~ 22.9, 32.9, 42.3, 51.1 dB for $N = 2, 3, 4, 5$. Hence, increasing the number of users by one leads to a 10 dB degradation for U_N. Moreover, it is noted that the J-Gray performance gain increases slightly as the number of users increases. For instance, at a BER threshold of 10^{-2}, the

U_1 performance gain increases from 1.1 dB for $N = 2$ to 1.9 dB for $N = 5$. In addition, such performance gain can be shared with the highest power user by allowing power coefficient tuning.

5.6 Conclusions

This chapter presented a generalized Gray-mapper/demapper for NOMA with arbitrary modulation orders and number of users. It derived closed-form instantaneous and average BER expressions over Rician fading channels. The expressions are presented in tabulated form for the square QAM, where general formulas are presented for the two-user scenario, which can be efficiently evaluated using any mathematical software package. The derived analytical expressions are verified by Monte Carlo simulation results. The derived expressions are used to provide insights about the average BER performance for various conditions and system configurations, including some extreme scenarios in terms of the number of users and modulation orders. For example, it is shown that the performance gain of the joint-multiuser Gray-mapping can be as high as 10 dB in certain scenarios, whereas the feasibility region's area can have a 400% improvement when compared to the disjoint-multiuser Gray-mapping. A limitation of the joint-multiuser Gray-mapping is that the performance gain in AWGN channels is limited, while it improves in fading channels. Another riveting finding is that even though the performance degrades by increasing the number of users, the performance gain keeps increasing slightly.

Chapter 6

Cognitive NOMA with Blind Transmission Mode Identification

6.1 Chapter Introduction

This chapter presents a novel non-orthogonal multiple access (NOMA) cognitive radio (CR) system where the base station (BS) opportunistically multiplexes the secondary user (SU) with the primary user (PU) using power-domain NOMA. As the PU has the priority to transmit and SU is satisfied on best-effort basis, four different transmission modes (TMs) are produced at the BS, which are PU-orthogonal multiple access (OMA), SU-OMA, PU/SU-NOMA, and silent mode. Consequently, the considered protocol can be classified as a hybrid underlay-interweave cognitive-radio inspired NOMA (CR-NOMA). The TM adaptation should be seamless for the PU where its detector configuration remains unchanged regardless of the active TM. In contrast, the SU has to identify the active TM blindly, i.e. without side information, to select the appropriate detector. The identification process is performed using a classifier that is designed based on the maximum likelihood (ML) criterion. The performance of the proposed system is analysed in terms of throughput, packet error rate (PER), and classification error. The Binomial and Multinomial theorems are utilised to simplify and allow a tractable analysis. The derived closed-form expressions, corroborated by Monte-Carlo simulation results, show that the hybrid CR-NOMA can provide substantial throughput improvement over conventional NOMA, which is about a 100%.

6.1.1 Chapter Organisation

The remaining contents of the chapter are organised as follows. In Section 6.2, the related work and motivations and contributions are mentioned. In Section 6.3, the system and channel models are introduced. In Section 6.4, the communications protocol is discussed. In Section 6.5, the performance of the multiple access classifier is analysed, while Section 6.6 presents the analysis of the considered counting rules. Then, Section 6.7 derives the PER expressions for NOMA and OMA, while Section 6.8 shows the derivations of the system throughput. Section 6.9 analyses the system's computational complexity. Section 6.10 outlines the extension to a higher number of users. Section 6.11 presents the analytical and Monte-Carlo simulation results, and finally Section 6.12 concludes the chapter.

6.1.2 Notations

The notations used throughout the chapter are as follows. Boldface uppercase and lowercase symbols, such as $\mathbf{X}$ and $\mathbf{x}$ will denote matrices and row/column vectors, respectively. The transpose is denoted by $(\cdot)^T$ and the set of complex numbers is denoted by $\mathbb{C}$. $\Pr(\cdot)$ is the probability of an event, $f(\cdot)$ is the probability density function (PDF) of a random variable, $F(\cdot)$ is the cumulative distribution function (CDF) of a random variable, $\mathbb{E}[\cdot]$ is the statistical expectation, $|\cdot|$ is the absolute value, $\mathrm{Re}[\cdot]$ and $\mathrm{Im}[\cdot]$ denotes the real and imaginary components, $\binom{n}{k}$ denotes the binomial coefficients, and $\binom{n}{k_1,k_2,\ldots,k_N}$ denotes the multinomial coefficients. The diagonal matrix with identical elements is represented as $\mathrm{diag}(a)$, where a is the diagonal. The identity $N \times N$ matrix is denoted as $\mathbf{I}_N$. The complex Gaussian random variable with a zero mean and σ^2 variance is denoted as $\mathcal{CN}(0,\sigma^2)$. The incomplete upper Gamma function is denoted by $\Gamma(a,x) = \int_x^\infty \exp(-t)t^{a-1}\,\mathrm{d}t$.

6.2 Related Work

NOMA is a promising candidate for future radio networks [140, 141] due to its ability to improve connectivity and provide high spectral efficiency [83]. It can also support diverse quality of service (QoS) requirements when users with distinct channel conditions are paired to transmit over the same transmission resources. In power-domain NOMA, the users share the same transmission time slot or frequency band, while the multiplexing is performed by assigning each user a different power coefficient. The

power coefficients might be allocated based on the users' channel conditions or QoS requirements. Successive interference cancellation (SIC) can be used at the receiver to separate the signal of each user and eliminate the interference of other users. The NOMA spectral efficiency gain typically comes at the expense of bit error rate (BER) degradation due to the inter-user interference (IUI), which is not the case for OMA techniques [24, 48, 95].

CR is another prominent technology that can be used to combat the inefficient spectrum utilization problem [142–146]. The main principle of CR is to allow unlicensed users, called SUs or cognitive users, to opportunistically or collaboratively share the spectrum of the licensed users, called PUs. The spectrum sharing process should be constrained by the PUs QoS requirements. CR has three main types, which are: interweave, underlay and overlay CR. In interweave CR, the secondary transmitter transmits to the secondary receivers when the spectrum is not occupied by PU. This requires certain activity detection technology, such as spectrum sensing, to identify idle spectrum bands [142]. In underlay CR, the PU and SU transmit simultaneously given that the SU interference is tolerable by the PU. In overlay CR, the SU collaborates with the PU to relay its messages. Therefore, it can be noted that spectrum sharing is the common principle between NOMA and CR, and also the objectives of both technologies are aligned, which triggered several research attempts to combine NOMA and CR technologies to improve the spectrum sharing process. The main literature that considered the integration of CR and NOMA is summarised in the following subsection.

6.2.1 Cognitive Radio and NOMA

Because CR and NOMA technologies typically introduce interference, the performance degradation and QoS requirements should be considered while integrating CR and NOMA [144, 147–150]. For example, four schemes are proposed in [147], which are the interweave cognitive NOMA, underlay cognitive NOMA, overlay cognitive NOMA and CR-NOMA. The first three are similar to the conventional CR, except that NOMA is employed in the secondary network. However, CR-NOMA is based on a single transmitter that prioritizes the users and tries to guarantee the QoS requirements for single or multiple users through a unique power assignment policy that gives higher priority to the PU while serving the SU opportunistically. A cooperative framework is introduced to improve the system performance [145–150], where hardware imperfections are considered in [148, 150].

Underlay cognitive NOMA is studied in [151, 152]. Liu *et al.* [151] used stochastic geometry to derive closed-form outage expressions of large-scale randomly deployed users with fixed power coefficients, where certain power constraints at the primary transmitter are considered. It is shown that underlay cognitive NOMA outperforms underlay cognitive OMA for carefully selected power coefficients and rate requirements. Similarly, Arzykulov *et al.* [152] proposed cooperative relaying for the secondary network that employs a decode-and-forward half-duplex relay. The power coefficients are optimized based on the channel quality to achieve fairness among users in terms of outage probability. Also, imperfect channel state information (CSI) is assumed and the outage-based throughput is evaluated. Moreover, an overlay cognitive NOMA network with imperfect SIC is investigated in [153], where asymptotic expressions for outage and throughput are derived, and the optimal power allocation that maximises the system throughput is computed. Bhowmick *et al.* [154] proposed an unmanned areal vehicle (UAV) assisted cooperative interweave CR network based on NOMA signalling, where spectrum sensing is performed cooperatively among the ground users and the UAV, while a global decision is made at the fusion centre. The false alarm probability of the spectrum sensing is optimized besides the NOMA power assignment to maximise the throughput.

Ding *et al.* [155] proposed CR-NOMA to strictly satisfy the PU rate QoS requirement via a unique power assignment. Unlike fixed power NOMA which tends to pair users with the most distinct channel conditions, CR-NOMA tends to pair the best channel user with the second best channel user. The power assignment scheme is generalized to the downlink and uplink in [156]. CR-NOMA is extended to multiple-input multiple-output (MIMO) systems in [157], where a precoder is designed to distinguish between the users' channels. Two power assignment policies are set to ensure the rate QoS requirement of the degraded user instantaneously or over the long term. While CR-NOMA is studied for unicast transmissions in [155–157], it is also considered for mixed multicast and unicast transmissions in [158, 159]. Two power assignment policies are proposed in [158] to satisfy the rate QoS requirements of the multicast users and the unicast user, respectively. Cooperative multicast CR-NOMA is investigated in [159] where the secondary multicast users act as relays to improve the performance of the primary unicast user. The proposed scheduling strategies require knowledge of CSI at the transmitter side.

6.2.2 Hybrid NOMA/OMA

The principle of hybrid OMA/NOMA (H-NOMA) is related to the concept of dynamic multiple access [71, 160–167], unlike the CR-NOMA which focuses on the power assignment policy. In [160], a dynamic H-NOMA scheme for the downlink is proposed to overcome the diversity gain limitation, provide a more balanced trade-off between the users' rate QoS requirements, and avoid an outage probability of one as in fixed power NOMA. This is achieved by adapting the power coefficients based on the instantaneous channel conditions. The authors of [161] propose an H-NOMA system for downlink multiple-input single-output (MISO) channels, where the BS uses zero-forcing spatial multiplexing precoding if the channel is not quasi-degraded, while NOMA precoding is used if the channel is quasi-degraded. Assaf *et al.* [71] proposed an efficient bit-loading algorithm for a downlink multi-carrier NOMA system, where the system inherently uses H-NOMA to maximise the throughput. Moreover, energy-efficient power allocation is investigated in [162] for downlink H-NOMA systems where optimal and heuristic power allocation algorithms are proposed. The authors in [163] investigate the throughput of H-NOMA for visible light communications and show that H-NOMA achieves better throughput and fairness compared to static multiple access. In addition, [164–167] propose employing cooperative techniques with H-NOMA to improve the sum rate and reduce outage. For example, decode-and-forward buffer-aided relays for downlink is proposed in [164] where practical issues such as outdated CSI, packets' retransmission and TM signalling are addressed. Further, [165] obtains the optimal TM selection by the relay node to maximise the sum throughput. Janghel *et al.* [166] derived the outage probability of downlink H-NOMA system, which employs the near user as a relay to assist the far user, where pilots are used for the NOMA/OMA switching. Fixed and dynamic power allocation schemes are proposed for the uplink H-NOMA in [167] where the multiple access switching is based on CSI and buffer state.

As can be noted from the surveyed literature, H-NOMA increases the signalling overhead because the transmitter should inform the receivers about the active TM. However, very few articles have considered the signalling overhead and knowledge of the mode switching at the receiver [161, 164–166]. To overcome the signalling overhead problem, TM identification using blind signal classification is proposed in [168–172]. For example, the work in [168, 169] considers the ML based classifiers, where Choi *et al.* [168] proposed a phase rotation-based and a pilot-based assisted blind classification that estimates the multiple access technique, modulation orders and

power coefficients. In addition, the effect of classification errors on the capacity and signal-to-interference-plus-noise ratio (SINR) is studied. On the other hand, feature-extraction based classifiers are considered in [170–172]. For instance, [170] studied the BER performance of the classifier-detector. A classification algorithm based on the fourth-order cumulant is proposed in [171]. Zhang *et al.* [172] proposed a machine-learning algorithm to determine the modulation order of the interfering user. In our work [173, 174], we present the basic concepts of CR-NOMA with blind receiver identification. The performance is evaluated in terms of throughput. However, no theoretical analysis is considered for the throughput where the results were obtained using Monte-Carlo simulation. Moreover, the PER and classification errors are not evaluated either analytically nor via simulation.

6.2.3 Motivation and Contribution

As can be noted from the aforementioned discussion, NOMA and CR are inherently aligned because NOMA is originally designed to handle interference. In particular, the BS in downlink NOMA typically has knowledge of the channel and other system parameters. Nevertheless, efficient CR systems should refrain from using extensive signalling between the primary and secondary users. As such, the BS is not supposed to inform the SUs about the PU activities. Consequently, the cognitive operation is transferred to the SU receivers who should keep listening to the channel to extract the relevant information. To the best of the authors' knowledge, and based on the surveyed literature, there is no work in the literature that considers the performance analysis of blind TM identification and imperfect SIC for the downlink hybrid underlay-interweave CR-NOMA network with BER constraints. Since the aim of the network is to increase the throughput while satisfying the users' BER QoS requirements, our objective is to derive the closed-form expressions of the users' throughput. Furthermore, we analyse the probability of false alarm and misdetection of the blind cognitive receiver, which is considered to reduce signalling overhead that is associated with the side information sent from the transmitter. Besides, the closed-form expressions of the PER of NOMA and OMA are derived. The obtained results show that the proposed scheme can significantly increase the network throughput, even when considering imperfect classification and identification at the SU. More specifically, the contribution of this chapter can be summarised as follows:

1. Analyse the performance of the uncoordinated downlink CR-NOMA network

with hybrid underlay-interweave CR and users' BER constraints proposed in [173, 174]. For brevity, we refer to the network as hybrid CR-NOMA.

2. Derive average and instantaneous closed-form expressions of the local false alarm and misdetection probabilities for the proposed multiple access classifier at the blind cognitive receiver.

3. Derive average and instantaneous closed-form expressions of the global false alarm and misdetection probabilities for various counting rules to fuse the symbol-based local observations and find the global decision on a packet.

4. Derive closed-form expressions of the NOMA and OMA average and instantaneous PERs.

5. Derive closed-form expressions of the system's throughput while considering imperfect detection at the PU and imperfect classification and detection at the SU.

6. Use Monte-Carlo simulation to corroborate the derived analytical expressions.

6.3 System and Channel Models

6.3.1 System Model

This work considers the downlink transmission of a packet-based point-to-multi-point hybrid CR-NOMA network. Each packet consists of L symbols and the packet duration is assumed to be less than the channel coherence time, hence, the transmission takes place in block fading channels. The packet structure follows the conventional structure where each packet has a header that contains fields for synchronization bits, packet number, and source and destination addresses. The payload field is dedicated to the user information symbols, while the trailer is used for error flow control cyclic redundancy check (CRC). The main system elements are shown in Figure 6.1a, which include the BS, paired primary and secondary users, where the users are denoted by U_j, $j \in \{P, S\}$, i.e. primary and secondary. In this work, we consider the case where the pairing is performed such that the PU is the far user at distance Υ_P from the BS while the SU is the near user at distance Υ_S from the BS. Such configuration grants the SU a high probability to access the licensed spectrum of the PU [159]. However, the system can also support other scenarios. For example, if the SU is the far while the

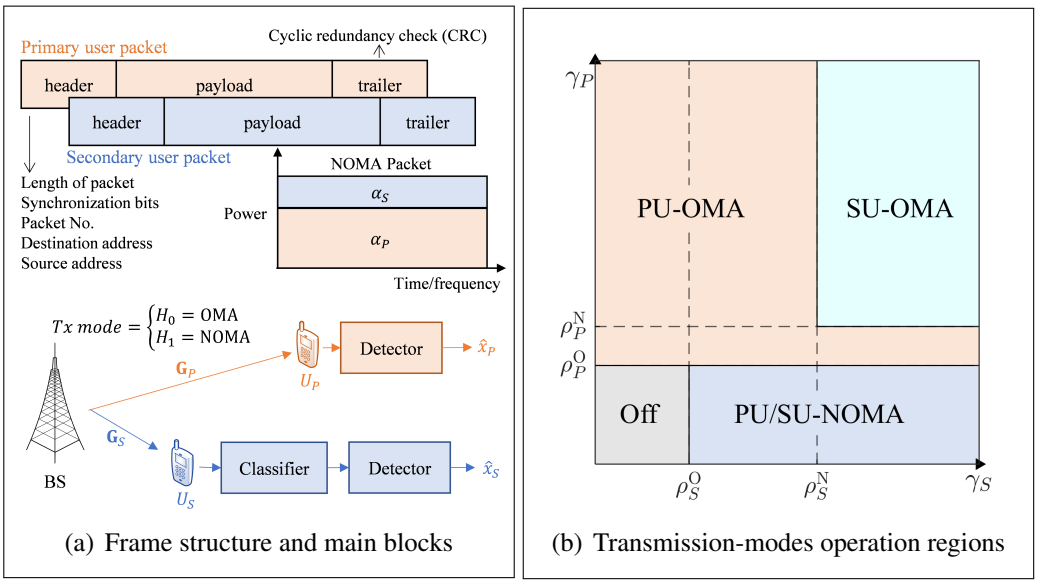


Figure 6.1: Illustrative diagram of the downlink hybrid CR-NOMA system model.

Table 6.1: Summary of Transmission-Modes.

	Packet	BER Constraints
TM_1	PU-OMA	$\{\{P_{B_P}^{\mathrm{O}} \leq \tau_P\} \cap \{P_{B_P}^{\mathrm{N}} > \tau_P\}\} \cup \{\{P_{B_P}^{\mathrm{N}} \leq \tau_P\} \cap \{P_{B_S}^{\mathrm{N}} > \tau_S\}\}$
TM_2	PU/SU-NOMA	$\{P_{B_P}^{\mathrm{N}} \leq \tau_P\} \cap \{P_{B_S}^{\mathrm{N}} \leq \tau_S\}$
TM_3	SU-OMA	$\{P_{B_P}^{\mathrm{O}} > \tau_P\} \cap \{P_{B_S}^{\mathrm{O}} \leq \tau_S\}$
TM_4	Off	$\{P_{B_P}^{\mathrm{O}} > \tau_P\} \cap \{P_{B_S}^{\mathrm{O}} > \tau_S\}$

PU is near/far, then the SU access to the licensed spectrum will be limited due to its poor channel conditions. If both users have similar and short distances from the BS, i.e. both are near, then the SU will have a high probability to access the PU spectrum because of the good channel conditions of both users. Having said that, the users in practical scenarios are uniformly distributed in a cell, which indicates that it is very likely to find users that satisfy the considered pairing condition [175]. The effect of changing the distance is shown in Section 6.11.

The PU is given the highest priority because it is the licensed user, whereas the SU is assumed to be delay-tolerant and opportunistic. The hybrid CR-NOMA aims at improving the spectral efficiency while satisfying the BER constraints of both users. Therefore, the BS adapts the TM on a packet-by-packet basis while considering the users' BER thresholds, τ_j, and achieved BERs, $P_{B_j}^{m}$, where $m \in (\mathrm{N}, \mathrm{O})$, i.e. NOMA and OMA. It is worth noting that the achieved BER throughout the packet is equivalent to the BER of the lth symbol in the packet due to the block fading channel, where $l \in \{1, \ldots, L\}$. Table 6.1 summarises all possible TMs that can be adopted by the BS. The table can be described as follows:

- TM_1: Only PU packet is sent using OMA when OMA strictly satisfies PU or if NOMA satisfies PU but does not satisfy SU.
- TM_2: PU and SU packets are multiplexed and transmitted using NOMA if the BER requirements of of both users can be satisfied.
- TM_3: Only SU packet is sent using OMA when OMA can only satisfy SU.
- TM_4: No packets are sent if neither user is satisfied by OMA.

The selection of a particular TM depends on the BER constraints, where the achieved BER by each user is affected by the power coefficients, modulation orders, and channel quality. In this work, we assume that channel state information at the transmitter (CSIT) is available for all users, which can be achieved through a delay-free feedback channel, or using the channel reciprocity concept in time-division duplexing (TDD). However, due to imperfections and noise, CSIT can be imperfect. Furthermore, we assume fixed modulation orders and power coefficients because adopting a fully flexible configuration with adaptive power and modulation orders while considering blind TM identification requires a highly complex transmitter and receiver designs [168–170, 173, 174]. Also, adapting the modulation order might result in limiting the access of the SUs to the spectrum. Hence, the number of served users will be limited. In slowly varying channels, the power coefficients for the NOMA scenario can be estimated in the initialization phase, whereas the power coefficients have to be optimized in fast fading channels, which is beyond the scope of the paper. Moreover, the modulation orders are fixed to binary phase-shift keying (BPSK) or quadrature phase-shift keying (QPSK). Figure 6.1b shows the TMs operation regions on the γ_S-γ_P plane, where colour code and pattern are used to distinguish the different TMs. The signal to noise ratio (SNR) thresholds, ρ_j^m, shown in Figure 6.1b are computed numerically using the instantaneous BER expressions derived in [95], where the ρ_j^m corresponds to the BER constraints in Table 6.1.

6.3.2 Received Signal Model

The received packet-based signal can be written as

$$\mathbf{y}_j = \begin{cases} \mathbf{G}_j\mathbf{x}_P + \mathbf{w}_j, & \text{TM}_1 \\ \mathbf{G}_j(\sqrt{\alpha_P}\mathbf{x}_P + \sqrt{\alpha_S}\mathbf{x}_S) + \mathbf{w}_j, & \text{TM}_2 \\ \mathbf{G}_j\mathbf{x}_S + \mathbf{w}_j, & \text{TM}_3 \\ \mathbf{w}_j, & \text{TM}_4 \end{cases} \tag{6.1}$$

where $\mathbf{y}_j \in \mathbb{C}^{L\times 1}$, and the power coefficients are selected such that $\alpha_P + \alpha_S = 1$. The PU data packet is $\mathbf{x}_P = [x_P^{(1)}, x_P^{(2)}, \ldots, x_P^{(L)}]^T$ and the SU packet $\mathbf{x}_S = [x_S^{(1)}, x_S^{(2)}, \ldots, x_S^{(L)}]^T$, where $\mathbb{E}[|x_j^{(l)}|^2] = 1$. The additive white Gaussian noise (AWGN) vector $\mathbf{w}_j \sim \mathcal{CN}(0, N_0\mathbf{I}_L)$ where $\sigma^2_{w_j} = N_0/2$. The complex channel gain for a single-input-single-output (SISO) channel is given $\mathbf{G}_j \in \mathbb{C}^{L\times L}$, which corresponds to block fading over one packet interval, that is $\mathbf{G}_j = g_j\mathbf{I}_L$, $g_j = \check{g}_j \times \Upsilon_j^{-\lambda/2}$, $|\check{g}_j|$ follows the Nakagami distribution with $\mathcal{M}_j$ fading parameter and normalized average power of the signal envelope, while the phase of $\check{g}_j$ is uniformly distributed between 0 and 2π. Υ_j is the distance between the BS and the jth user and λ is the pathloss exponent. The small-scale channel gains are assumed to be mutually independent and identically distributed (i.i.d). The instantaneous SNR is $\gamma_j \triangleq \frac{2|g_j|^2}{N_0}$, while the average SNR is $\bar{\gamma}_j \triangleq \frac{2\mathbb{E}[|g_j|^2]}{N_0}$. It is worth noting that the power allocation for NOMA transmission is performed such that $\alpha_P > \alpha_S$ to ensure that U_P is given the highest priority and its QoS requirements are guaranteed [175]. Hence, the NOMA decoding is performed based on the descending order of the power coefficient. Therefore, the TM adaptation will be seamless to the PU receiver as in cognitive systems. We assume in this work that the fading is slow, which implies that the power coefficients can be fixed for a packet interval. Considering multiple antennas can improve the system performance, where the system model can be directly extended to such scenarios. For example, when considering multiple antennas at the users' sides, the equivalent γ_j should be evaluated where its PDF and CDF would depend on the combining method at the receiver [24]. The same approach can be followed when considering multiple antennas at the BS or at all users' sides, as in the case where space-time block codes [129]. Alternatively, NOMA beamforming can be applied such that each pair gets one beam from the BS, where inter-pair interference needs to be accounted for in the case of non-orthogonal beams [176].

6.3.3 PU Receiver

Since PU is given the maximum power coefficient, its detector is a conventional maximum likelihood detector (MLD) regardless of the active TM because the interference from the SU signal is treated as unknown additive noise. Therefore, assuming perfect channel state information at the receiver (CSIR), a symbol-based MLD can be applied to detect the packet. Thus

$$\widehat{x}_P^{(l)} = \arg \min_{x_P^{(l)} \in \chi_P} |y_P^{(l)} - g_P \sqrt{\alpha_P} x_P^{(l)}|^2 \tag{6.2}$$

where $x_P^{(l)}$ is the trail symbol taken from χ_P codebook of PU. For notational simplicity, the symbol index l will be dropped unless it is necessary to write it explicitly. The PU receiver will identify its idle cases, i.e TM_3 or TM_4, by receiving control signals from the BS which will evaluate the feasibility of TM_1 by testing $P_{B_P}^{\mathrm{O}} > \tau_P$, where it should not be assigned packets if TM_1 is infeasible. It is worth mentioning that the CSI perfect knowledge assumption is widely used in the literature, e.g. [48, 151, 153, 156, 158, 162, 165], as there exist many efficient CSI estimation algorithms that can provide accurate CSI estimates [177]. Consequently, the system performance in the presence of CSI imperfections is generally close to those with perfect CSI as will be shown in Section 6.11.

6.3.4 SU Receiver

In the absence of side information from the BS to the SU about TM, the SU has to identify the TM blindly [168]. The TM identification process starts by testing if $P_{B_S}^{\mathrm{O}} > \tau_S$. If the test result is positive, then SU has no packets. Otherwise, the SU receiver should decide if the received signal corresponds to OMA (TM_1 or TM_3), or NOMA (TM_2). Therefore, the process can be considered as a binary hypothesis classification [169], i.e.,

$$H = \begin{cases} H_0, & \mathrm{TM}_1 \text{ or } \mathrm{TM}_3 \\ H_1, & \mathrm{TM}_2 \end{cases} \tag{6.3}$$

where H_0 is the null hypothesis while H_1 is the alternative hypothesis. If the classifier decides in favour of H_0, which corresponds to TM_1 or TM_3, the SU receiver will be configured to use a single-user MLD because both TMs correspond to an OMA signal. Then the receiver can use cross-layer design approaches to identify if the packet is its

own. More specifically, the SU may request the packet destination address from the network layer to either keep or drop the packet. On the other hand, if the classifier decides in favour of H_0, which corresponds to TM_2, the SU receiver will be configured to use SIC. Assuming perfect CSIR, the detection using MLD is similar to that in (6.2), while the SIC detection is expressed as [48],

$$\widehat{x}_S = \arg\min_{x_S \in \chi_S} \left| y_S - g_S\sqrt{\alpha_P}\widehat{x}_P - g_S\sqrt{\alpha_S}x_S \right|^2 \tag{6.4}$$

where x_S is the trial symbol selected from χ_S codebook of SU, $\widehat{x}_P$ is the PU estimated symbol using the MLD similar to (6.2). Since the PU information symbols are estimated versions of x_P, the SIC process might result in residual interference reflecting the imperfect nature of the SIC in practical scenarios.

It is worth noting that in certain adaptive NOMA systems, the decoding order using SIC might be dynamic due to varying channel gains or adaptive power allocation. In such scenarios, the receiver should identify the decoding order to recover the data correctly [178]. In the proposed system, the PU receiver design and QoS requirements should not be affected by permitting the SU to access the PU spectrum, which is typically the case in CR applications. Consequently, the decoding order of the SIC process is designed to be fixed when the PU/SU-NOMA is adopted as the largest power coefficient is assigned to the PU. Therefore, the PU receiver design becomes generally independent of the SU transmission activity.

6.3.5 Instantaneous BER Expressions

The instantaneous BER equations are listed in this subsection. The OMA BER for BPSK or QPSK is $P^{\mathrm{O}}_{B_j} = Q\left(\sqrt{\frac{\gamma_j}{\kappa_j}}\right)$, where $\kappa_j = 1$ for BPSK and 2 for QPSK. Furthermore, the BER of the PU NOMA considering identical BPSK or QPSK modulations for both users is given in [48, 95] as $P^{\mathrm{N}}_{B_P} = \sum_{r=1}^{2} v_r Q\left(c_{r,P}\sqrt{\frac{\gamma_P}{\kappa_P}}\right)$, where $\mathbf{v} = \frac{1}{2}[1,1]$, $c_{1,P} = \sqrt{\alpha_P} - \sqrt{\alpha_S}$, and $c_{2,P} = \sqrt{\alpha_P} + \sqrt{\alpha_S}$. On the other hand, the SU-NOMA BER while considering imperfect SIC is expressed in [48,95] as $P^{\mathrm{N}}_{B_S} = \sum_{r=1}^{5} v_r Q\left(c_{r,S}\sqrt{\frac{\gamma_S}{\kappa_S}}\right)$, where $\mathbf{v} = \frac{1}{2}[2,1,1,-1,-1]$, $c_{1,S} = \sqrt{\alpha_S}$, $c_{2,S} = 2\sqrt{\alpha_P} + \sqrt{\alpha_S}$, $c_{3,S} = \sqrt{\alpha_P} - \sqrt{\alpha_S}$, $c_{4,S} = \sqrt{\alpha_P} + \sqrt{\alpha_S}$ and $c_{5,S} = 2\sqrt{\alpha_P} - \sqrt{\alpha_S}$.

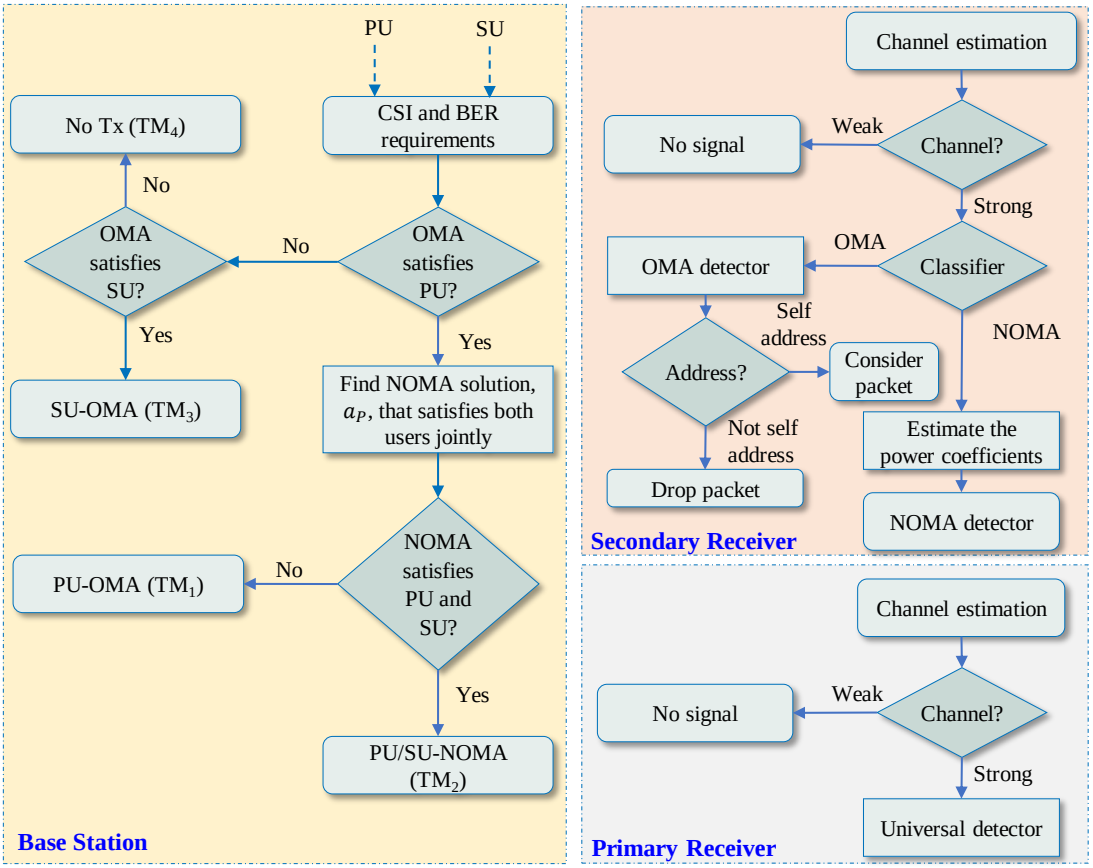


Figure 6.2: The communications protocol flowchart where adaptive power allocation and predefined modulation schemes are considered.

6.4 The Communications Protocol

The protocol for the proposed CR-NOMA network is shown in Figure 6.2. To simplify the process, it is assumed that the BS is allowed to adapt the TM and the power coefficients while using a predefined modulation order. While the adaptation process increases the complexity at the SU which has to identify the TM, the power adaptation would not change rapidly for slow fading channels. Hence, it is possible for the BS to inform the users about the selected power during the connection establishment phase. In this section, the main steps of the communications protocol will be described and discussed based on the network component, i.e., base station, PU and SU.

6.4.1 Base Station

As the BS transmission power must comply with a certain threshold, for public safety and also to minimise intercell interference, it is assumed that a fixed maximum transmission power. The algorithm starts with BS receiving CSI and the BER requirements of both users, which is a typical step in most communications networks. As mentioned previously, the BS has four TMs to choose from. At the beginning of the process, the BS computes $P^{\mathrm{O}}_{B_P}$. If $P^{\mathrm{O}}_{B_P} > \tau_P$, the PU will be muted during that packet period because it will be in an outage, and thus, the transmission would be a waste of energy. Then, the base station will compute $P^{\mathrm{O}}_{B_S}$, If $P^{\mathrm{O}}_{B_S} \leq \tau_S$, then TM_3 is selected and the SU packet will be transmitted with full power. If $P^{\mathrm{O}}_{B_S} > \tau_S$, then the SU will be muted as well. Therefore, the BS selects TM_4.

If $P^{\mathrm{O}}_{B_P} \leq \tau_P$, the BS evaluates the possibility of using NOMA to allow the SU to transmit its packet. In this case, $P^{\mathrm{N}}_{B_j}, \forall j$ should be evaluated and compared to τ_j, given that the appropriate power coefficients are selected. If the $P^{\mathrm{N}}_{B_j} > \tau_j$ for either users, then the BS reverts to TM_1. Otherwise, the BS selects TM_2. The transmitted power level in OMA cases can be optimized, to save energy and minimise interference, rather than transmitting at the maximum power level. However, it is assumed that it is fixed in this work to minimise the overhead signalling.

6.4.2 Primary User

The PU has a conventional single user detector regardless of the selected TM. This implies that the PU detector is not affected by adopting the CR-NOMA. While the PU should identify which TM the BS has selected, it should only decide if it has data or not, which can be performed with or without feedback from the BS. This is possible because the PU can simply compute $P^{\mathrm{O}}_{B_P}$ since TM_1 is selected independently of the SU conditions. If the channel conditions allow TM_1, then it should expect a packet in the next signalling period. Otherwise, it considers that no packet will be sent. Alternative methods to inform the receiver about whether it has data or not are suggested. For example, using two unique pilot sequence signatures for the two ambiguous OMA transmissions. As a result, the destination address can be deduced at the channel estimation phase. Another approach is by using rotated constellations for one of the cases. This can be easily inferred while doing the detection.

6.4.3 Secondary User

Similar to the PU, the process at the SU starts by channel estimation and $P_{B_S}^{\mathrm{O}}$ computation. If $P_{B_S}^{\mathrm{O}} > \tau_S$, then the SU considers that no packet will be sent. Otherwise, the classifier must be used to decide which TM is used and then use the corresponding detector. The classifier can be designed based on the minimum Euclidean distance over the packet, where each symbol is compared to all possible constellation points in the NOMA and OMA schemes. Then the point with the minimum Euclidean distance is selected and the TM for the packet is decided. If the classifier decision is in favour of TM_2, then SIC will be used to cancel the interference of PU, and then detect the SU packet. Otherwise, it will detect the packet directly and check the packet's destination address. If it is self-address, then it considers the packet. While if the address is not self-address, then it drops the packet. It is noteworthy to mention that the Euclidean distance-based classifier belongs to the likelihood-based classifiers which is optimal in the sense of minimizing the false alarm probability. However, a low complexity sub-optimal feature-extraction-based classifier was used in [170]. Therefore, the classifier performance-complexity trade-off must be taken into consideration when designing this stage.

As the processes of multiple access classification and interference cancellation are imperfect, applying error flow control protocols would be crucial to ensure reliable communication. For example, if one of the packets was missed, then the receiver must have the ability to know this information through the received packet's number. Also, in the case of a false alarm, the receiver can be informed about this through the packet number and the destination address. This indicates that there are several stages of protection against error caused by the imperfect classifier particularly.

6.5 Multiple Access Classifier

This section starts with an overview of the decision rules used for the multiple access classifier, which is deployed at the SU. Moreover, it defines the performance metrics for the classifier. In addition, it shows the analysis for the BPSK modulation scheme, which is considered for analytical tractability.

6.5.1 Overview of Decision Rules

The classifier tests the binary hypothesis in (6.3) and makes the decision based on the entire packet observations because the TM is adapted on a packet-by-packet basis. The optimal classifier is based on the likelihood-ratio test (LRT) [171], which uses the joint PDF of the L observation samples for the two hypotheses, denoted as $f_{\mathbf{y}_S}(\mathbf{y}_S|H;\mathbf{G}_S;\mathbf{x}_S;\mathbf{x}_P)$. By noting that the observation samples are conditionally Gaussian and mutually independent, then hypothesis likelihood $f_{\mathbf{y}_S}(\mathbf{y}_S|H;\mathbf{G}_S)$ can be expressed as [171, Eq. (4)],

$$f_{\mathbf{y}_S}(\mathbf{y}_S|H;\mathbf{G}_S) = \frac{1}{\zeta} \times \prod_{l=1}^{L} \sum_{m=1}^{M_i} \exp\left(-\left[2\sigma_w^2\right]^{-1} |y_S^{(l)} - g_S^{(l)} A_m^{(i)}|^2\right) \tag{6.5}$$

where $\zeta = (2\pi\sigma_w^2 M_i)^L$, M_i is the number of constellation points for a given multiple access technique, $y^{(l)}$ is the lth observation which is the lth received symbol in the packet and $A_m^{(i)}$ is the mth constellation point's amplitude. The mutual independence of the observation samples is due to the mutual independence of the AWGN samples and the data symbols.

The LRT is the ratio of likelihoods compared to a threshold related to the prior probabilities of the two hypotheses, i.e., $\Lambda_{\text{LRT}} = \frac{f_{\mathbf{y}_S}(\mathbf{y}_S|H_1;\mathbf{G}_S)}{f_{\mathbf{y}_S}(\mathbf{y}_S|H_0;\mathbf{G}_S)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\Pr(H_0)}{\Pr(H_1)} \triangleq \frac{\Pr(\text{TM}_3)}{\Pr(\text{TM}_2)}$. The computational complexity of the LRT is significant [171]. Therefore, an alternative suboptimal rule based on the ML estimation for each observation is proposed. The rule is suboptimal because in general $\Pr(H_0) \neq \Pr(H_1)$, whereas the maximum a posterior probability rule would give an optimal solution based on each observation. The ML rule finds the most likely multiple access by searching through all possible constellation points and selecting the point with the minimum Euclidean distance. The estimated hypothesis for the lth observation can be written as

$$\begin{aligned} \widehat{H}^{(l)} &= \arg \max_{m \in (1,M_i)} \frac{1}{2\pi\sigma_w^2} \exp\left(-\frac{|y_S^{(l)} - g_S^{(l)} A_m^{(k)}|^2}{2\sigma_w^2}\right) \\ &= \arg \min_{x_i \in (\chi_0,\chi_1)} |y_S^{(l)} - g_S^{(l)} x_i|^2 \end{aligned} \tag{6.6}$$

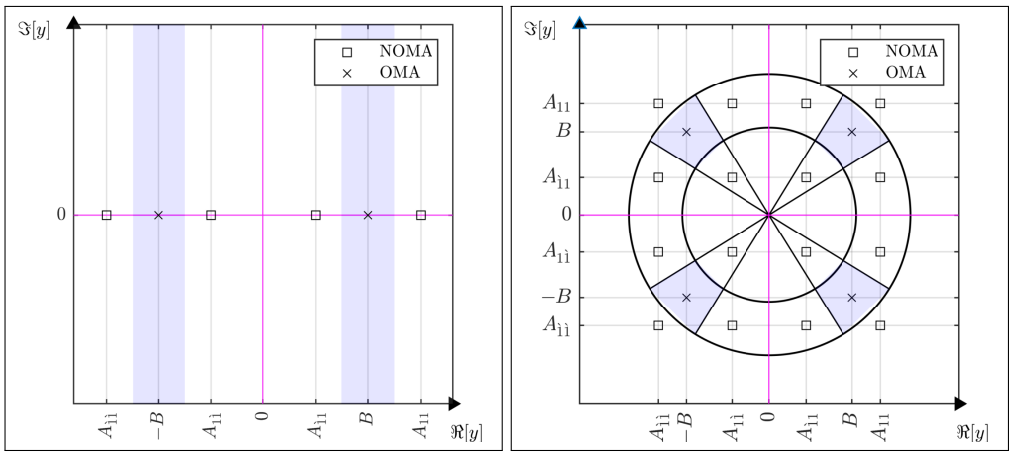
where χ_0 is a set of both users' codebooks such that $\chi_0 \in (\chi_P, \chi_S)$, while χ_1 is the NOMA codebook resulting from the SC of the two users' codebooks. The decisions on the local observations should be combined to produce the global decision. Consequently, hard-decision counting rules are proposed as discussed in the next section.

6.5.2 Performance Metrics for Classifier

The classifier is imperfect, therefore, erroneous decisions may degrade the performance, these errors are misdetection: $P_{\mathrm{m}}^{(l)} = \Pr(\widehat{H}^{(l)} = H_0 \,|\, H_1,\, \mathrm{TM}_2)$, false alarm 1 $P_{\mathrm{fa}_1}^{(l)} = \Pr(\widehat{H}^{(l)} = H_1 \,|\, H_0,\, \mathrm{TM}_1)$, and false alarm 2: $P_{\mathrm{fa}_2}^{(l)} = \Pr(\widehat{H}^{(l)} = H_1 \,|\, H_0,\, \mathrm{TM}_3)$. Note that all the previous metrics represent the decision on a single observation, which is known as the local decision and denoted with a small letter subscript. However, after combining the observations, the performance metric is denoted with a capital letter subscript. The false alarm 1 event can be ignored as the destination address field will be checked, while the misdetection and false alarm 2 (false alarm for short) events must be considered as both affect the throughput directly. The former indicates that the packet will be lost because SIC is not used, while the latter indicates that SIC will be used for decoding, which will reduce the signal strength leading to performance degradation. The misdetection is considered due to the following reasons: 1) The detector will be configured such that the average power of the PU reference symbols $\alpha_P < 1$, and thus, the probability of error for the first detection stage will increase. 2) Even if the first detection stage is successful, the SIC process will reduce the SNR of the SU by $10\log_{10}\alpha_S$ dB, which implies that the SU packet will be mostly detected incorrectly because the SNR degradation is generally more than 7 dB ($\alpha_P = 0.8$). 3) In the general scenario, the two users may have different modulation schemes/orders, hence, the first detection stage will never produce the SU signal as the detected packets will be modulated according to the PU modulation scheme.

6.5.3 Classifier Analysis

The classifier is based on the ML rule that divides the signal space into two regions representing NOMA and OMA. The regions are shown in Figure 6.3a where BPSK is used for OMA while NOMA is made of two users with BPSK. It is worth mentioning that selecting higher modulation order would make the regions polygons, resulting in intractable analysis. For instance, Figure 6.3b considers the QPSK case and the regions are trapezoids. The classification error analysis in AWGN channels can be found by considering the bivariate Gaussian PDF in polar coordinates, which includes the integration over the circles' radii and the regions' angles [179]. Further, the amplitudes of the NOMA symbols are $A_{11} = \frac{1}{\sqrt{\kappa_S}}\sqrt{\alpha_S} + \frac{1}{\sqrt{\kappa_P}}\sqrt{\alpha_P}$, $A_{\grave{1}1} = -\frac{1}{\sqrt{\kappa_S}}\sqrt{\alpha_S} + \frac{1}{\sqrt{\kappa_P}}\sqrt{\alpha_P}$, $A_{1\grave{1}} = -A_{\grave{1}1}$ and $A_{\grave{1}\grave{1}} = -A_{11}$. On the other hand, the amplitude of the OMA symbol is $B = \frac{1}{\sqrt{\kappa_j}}$. Here, conditional misdetection and false alarm probabilities will be derived.



(a) BPSK (b) QPSK

Figure 6.3: Signal space diagram for NOMA and OMA, where the shaded area is the OMA region.

For notational simplicity, the observation index $(\cdot)^{(l)}$ will be dropped unless it is necessary to include it, and thus, the probability of false alarm and miss detection will be denoted as P_{fa_2} and P_{m}, respectively.

6.5.3.1 $P_{\text{fa}_2}|\gamma_S$ analysis

The false alarm event takes place if OMA is transmitted but the classifier classifies it as a NOMA signal. Therefore, two cases should be considered.

Case 1: Amplitude B is transmitted by the BS. Therefore, the probability that the signal falls in the NOMA region represents the probability of a false alarm, which is

$$P_{\text{fa}_2}^{\text{Case 1}}|\gamma_S = \Pr\left(\tilde{w} > \frac{A_{11}-B}{2}\right) + \Pr\left(\tilde{w} < \frac{B-A_{\grave{1}1}}{-2}\right) - \Pr\left(\frac{3B+A_{11}}{-2} < \tilde{w} < \frac{3B+A_{\grave{1}1}}{-2}\right) \quad (6.7)$$

where $\tilde{w} \triangleq \text{Re}[w]$ and the superscript in P_{fa_2} represents the case number. The first term is evaluated below and all other terms can be evaluated by following the same approach. Therefore,

$$\Pr\left(\tilde{w} > \frac{A_{11}-B}{2}\gamma_S\right) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \int_{\frac{A_{11}-B}{2}\sqrt{\gamma_S}}^{\infty} \mathrm{e}^{-\tilde{w}^2/2}\,\mathrm{d}\tilde{w} = Q\left(c_{1,C}\sqrt{\gamma_S}\right) \quad (6.8)$$

where $Q(\cdot)$ is the Gaussian complementary CDF and $c_{1,C} = \frac{A_{11}-B}{2}$. Therefore,

$$P_{\text{fa}_2}^{\text{Case 1}}|\gamma_S = Q\left(c_{1,C}\sqrt{\gamma_S}\right) + Q\left(c_{2,C}\sqrt{\gamma_S}\right) - Q\left(c_{3,C}\sqrt{\gamma_S}\right) + Q\left(c_{4,C}\sqrt{\gamma_S}\right) \tag{6.9}$$

where $c_{2,C} = \frac{B-A_{\grave{1}1}}{2}$, $c_{3,C} = \frac{3B+A_{\grave{1}1}}{2}$ and $c_{4,C} = \frac{3B+A_{11}}{2}$.

Case 2: This case case is similar to *Case 1*, except that the amplitude B is replaced by -B. Consequently, $P_{\text{fa}_2}^{\text{Case 1}}|\gamma_S$ and $P_{\text{fa}_2}^{\text{Case 2}}|\gamma_S$ are equal due to symmetry. Therefore, the average of the two cases $P_{\text{fa}_2}|\gamma_S = P_{\text{fa}_2}^{\text{Case 1}}|\gamma_S = P_{\text{fa}_2}^{\text{Case 2}}|\gamma_S$.

6.5.3.2 $P_{\text{m}}|\gamma_S$ analysis

The event of misdetection takes place if NOMA is transmitted, but the classifier classifies it as an OMA signal. Therefore, four cases will result.

Case 1: Amplitude A_{11} is transmitted by the BS. Therefore, the probability that the signal falls in the OMA region represents the probability of misdetection, which is

$$P_{\text{m}}^{\text{Case 1}}|\gamma_S = \Pr\left(\frac{2A_{11}-A_{\grave{1}1}-B}{-2} < \tilde{w} < \frac{A_{11}-B}{-2}\right) + \Pr\left(\frac{3A_{11}+B}{-2} < \tilde{w} < \frac{2A_{11}+A_{\grave{1}1}+B}{-2}\right). \tag{6.10}$$

Evaluating (6.10) gives

$$P_{\text{m}}^{\text{Case 1}}|\gamma_S = Q\left(c_{1,C}\sqrt{\gamma_S}\right) - Q\left(c_{5,C}\sqrt{\gamma_S}\right) + Q\left(c_{6,C}\sqrt{\gamma_S}\right) - Q\left(c_{7,C}\sqrt{\gamma_S}\right) \tag{6.11}$$

where $c_{5,C} = \frac{2A_{11}-A_{\grave{1}1}-B}{2}$, $c_{6,C} = \frac{2A_{11}+A_{\grave{1}1}+B}{2}$ and $c_{7,C} = \frac{3A_{11}+B}{2}$.

Case 2: Amplitude $A_{\grave{1}1}$ is transmitted. Therefore, the probability of misdetection is given as

$$P_{\text{m}}^{\text{Case2}}|\gamma_S = \Pr\left(\frac{B-A_{\grave{1}1}}{2} < \tilde{w} < \frac{A_{11}+B-2A_{\grave{1}1}}{2}\right) + \Pr\left(\frac{A_{11}+2A_{\grave{1}1}+B}{-2} < \tilde{w} < \frac{3A_{\grave{1}1}+B}{-2}\right). \tag{6.12}$$

Evaluating (6.12) gives

$$P_{\text{m}}^{\text{Case 2}}|\gamma_S = Q\left(c_{2,C}\sqrt{\gamma_S}\right) - Q\left(c_{8,C}\sqrt{\gamma_S}\right) + Q\left(c_{9,C}\sqrt{\gamma_S}\right) - Q\left(c_{10,C}\sqrt{\gamma_S}\right) \tag{6.13}$$

where $c_{8,C} = \frac{A_{11}+B-2A_{\grave{1}1}}{2}$, $c_{9,C} = \frac{3A_{\grave{1}1}+B}{2}$ and $c_{10,C} = \frac{A_{11}+2A_{\grave{1}1}+B}{2}$.

Case 3 and *Case 4* analyses are identical to *Case 2* and *Case 1*, respectively. Therefore, the conditional probability of misdetection can be written as

$$P_{\mathrm{m}}|\gamma_S = \frac{1}{4}\sum_{r=1}^{4} P_{\mathrm{m}}^{\mathrm{Case}\,r}|\gamma_S = \sum_{r=1}^{10} v_r Q\left(c_{r,C}\sqrt{\gamma_S}\right) \tag{6.14}$$

where $\mathbf{v} = \frac{1}{2}[1,1,0,0,-1,1,-1,-1,1,-1]$.

6.5.3.3 Unconditional analysis

Computing P_{fa_2} and P_{m} using (6.9) and (6.14) requires integration over the PDF of γ_S. Therefore, it is necessary to evaluate the following integral,

$$I_1 = \acute{F}_{\gamma_S}(\rho)\sum_r v_r \int_{\rho}^{\infty} Q\left(c_{r,C}\sqrt{\gamma_S}\right) f_{\gamma_S}(\gamma_S)\,\mathrm{d}\gamma_S \tag{6.15}$$

where $\acute{F}_{\gamma_S}(\rho) \triangleq \frac{1}{1-F_{\gamma_S}(\rho)}$, the CDF and PDF of γ_j are respectively given by

$$F_{\gamma_j}(\gamma_j) = 1 - \Gamma\left(\mathcal{M}_j, \mathcal{M}_j\gamma_j\bar{\gamma}_j^{-1}\right)\acute{\Gamma}\left(\mathcal{M}_j\right) \tag{6.16}$$

where $\acute{\Gamma}(\cdot) = 1/\Gamma(\cdot)$ and

$$f_{\gamma_j}(\gamma_j) = \left(\mathcal{M}_j\bar{\gamma}_j^{-1}\right)^{\mathcal{M}_j}\gamma_j^{\mathcal{M}_j-1}\acute{\Gamma}(\mathcal{M}_j)\exp\left(-\mathcal{M}_j\gamma_j\bar{\gamma}_j^{-1}\right). \tag{6.17}$$

It is worth noting that the integral lower limit shown in (6.15) is ρ rather than 0 because there is a condition on γ_S that is related to the selection of TM_2 or TM_3 as shown in Figure 6.1b. In addition, the subscript is dropped from ρ to be generic for both metrics. The value of ρ depends on the SNR threshold shown in Figure 6.1b where $\rho = \rho_S^{\mathrm{N}}$ for P_{m} and ρ_S^{O} for P_{fa_2}. More details about the thresholds are given in Section 6.8. The exact closed-form solution for (6.15) for the special case of $\mathcal{M}_j = 1$ can be found in [180, Eq. (29)]. Nonetheless, the exact closed-form solution considering the general case of $\mathcal{M}_j$ is intractable since the integral's lower limit is not 0. Therefore, we use the following approximation for $Q(\cdot)$,

$$Q(z) \approx \sum_n a_n \exp\left(-b_n z^2\right) \tag{6.18}$$

Table 6.2: Summary of the decision rules.

AND Rule	**OR Rule**	**$\bar{K}$-out-of-L Rule**
$\widehat{H} = \begin{cases} H_0, & 0 \leq O < L \\ H_1, & O = L \end{cases}$	$\widehat{H} = \begin{cases} H_0, & O = 0 \\ H_1, & O \geq 1 \end{cases}$	$\widehat{H} = \begin{cases} H_0, & O < \bar{K} \\ H_1, & O \geq \bar{K} \end{cases}$

where $n \geq 3$, $a_n = \frac{1}{2n}$ and b_n values can be found in [181, Table 2] while noting that $Q(z) = \frac{1}{2}\text{erfc}\left(\frac{x}{\sqrt{2}}\right)$. Consequently, the integral in (6.15) can be approximated as

$$I_1 = \acute{F}_{\gamma_S}(\rho) \sum_r \sum_n v_r \theta_n \int_{\rho}^{\infty} \gamma_S^{\mathcal{M}_S - 1} \exp(-\xi_{r,n}\gamma_S)\, \mathrm{d}\gamma_S \tag{6.19}$$

where $\theta_n = a_n \left(\mathcal{M}_S \bar{\gamma}_S^{-1}\right)^{\mathcal{M}_S} \acute{\Gamma}(\mathcal{M}_S)$ and $\xi_{r,n} = b_n c_{r,C}^2 + \frac{\mathcal{M}_S}{\bar{\gamma}_S}$. It is noted that the integral in (6.19) has a closed-form solution [182, Eq. 3.381.3], and hence, it can be expressed as

$$I_1 = \acute{F}_{\gamma_S}(\rho) \sum_r \sum_n v_r \theta_n \xi_{r,n}^{-\mathcal{M}_S} \Gamma(\mathcal{M}_S, \rho\xi_{r,n}). \tag{6.20}$$

Therefore, $P_{\text{fa}_2} = I_1 | \{\rho = \rho_S^{\text{N}}, \mathbf{v} = [1, 1, -1, 1]\}$ and $P_{\text{m}} = I_1 | \{\rho = \rho_S^{\text{O}}, \mathbf{v} = \frac{1}{2}[1, 1, 0, 0, -1, 1, -1, -1, 1, -1]\}$.

6.6 Counting Rules Analysis

Several counting rules are studied in this section which are: AND, OR, Majority and $\bar{K}$-out-L rules. The counting rules are based on the hard decision of the local observations. This section starts by stating the decision-rule statistics of each counting rule, then it derives the conditional and unconditional misdetection and false alarm global probabilities.

6.6.1 Decision-Rule Statistics

The indicator function is defined to denote $\widehat{H}^{(l)}$ numerically, which can be expressed as

$$\mathbf{1}_{H_1}^{(l)}\left(\widehat{H}^{(l)}\right) = \begin{cases} 1, & \widehat{H}^{(l)} = H_1 \\ 0, & \widehat{H}^{(l)} \neq H_1 \end{cases}. \tag{6.21}$$

In addition, the total number of local observations decided as H_1 is $O = \sum_{l=1}^{L} \mathbf{1}_{H_1}^{(l)}$. Therefore, the decision-rule statistics can be written for each counting rule as shown in Table 6.2 and detailed as follows:

1. AND Rule: All local observations must be NOMA to have a global NOMA observation, otherwise it is a global OMA observation.

2. OR Rule: At least one local observation must be NOMA to have a global decision of NOMA. Otherwise, it is a global OMA observation.

3. $\bar{K}$-out-of-L Rule: At least $\bar{K}$ local observations must be NOMA to decide a global NOMA TM, otherwise it is a global OMA. Note that the Majority rule is a special case of the $\bar{K}$-out-of-L rule when $\bar{K} = L/2$ given that L is even.

6.6.2 Conditional Performance Analysis

In general, the conditional global probability of the false alarm is given as $P_{\mathrm{FA}_2}|\gamma_S = \Pr(\widehat{H} = H_1 \,|\, H_0,\ \mathrm{TM}_3) = 1 - \Pr(\widehat{H} = H_0 \,|\, H_0,\ \mathrm{TM}_3)$, while the conditional global probability of misdetection is given as $P_{\mathrm{M}}|\gamma_S = \Pr(\widehat{H} = H_0 \,|\, H_1) = 1 - \Pr(\widehat{H} = H_1 \,|\, H_1)$.

6.6.2.1 AND rule

The global decision regarding a packet will be considered a false alarm if all the local observations decisions are false alarms. Therefore, $P_{\mathrm{FA}_2}|\gamma_S = \prod_{l=1}^{L} P_{\mathrm{fa}_2}^{(l)}|\gamma_S = (P_{\mathrm{fa}_2}|\gamma_S)^L$. On the other hand, $P_{\mathrm{M}}|\gamma_S = 1 - \prod_{l=1}^{L}\left(1 - P_{\mathrm{m}}^{(l)}|\gamma_S\right) = 1 - (1 - P_{\mathrm{m}}|\gamma_S)^L$ which can be expanded using the Binomial theorem, $(a+b)^n = \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k$. Therefore, $P_{\mathrm{M}}|\gamma_S = 1 - \sum_{r=0}^{L}\binom{L}{r}(-1)^r\,(P_{\mathrm{m}}|\gamma_S)^r$.

6.6.2.2 OR rule

The global decision is considered a false alarm if at least one local observation decision is a false alarm. Therefore, $P_{\mathrm{FA}_2}|\gamma_S$ can be expressed as $P_{\mathrm{FA}_2}|\gamma_S = 1 - \prod_{l=1}^{L}(1 - P_{\mathrm{fa}_2}^{(l)}|\gamma_S) = 1 - (1 - P_{\mathrm{fa}_2}|\gamma_S)^L$, which can be written as $P_{\mathrm{FA}_2}|\gamma_S = 1 - \sum_{r=0}^{L}\binom{L}{r}(-1)^r(P_{\mathrm{fa}_2}|\gamma_S)^r$. On the other hand, the whole packet will be in a misdetection event if all observations are in misdetection. As a result, $P_{\mathrm{M}}|\gamma_S$ is expressed as $P_{\mathrm{M}}|\gamma_S = \prod_{l=1}^{L} P_{\mathrm{m}}^{(l)}|\gamma_S = (P_{\mathrm{m}}|\gamma_S)^L$.

Table 6.3: Summary of the approximate closed-form expressions for the global false alarm and misdetection.

	P_{FA_2}	P_{M}
AND	$\frac{1}{1-F_{\gamma_S}(\rho_S^{\mathrm{O}})}\mathcal{F}\left(L,\rho_S^{\mathrm{O}}\right)$	$1-\frac{1}{1-F_{\gamma_S}(\rho_S^{\mathrm{N}})}\sum_{r=0}^{L}\binom{L}{r}(-1)^r\mathcal{F}\left(r,\rho_S^{\mathrm{N}}\right)$
OR	$1-\frac{1}{1-F_{\gamma_S}(\rho_S^{\mathrm{O}})}\sum_{r=0}^{L}\binom{L}{r}(-1)^r\mathcal{F}\left(r,\rho_S^{\mathrm{O}}\right)$	$\frac{1}{1-F_{\gamma_S}(\rho_S^{\mathrm{N}})}\mathcal{F}\left(L,\rho_S^{\mathrm{N}}\right)$
$\bar{K}$-out-of-L Rule	$\frac{1}{1-F_{\gamma_S}(\rho_S^{\mathrm{O}})}\sum_{l=\bar{K}}^{L}\sum_{r=0}^{L-l}\binom{L}{l}\binom{L-l}{r}(-1)^r\mathcal{F}\left(l+r,\rho_S^{\mathrm{O}}\right)$	$1-\frac{1}{1-F_{\gamma_S}(\rho_S^{\mathrm{N}})}\sum_{l=\bar{K}}^{L}\sum_{r=0}^{l}\binom{L}{l}\binom{l}{r}(-1)^r\mathcal{F}\left(L+r-l,\rho_S^{\mathrm{N}}\right)$

6.6.2.3 $\bar{K}$-out-of-L rule

The global decision is considered a false alarm if at least $\bar{K}$ local observations decisions are false alarms. Therefore, $P_{\mathrm{FA}_2}|\gamma_S$ can be expressed as

$$\begin{aligned} P_{\mathrm{FA}_2}|\gamma_S &= \sum_{l=\bar{K}}^{L}\binom{L}{l}(P_{\mathrm{fa}_2}|\gamma_S)^l(1-P_{\mathrm{fa}_2}|\gamma_S)^{L-l} \\ &= \sum_{l=\bar{K}}^{L}\sum_{r=0}^{L-l}\binom{L}{l}\binom{L-l}{r}(-1)^r(P_{\mathrm{fa}_2}|\gamma_S)^{l+r}. \end{aligned} \tag{6.22}$$

Moreover, the global decision regarding a packet will be considered misdetection if at most $\bar{K}$ local observations decisions are misdetections. Therefore, $P_{\mathrm{M}}|\gamma_S$ is

$$\begin{aligned} P_{\mathrm{M}}|\gamma_S &= 1-\sum_{l=\bar{K}}^{L}\binom{L}{l}(1-P_{\mathrm{m}})^l(P_{\mathrm{m}})^{L-l} \\ &= 1-\sum_{l=\bar{K}}^{L}\sum_{r=0}^{l}\binom{L}{l}\binom{l}{r}(-1)^r(P_{\mathrm{m}})^{L+r-l}. \end{aligned} \tag{6.23}$$

6.6.3 Unconditional Performance Analysis

Removing the conditioning on γ_S from $P_{\mathrm{FA}_2}|\gamma_S$ and $P_{\mathrm{M}}|\gamma_S$ requires integration over the PDF of γ_S. Therefore, P_{FA_2} and P_{M} can be found by considering the following integral,

$$I_2 = \acute{F}_{\gamma_S}(\rho)\int_{\rho}^{\infty}\left[\sum\nolimits_r v_r Q\left(c_{r,C}\sqrt{\gamma_S}\right)\right]^{\nu} f_{\gamma_S}(\gamma_S)\,\mathrm{d}\gamma_S \tag{6.24}$$

where $\nu \geq 0$. The evaluation of the integral in (6.24) is intractable due to the exponent factor ν, which leads to producing terms that consist of the product of multiple Q-function. Therefore, an approximate solution can be developed by approximating the $Q(\cdot)$ function using (6.18). Furthermore, the local probabilities of misdetection and false alarm are approximated using the most dominant terms only to generate a

tractable solution. Therefore, the integral in (6.24) can be approximated as

$$I_2 = \acute{F}_{\gamma_S}(\rho) \int_{\rho}^{\infty} \underbrace{\left[\sum_{r=1}^{2} \sum_{n} v_r a_n \exp\left(-c_{r,C}^2 b_n \gamma_S\right)\right]^{\nu}}_{\Xi} \times f_{\gamma_S}(\gamma_S)\, \mathrm{d}\gamma_S. \tag{6.25}$$

The term Ξ can be expanded using the Multinomial theorem as [183, p. 70]

$$\Xi = \sum_{N=\nu} \binom{\nu}{N_1, \ldots, N_{2n}} \times \prod_{q=1}^{n} \prod_{t=2q-1}^{2q} \left(v_\zeta a_q\right)^{N_t} \exp\left(-b_q N_t c_{\zeta,C}^2 \gamma_S\right) \tag{6.26}$$

where $N = N_1 + N_2 + \cdots + N_{2n}$, $\zeta = 1$ if t is odd, otherwise $\zeta = 2$. By substituting (6.26) in (6.25), we obtain

$$I_2 = \acute{F}_{\gamma_S}(\rho) \sum_{N=\nu} \binom{\nu}{N_1, \ldots, N_{2n}} \times \int_{\rho}^{\infty} \prod_{q=1}^{n} \prod_{t=2q-1}^{2q} \left(\frac{\mathcal{M}_S}{\bar{\gamma}_S}\right)^{\mathcal{M}_S} \acute{\Gamma}(\mathcal{M}_S) \left(v_\zeta a_q\right)^{N_t} \times \exp\left(-\gamma_S \left[b_q N_t c_{\zeta,C}^2 + \mathcal{M}_S \bar{\gamma}_S^{-1}\right]\right) \gamma_S^{\mathcal{M}_S - 1}\, \mathrm{d}\gamma_S. \tag{6.27}$$

By simplifying the product of exponential functions to a single exponential with a sum of exponents, the expression can be further simplified to

$$I_2 = \acute{F}_{\gamma_S}(\rho) \sum_{N=\nu} \binom{\nu}{N_1, \ldots, N_{2n}} \Psi \times \int_{\rho}^{\infty} \gamma_S^{\mathcal{M}_S - 1} \exp\left(-\Delta \times \gamma_S\right) \mathrm{d}\gamma_S \tag{6.28}$$

where $\Psi = \prod_{q=1}^{n} \prod_{t=2q-1}^{2q} \frac{\left(\frac{\mathcal{M}_S}{\bar{\gamma}_S}\right)^{\mathcal{M}_S}}{\Gamma(\mathcal{M}_S)} \left(v_\zeta a_q\right)^{N_t}$ and $\Delta = \sum_{q=1}^{n} \sum_{t=2q-1}^{2q} b_q N_t c_{\zeta,C}^2 + \frac{\mathcal{M}_S}{\bar{\gamma}_S}$. It is noted that the integral in (6.28) has a closed-form solution [182, Eq. 3.381.3], and hence, it can be expressed as

$$I_2 = \acute{F}_{\gamma_S}(\rho) \sum_{N=\nu} \binom{\nu}{N_1, \ldots, N_{2n}} \Psi \times \Delta^{-\mathcal{M}_S} \Gamma(\mathcal{M}_S, \Delta \times \rho) \triangleq \acute{F}_{\gamma_S}(\rho)\, \mathcal{F}(\nu, \rho). \tag{6.29}$$

Table 6.3 summarises the unconditional global probability of the false alarm for the considered counting rules using the expression derived in (6.29).

6.7 Packet Error Rate Analysis

In this section, the PER is derived to quantify the detector performance for each user. The conditional PER in a block fading channel can be found as,

$$\begin{aligned} P_{e_j}^{m}|\gamma_j &= 1-\left(1-P_{B_j}^{m}\right)^{L} \\ &= 1-\sum\nolimits_{r=0}^{L}\binom{L}{r}(-1)^{r}\left(P_{B_j}^{m}\right)^{r}. \end{aligned} \tag{6.30}$$

Removing the conditioning on γ_j from the conditional PER requires integration over the PDF of γ_S. Nonetheless, the integration with (6.30) in its current is intractable. Thus, the approximation of $Q(\cdot)$ in (6.18) should be used.

6.7.1 OMA PER

The PER of PU-OMA TM consists of two cases because there are two conditions that each of which may lead to PU-OMA transmission. Therefore, the PER can be found as $P_{e_P}^{\mathrm{O}} = \Pr(\text{Case}\,1)\times P_{e_P}^{\mathrm{O},\text{Case}\,1} + \Pr(\text{Case}\,2)\times P_{e_P}^{\mathrm{O},\text{Case}\,2}$. Case 1 corresponds to $\rho_P^{\mathrm{O}} \le \gamma_P < \rho_P^{\mathrm{N}}$, whereas Case 2 corresponds to $\gamma_P \ge \rho_P^{\mathrm{N}} \cap \gamma_S < \rho_S^{\mathrm{N}}$. Hence, $\Pr(\text{Case}\,1) = F_{\gamma_P}\left(\rho_P^{\mathrm{N}}\right) - F_{\gamma_P}\left(\rho_P^{\mathrm{O}}\right)$, while $\Pr(\text{Case}\,2) = \left(1-F_{\gamma_P}\left(\rho_P^{\mathrm{N}}\right)\right)F_{\gamma_S}\left(\rho_S^{\mathrm{N}}\right)$. Note that the probabilities of these two events should be normalized by the space probability, which is $\Pr(\text{Case}\,1) + \Pr(\text{Case}\,2)$.

To compute the PER, the approximation in (6.18) should be used. By noting the change of γ subscript, it can be seen that $P_{e_P}^{\mathrm{O},\text{Case}\,2}$ follows (6.24), where (6.29) can be used as its solution with the Multinomial coefficient being $\binom{\nu}{N_1,\ldots,N_n}$, $\nu = L$, $\rho = \rho_P^{\mathrm{N}}$, $\Psi = \prod_{q=1}^{n}\frac{\left(\frac{\mathcal{M}_P}{\bar{\gamma}_P}\right)^{\mathcal{M}_P}}{\Gamma(\mathcal{M}_P)}\left(a_q\right)^{N_t}$ and $\Delta = \sum_{q=1}^{n} b_q N_t + \frac{\mathcal{M}_P}{\bar{\gamma}_P}$. Hence, $P_{e_P}^{\mathrm{O},\text{Case}\,2} = 1 - \frac{1}{1-F_{\gamma_P}\left(\rho_P^{\mathrm{N}}\right)}\sum_{r=0}^{L}\binom{L}{r}(-1)^{r}\mathcal{F}\left(r,\rho_P^{\mathrm{N}}\right)$. Furthermore, $P_{e_P}^{\mathrm{O},\text{Case}\,1}$ has the lower and upper integration limits as ρ_P^{O} and ρ_P^{N}, respectively. Therefore, the integral in (6.24) should be adjusted to fit this case. Consequently,

$$P_{e_P}^{\mathrm{O},\text{Case}\,1} = 1 - \frac{1}{F_{\gamma_P}\left(\rho_P^{\mathrm{N}}\right) - F_{\gamma_P}\left(\rho_P^{\mathrm{O}}\right)} \times \sum\nolimits_{r=0}^{L}\binom{L}{r}(-1)^{r}\left[\mathcal{F}\left(r,\rho_P^{\mathrm{O}}\right) - \mathcal{F}\left(r,\rho_P^{\mathrm{N}}\right)\right]. \tag{6.31}$$

Similarly, $P_{e_S}^{\mathrm{O}}$ follows (6.24) and (6.29) can be used as its solution with the Multinomial

coefficient being $\binom{\nu}{N_1,\ldots,N_n}$, $\nu = L$, $\rho = \rho_S^{\mathrm{O}}$,

$$\Psi = \prod_{q=1}^{n} \left(\mathcal{M}_S \bar{\gamma}_S^{-1}\right)^{\mathcal{M}_S} \acute{\Gamma}\left(\mathcal{M}_S\right) \left(a_q\right)^{N_t} \tag{6.32}$$

and

$$\Delta = \sum_{q=1}^{n} b_q N_t + \mathcal{M}_S \bar{\gamma}_S^{-1}. \tag{6.33}$$

Therefore,

$$P_{e_S}^{\mathrm{O}} = 1 - \frac{1}{1 - F_{\gamma_S}\left(\rho_S^{\mathrm{O}}\right)} \sum_{r=0}^{L} \binom{L}{r} (-1)^r \times \mathcal{F}\left(r, \rho_S^{\mathrm{O}}\right). \tag{6.34}$$

6.7.2 NOMA PER

The NOMA PER analysis would be intractable if the instantaneous BER shown in Section 6.3.5 are used in the form of a sum of $Q(\cdot)$ terms. Therefore, it is proposed to approximate these expressions considering the $Q(\cdot)$ approximation in (6.18). Further, the two most dominant $Q(\cdot)$ terms are considered for the SU. Hence, the instantaneous BER for the PU and SU can be written as $P_{B_P}^{\mathrm{N}} \approx \sum_{r=1}^{2} \sum_n v_r a_n \exp\left(-c_{r,P}^2 b_n \gamma_P\right)$ and $P_{B_S}^{\mathrm{N}} \approx \sum_{r\in\{1,3\}} \sum_n v_r a_n \exp\left(-c_{r,S}^2 b_n \gamma_S\right)$. Therefore, $P_{e_P}^{\mathrm{N}}$ follows (6.24) and (6.29) can be used as its solution with $\nu = L$, $\rho = \rho_P^{\mathrm{N}}$, $\Psi = \prod_{q=1}^{n} \prod_{t=2q-1}^{2q} \left(\mathcal{M}_P \bar{\gamma}_P^{-1}\right)^{\mathcal{M}_P} \acute{\Gamma}\left(\mathcal{M}_P\right) \left(v_\zeta a_q\right)^{N_t}$ and $\Delta = \sum_{q=1}^{n} \sum_{t=2q-1}^{2q} b_q N_t c_{\zeta,P}^2 + \mathcal{M}_P \bar{\gamma}_P^{-1}$, where $\zeta = 1$ if t is odd, otherwise $\zeta = 2$. Therefore, $P_{e_P}^{\mathrm{N}} = 1 - \frac{1}{1-F_{\gamma_P}\left(\rho_P^{\mathrm{N}}\right)} \sum_{r=0}^{L} \binom{L}{r} (-1)^r \times \mathcal{F}\left(r, \rho_P^{\mathrm{N}}\right)$. Similarly, $P_{e_S}^{\mathrm{N}}$ follows (6.24) and (6.29) can be used as its solution with $\nu = L$, $\rho = \rho_S^{\mathrm{N}}$,

$$\Psi = \prod_{q=1}^{n} \prod_{t=2q-1}^{2q} \left(\mathcal{M}_S \bar{\gamma}_S^{-1}\right)^{\mathcal{M}_S} \acute{\Gamma}\left(\mathcal{M}_S\right) \left(v_\zeta a_q\right)^{N_t} \tag{6.35}$$

and

$$\Delta = \sum_{q=1}^{n} \sum_{t=2q-1}^{2q} b_q N_t c_{\zeta,S}^2 + \mathcal{M}_P^{-1} \bar{\gamma}_P \tag{6.36}$$

where $\zeta = 1$ if t is odd, otherwise $\zeta = 3$. Therefore,

$$P_{e_S}^{\mathrm{N}} = 1 - \frac{1}{1 - F_{\gamma_S}\left(\rho_S^{\mathrm{N}}\right)} \sum_{r=0}^{L} \binom{L}{r} (-1)^r \times \mathcal{F}\left(r, \rho_S^{\mathrm{N}}\right). \tag{6.37}$$

6.8 Throughput Analysis

The user's throughput, R_j, is the ratio of the total packets received correctly to the total transmitted packets multiplied by the number of bits per symbol, i.e. $\log_2$ of modulation order. Therefore, it is fundamental to account for the receiver's imperfection. For example, the PU throughput depends on the PER only, whereas the throughput of the SU depends on the PER and the errors caused by the classifier and counting rules. In addition, the probabilities of the TMs are crucial to evaluate the users and system throughput.

To compute the TMs probabilities, the SNR thresholds ρ_j^m for the BER constraints shown in Table 6.1 must be found, and then integrated over the PDF of the instantaneous SNR. Since the instantaneous BER of NOMA is a sum of $Q(\cdot)$ function terms with different arguments, it is difficult to find SNR analytically. Therefore, it is suggested to find $\rho_j^m, \forall\{j,m\}$ numerically. By referring to Figure 6.1b, the TMs operation regions can be identified, and hence the conditions can be written in terms of instantaneous SNR and SNR thresholds as follows,

$$\mathrm{TM}_1 : \{\{\gamma_P \geq \rho_P^{\mathrm{O}}\} \cap \{\gamma_P < \rho_P^{\mathrm{N}}\}\} \cup \{\{\gamma_P \geq \rho_P^{\mathrm{N}}\} \cap \{\gamma_S < \rho_S^{\mathrm{N}}\}\} \tag{6.38}$$

$$\mathrm{TM}_2 : \{\gamma_P \geq \rho_P^{\mathrm{N}}\} \cap \{\gamma_S \geq \rho_S^{\mathrm{N}}\} \tag{6.39}$$

$$\mathrm{TM}_3 : \{\gamma_P < \rho_P^{\mathrm{O}}\} \cap \{\gamma_S \geq \rho_S^{\mathrm{O}}\} \tag{6.40}$$

$$\mathrm{TM}_4 : \{\gamma_P < \rho_P^{\mathrm{O}}\} \cap \{\gamma_S < \rho_S^{\mathrm{O}}\}. \tag{6.41}$$

Therefore, the probability of the TMs can be expressed as $\Pr(\mathrm{TM}_1) = F_{\gamma_P}(\rho_P^{\mathrm{N}}) - F_{\gamma_P}(\rho_P^{\mathrm{O}}) + (1 - F_{\gamma_P}(\rho_P^{\mathrm{N}}))(F_{\gamma_S}(\rho_S^{\mathrm{N}}))$, $\Pr(\mathrm{TM}_2) = (1 - F_{\gamma_P}(\rho_P^{\mathrm{N}}))(1 - F_{\gamma_S}(\rho_S^{\mathrm{N}}))$, $\Pr(\mathrm{TM}_3) = F_{\gamma_P}(\rho_P^{\mathrm{O}})(1 - F_{\gamma_S}(\rho_S^{\mathrm{O}}))$ and $\Pr(\mathrm{TM}_4) = F_{\gamma_P}(\rho_P^{\mathrm{O}})F_{\gamma_S}(\rho_S^{\mathrm{O}})$. Consequently, the PU throughput depends on the probability of TM_1 and TM_2, in addition to $P_{e_P}^m$. Therefore, it can be expressed as $R_P = \Pr(\mathrm{TM}_1)(1 - P_{e_P}^{\mathrm{O}}) + \Pr(\mathrm{TM}_2)(1 - P_{e_P}^{\mathrm{N}})$ [184, 185].

On the other hand, while assuming that the throughput is zero when a classification error occurs [168], the SU throughput can be written as $R_S = \Pr(\mathrm{TM}_2)(1 - P_{\mathrm{M}})(1 - P_{e_S}^{\mathrm{N}}) + \Pr(\mathrm{TM}_3)(1 - P_{\mathrm{FA}_2})(1 - P_{e_S}^{\mathrm{O}})$. It is noted that besides $P_{e_S}^m$, R_S is also affected by the imperfections of the classifier and the counting rule. Finally, the system's throughput can be found by summing the users' throughput, i.e. $R = R_P + R_S$. It is worth mentioning that the unit of the throughput is bit/symbol, which translates to the average number of error-free bits received per channel use, where the channel use considers all TMs including the idle mode.

Table 6.4: Normalized complexity.

		$\frac{R_A}{ML}$	$\frac{C}{(M-1)L}$
U_S	OMA -ML	26	1
	NOMA-ML	$26M$	$M+1$
	SIC	43	2
	Total	$69+26M$	$M+4$
U_P	MLD	26	1

6.9 Computational Complexity Analysis

The transmitter and receiver complexity can be evaluated as follows. At the transmitter, the additional complexity is mainly encountered by the TM adaptation process. This requires computing the instantaneous BER for: PU-OMA, SU-OMA and PU/SU-NOMA. The worst-case scenario in terms of complexity occurs when PU/SU-NOMA is evaluated using the formulas in Section 6.3.5, where the BERs are evaluated and compared to the thresholds. Alternatively, the SNR thresholds for each TM can be computed numerically, and hence, the instantaneous SNR for each user can be used to select the appropriate TM as described in (6.38)–(6.41) and Figure 6.1b.

At the PU receiver, the detector is similar to the far user detector in a NOMA system given by (6.2), and thus, its complexity is generally equal to the OMA case. However, this is not the case for the SU because it requires two processing stages, which are classification and detection. Based on the classifier decision, the proper detector will be selected. Table 6.4 summarises the computational complexity of the SU classifier given in (6.6), and PU and SU detectors given in (6.2) and (6.4), respectively. In the table, R_A and C denote the number of real addition and comparison operations, which are normalized to ML and $(M-1)L$, respectively. The number of real multiplications is converted to real additions using the mapping described in [186]. The complexity of the counting rules is negligible, hence they are ignored in the analysis. Further, to reduce the complexity at the SU, the detection can be performed by considering the Euclidean distance computed by the classifier from either the OMA-ML or the NOMA-ML based on the classifier decision. As can be noted from the table, the computational complexity of the proposed system is generally low even for large values of M and L.

6.10 Extension to Higher Number of Users

Extending the proposed scheme for multiple users generally follows the same approach for the sine PU/SU case. As an example, we discuss the case where we have one PU

Table 6.5: Transmission-modes for three users.

	Packet	BER Constraints
TM_2	$PU/SU_1/SU_2$	$\{P_{B_P}^{N_3} \leq \tau_P\} \cap \{P_{B_{S1}}^{N_3} \leq \tau_{S1}\} \cap \{P_{B_{S2}}^{N_3} \leq \tau_{S2}\}$
TM_3	PU/SU_1	$\{P_{B_P}^{N_2} \leq \tau_P\} \cap \{P_{B_{S1}}^{N_2} \leq \tau_{S1}\} \cap \{P_{B_{S2}}^{N_2} > \tau_{S2}\}$
TM_4	PU/SU_2	$\{P_{B_P}^{N_2} \leq \tau_P\} \cap \{P_{B_{S2}}^{N_2} \leq \tau_{S2}\} \cap \{P_{B_{S1}}^{N_2} > \tau_{S1}\}$
TM_5	SU_1/SU_2	$\{P_{B_P}^{N_2} > \tau_P\} \cap \{P_{B_{S1}}^{N_2} \leq \tau_{S1}\} \cap \{P_{B_{S2}}^{N_2} > \tau_{S2}\}$
TM_6	SU_1	$\{P_{B_P}^{O} > \tau_P\} \cap \{P_{B_{S1}}^{O} \leq \tau_{S1}\} \cap \{P_{B_{S2}}^{O} > \tau_{S2}\}$
TM_7	SU_2	$\{P_{B_P}^{O} > \tau_P\} \cap \{P_{B_{S1}}^{O} > \tau_{S1}\} \cap \{P_{B_{S2}}^{O} \leq \tau_{S2}\}$
TM_8	Off	$\{P_{B_P}^{O} > \tau_P\} \cap \{P_{B_{S1}}^{O} > \tau_{S1}\} \cap \{P_{B_{S2}}^{O} > \tau_{S2}\}$

and two SUs, SU_1 and SU_2. While the PU is given the highest priority, the SUs can have similar or different priorities. In the case where different priorities are assumed, SU_1 is given a higher priority than SU_2. Therefore, the BS will select one of the following TMs: PU-OMA, $PU/SU_1/SU_2$-NOMA, PU/SU_1-NOMA, PU/SU_2-NOMA, SU_1/SU_2-NOMA, SU_1-OMA, SU_2-OMA, and Off. The BER constraints for the TMs are given in Table 6.5, where the superscript N_2 and N_3 stands for NOMA with two and three users. The case of TM_1 corresponds to PU-OMA and its probability is given by $\{\{P_{B_P}^{O} \leq \tau_P\} \cap \{P_{B_P}^{N_2} > \tau_P\}\} \cup \{\{P_{B_P}^{N_2} \leq \tau_P\} \cap \{P_{B_{S1}}^{N_2} > \tau_{S1}\} \cap \{P_{B_{S2}}^{N_2} > \tau_{S2}\}\} \cup \{\{P_{B_P}^{N_3} \leq \tau_P\} \cap \{P_{B_{S1}}^{N_3} > \tau_{S1}\} \cap \{P_{B_{S2}}^{N_3} > \tau_{S2}\}\}$.

At the receiving sides, the cognition should be seamless to the PU while a classifier should be used by the SUs to select the proper detector where the packet destination address will be used to either keep or discard the packet. While the SIC decoding order and power coefficients are fixed based on the number of users, the classifier has to distinguish between the OMA, two-user NOMA, and three-user NOMA constellation. As per the analysis of the classifier, the various false alarms and misdetection probabilities have to be found and averaged.

6.11 Numerical Results and Discussion

This section presents the analytical and simulation results of the considered system in terms of throughput, classification errors, and PER where the main performance trends are highlighted and discussed. The results are evaluated considering two values of Nakagami factor which are $\mathcal{M}_j = \mathcal{M} \in \{1,2\} \forall j$. The proposed system results are bench-marked with conventional NOMA. Because all counting rules are special cases of the $\bar{K}$-out-of-L rule, we use the following notations to define all the counting rules with respect to the $\bar{K}$-out-of-L rule, where $\frac{\bar{K}}{L} \triangleq K$. Therefore, the OR rule is denoted as $K = 1/L$, Majority rule is denoted as $K = 1/2$, and AND rule is denoted as $K = 1$. Unless otherwise stated, the power coefficients are fixed to $\alpha_S = 0.1$ and $\alpha_P = 0.9$ for

Table 6.6: Simulation parameters.

Parameter	Value
Modulation scheme	BPSK
Pathloss exponent	$\lambda = 2.7$
Distance from BS	$\Upsilon_P = 1.67d$, $\Upsilon_S = d$
Channel Model	Rayleigh block fading
Block Length	$L = 32$
BER threshold	$\tau_P = \tau_S = 10^{-2}$ [71]
Simulation runs	300×10^3
Address field	Perfect
Classifier	Imperfect
CSIR and CSIT	Perfect/Imperfect

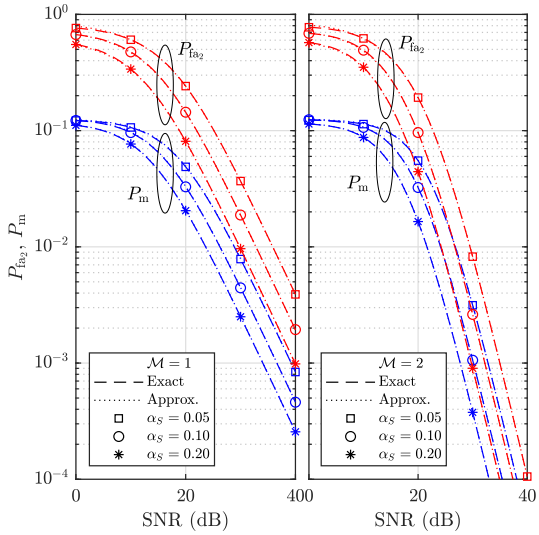


Figure 6.4: Exact and approximate classifier performance versus SNR.

the PU/SU-NOMA transmission, $n = 3$ for the $Q(\cdot)$ function approximation in (6.18) and perfect CSI is assumed at the BS and users. For the case of imperfect CSI at all nodes, the estimated CSI is modelled as $\widehat{g}_j = g_j + \tilde{g}_j$, where $\tilde{g}_j \sim \mathcal{CN}(0, N_0)$ [177]. Table 6.6 summarises the simulation parameters and main assumptions. The SNR in the figures is defined as SNR $\triangleq \frac{1}{2\sigma_{w_j}^2}$ to have a common metric for all the users. It is important to mention that, where applicable, the simulation results are presented in the figures using markers whereas the analytical results are represented by the lines.

Figure 6.4 shows the exact classifier performance for a local observation using the numerical evaluation of (6.15) and the approximate expression using (6.20). The figure shows that the approximate results match closely with the exact results for all SNR and various α_S values. Additionally, it is shown that P_{m} has a lower value compared to P_{fa_2} which is expected because of the weighting factor $1/2$ that appears in $P_{\mathrm{m}}|\gamma_S$ while it vanishes from the $P_{\mathrm{fa}_2}|\gamma_S$. Furthermore, it is seen that P_{m} and P_{fa_2} improve at moderate

and high SNRs for $\mathcal{M} = 2$, where this trend is observed in the next figures.

Figures 6.5 and 6.6 show the exact global misdetection and false alarm performances using the numerical evaluation of (6.24) and the approximate expressions derived in Table 6.3. The figure shows that the approximation matches closely the actual performance for all ranges of K, SNR and α_S. Figure 6.5a shows that P_{M} is negligible for $K = 1/L$, and hence, it can be ignored. It is noted that P_{M} performance can never be worse than the $K = 1$ performance which can be seen as an upper bound. Also, as $K \to 1/L$, P_{M} improves significantly. On the other hand, Figure 6.6a shows that P_{FA_2} is considered negligible for $K = 1$, and hence, it can be ignored. It is noted that P_{FA_2} performance can never be worse than the $K = 1/L$ case which can be seen as an upper bound. Also, as $K \to 1$, P_{FA_2} improves significantly. The effect of imperfect CSI at the BS and the users is quantified in Figure 6.7 in terms of misdetection and false alarm probabilities. It can be seen that at up to 10 dB, the misdetection probability for imperfect CSI $\forall K$ is better than the perfect CSI with $K = 15/16$ or 1, which is due to the fact that at such SNR values the misdetection is mainly due to the assumption that the BS is idle. Nonetheless, the misdetection for imperfect CSI starts degrading after 10 dB as the SU carries the classification process and gets two types of misdetection: 1) Due to classification errors. 2) Due to idle mode assumption. Eventually, the SU will end up mainly with the classification errors only and the reason for worse performance when compared to perfect CSI is typically due to the estimated CSI affecting the classification process. The same can be extended to false alarm probability which is better than perfect CSI with $K = 1/L$ or $1/2$. Further, the types of false alarm are: 1) Due to classification errors. 2) Due to on-mode assumption.

Figure 6.8 shows the exact and approximate PER results of PU and SU for the NOMA and OMA cases. The figure shows that the exact and approximate results match closely for all SNR and α_S ranges. Figure 6.8a shows that the SU OMA achieves the best PER performance, whereas the PU OMA PER saturates at high SNRs, which is justified by the fact that the PU-OMA TM is activated only when the instantaneous SNR is bounded between two particular values for one of the cases as seen in (6.31). Additionally, even though it is an OMA transmission, its performance depends on the selected power coefficients for the NOMA case as ρ_P^{N} is a factor in the PER expression. Furthermore, it is noted from Figure 6.8b that the PU NOMA PER degrades as a higher power coefficient is given to the SU, which is aligned with its BER behaviour. The case is the same for SU NOMA PER, which follows the BER behaviour. It is worth highlighting that the PER is less sensitive to power allocation compared to P_{FA_2} and

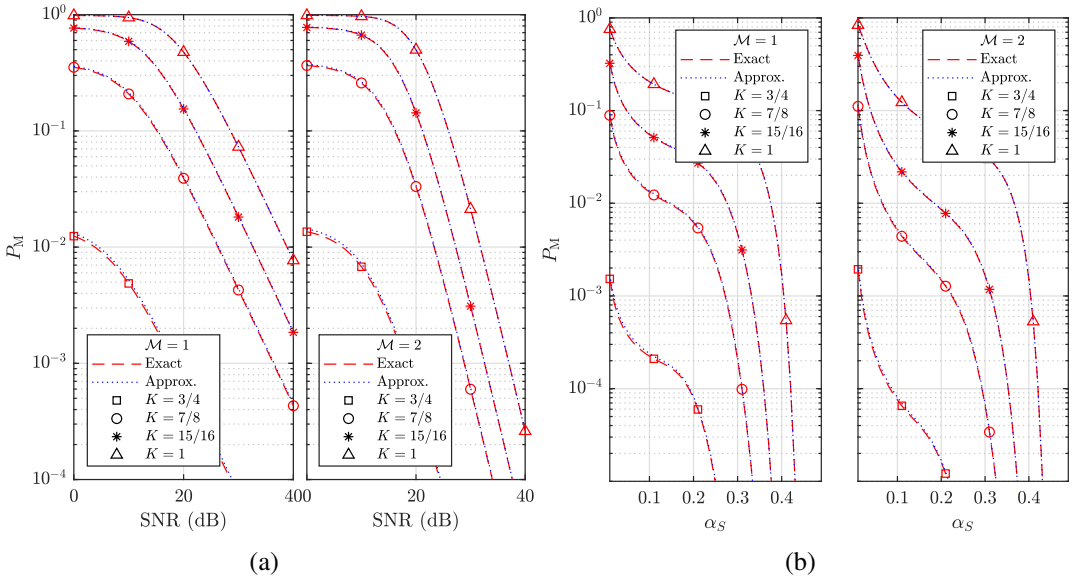


Figure 6.5: Exact and approximate global probability of misdetection: a) versus SNR. b) versus α_S for SNR = 15 dB.

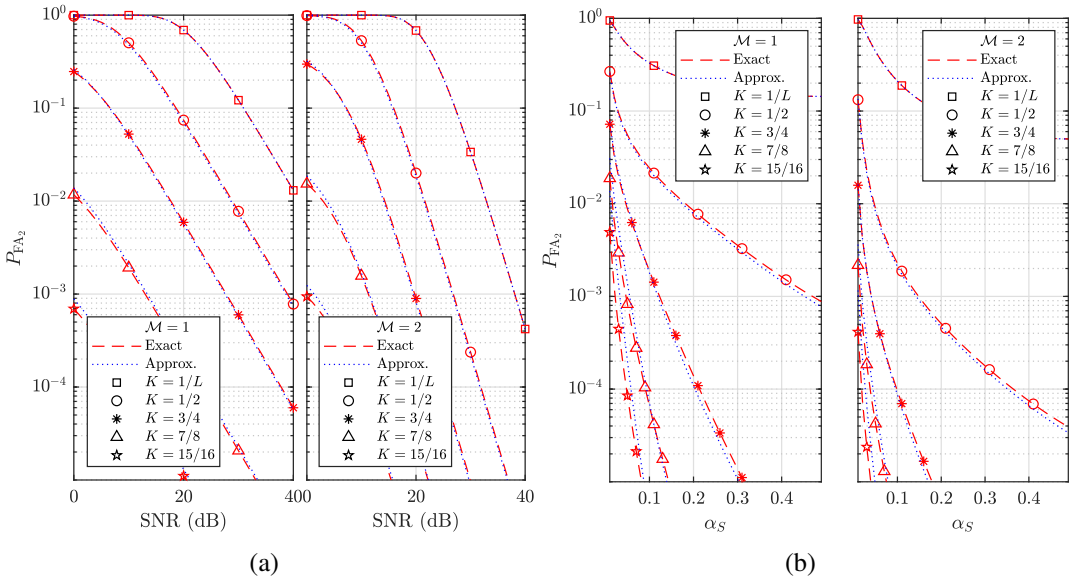


Figure 6.6: Exact and approximate global probability of false alarm: a) versus SNR. b) versus α_S for SNR = 15 dB.

P_{M}. Hence, the power coefficients should be selected to achieve a particular P_{FA_2}, P_{M} and the PER.

The analytical and simulated TMs probabilities are shown in Figure 6.9. It is noted from Figure 6.9a that at low SNRs, TM_4 is dominant because the BER requirement cannot be satisfied for any user. However, as the SNR increases to 6 dB, TM_3 becomes dominant because the SU is closer to the BS, and thus, its BER requirement can be satisfied with OMA, whereas the PU BER requirement cannot be satisfied with OMA

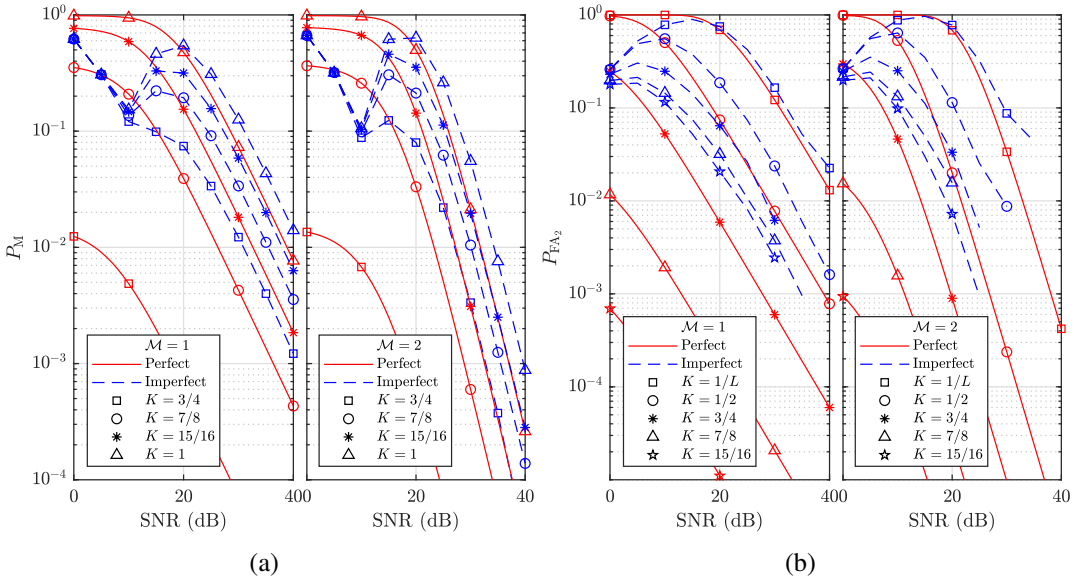


Figure 6.7: Global probability of misdetection and false alarm versus SNR for perfect and imperfect CSI.

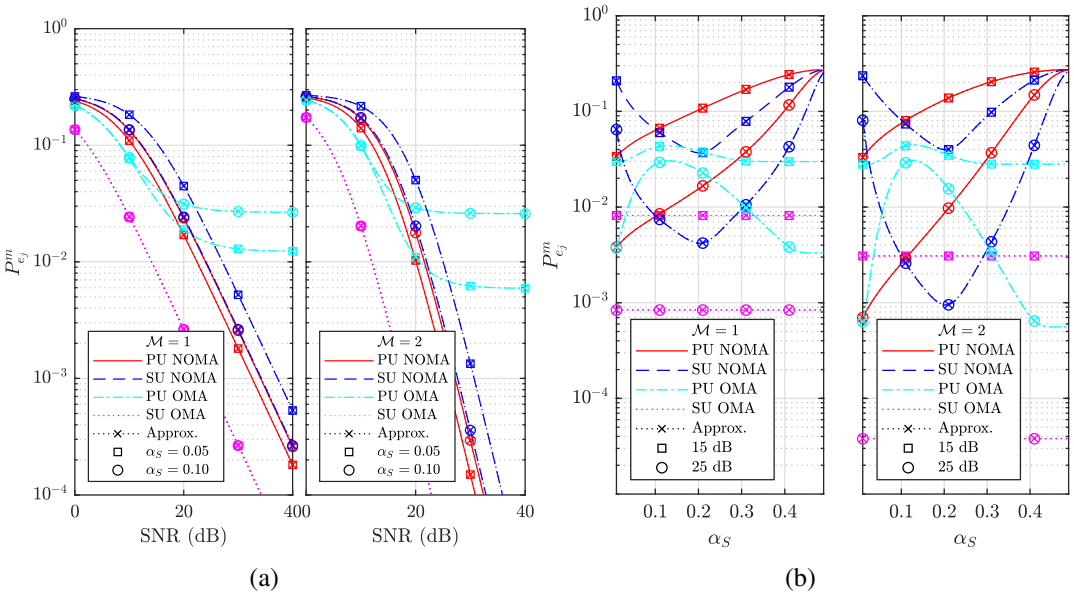


Figure 6.8: Exact and approximate packet error rate: a) versus SNR. b) versus α_S for SNR = 15 and 25 dB.

due to the experienced pathloss. When the PU pathloss is compensated for by the increased SNR, TM_1 takes over TM_3 because the PU is given a higher priority than the SU. It can be seen that when SNR is more than 15 dB, $\Pr(\mathrm{TM}_2)$ becomes greater than $\Pr(\mathrm{TM}_1)$ because NOMA packets can be adopted while satisfying the BER requirements of both users. Moreover, Figure 6.9b shows that $\Pr(\mathrm{TM}_3)$ and $\Pr(\mathrm{TM}_4)$ are independent of the selected power coefficients for NOMA mode as seen in (6.40) and (6.41). Also, improving the Nakagami factor enhanced the probability of TM_2 and the peak of TM_1 and TM_3 which is due to the channels‘ improvements. Additionally, it is

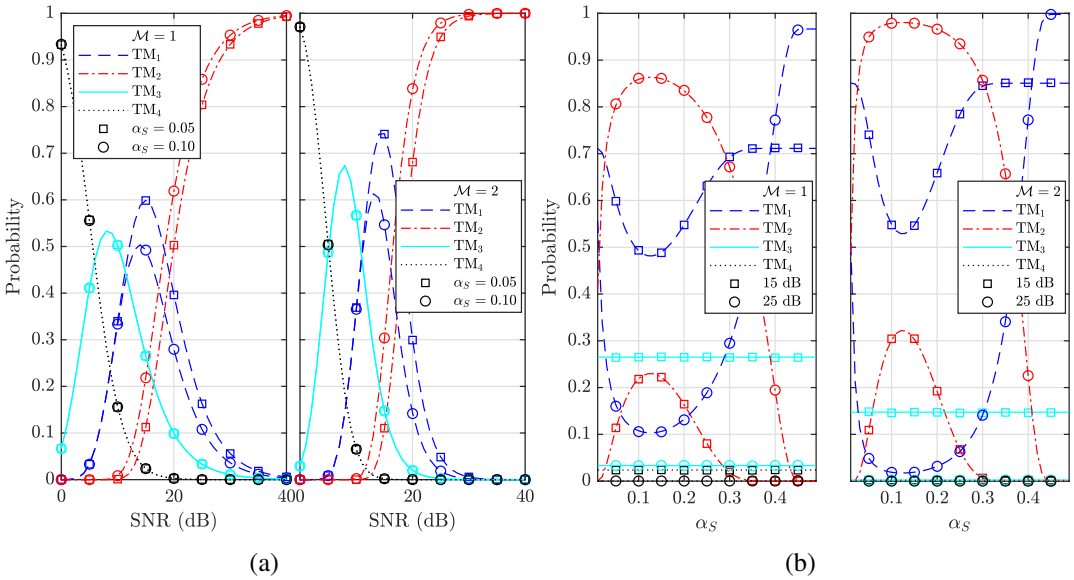
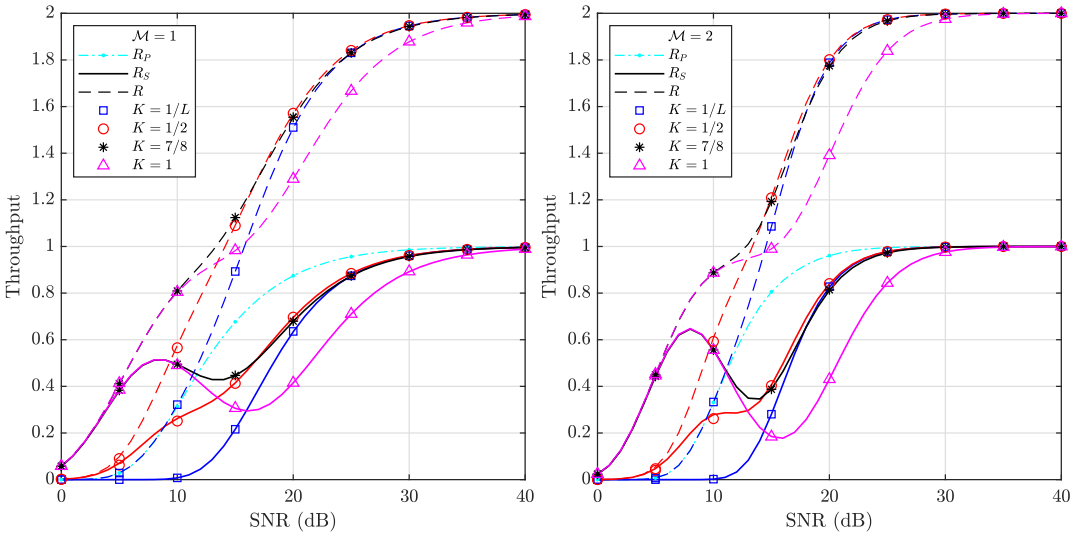


Figure 6.9: Analytical and simulated probability of TMs: a) versus SNR. b) versus α_S for SNR = 15 and 25 dB.

noted that $\alpha_S = 0.1$ maximises $\Pr(\mathrm{TM}_2)$ which is favourable to maximise the system throughput.

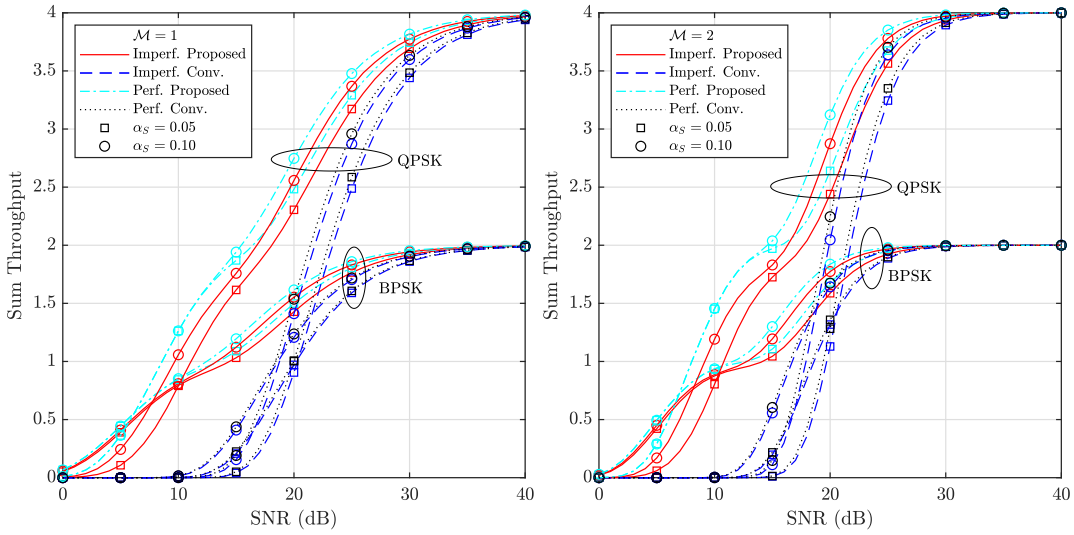
Figure 6.10 shows the analytical and simulated throughput for different K values. It can be seen that optimal R_S performance at low SNRs is achieved with $K = 1$, whereas $K = 1/L$ achieves an optimal R_S performance at high SNRs. For the moderate SNR region, the optimal R_S performance is achieved with $K = 7/8$. This indicates that K should be tuned by the SU based on the average SNR to achieve optimal performance. Furthermore, it can be seen that R_S is higher at low SNRs while R_P is better at moderate and high SNRs before saturation.

Finally, Figure 6.11 shows the sum throughput results for the proposed and conventional NOMA system with two users, where perfect and imperfect receivers are considered, and both BPSK and QPSK results are presented. The figure shows that $\alpha_S = 0.1$ offers higher sum throughput as compared to $\alpha_S = 0.05$. In addition, it shows that the performance gap between the perfect and imperfect receivers is wider for the proposed system because of the classifier and counting rules imperfections. Further, the gap for the QPSK case compared to the BPSK case is wider indicating more frequent classification errors. It is noted that the sum throughput is higher for BPSK case at low SNR values, which is due to its lower SNR threshold requirements compared to the QPSK. Furthermore, the proposed system shows a performance gain over the conventional system for all SNRs before saturation where the performance gain reduces as the SNR improves. For instance, when $\mathcal{M} = 1$ the performance gain is $\sim 9.4 - 11$ dB



(a) (b)

Figure 6.10: Analytical and simulated throughput versus SNR for various K values: a) $\mathcal{M}=1$. b) $\mathcal{M}=2$.



(a) (b)

Figure 6.11: Sum throughput versus SNR for BPSK and QPSK cases, where $K=7/8$: a) $\mathcal{M}=1$. b) $\mathcal{M}=2$.

and $\sim 10-10.1$ dB at 0.5 bit/symbol for BPSK and QPSK, respectively. It reduces to $\sim 5.3-5.8$ dB and $\sim 9.1-9.2$ at 1 bit/symbol for BPSK and QPSK, respectively. The same trend is observed for $\mathcal{M}=2$ but with performance enhancement for both systems due to the channel improvement. The reason for such a performance gap between the proposed and conventional system at moderate SNRs is that system outage is highly likely for the conventional system [174]. Figure 6.12 shows the sum throughput results considering SU and PU at various distance combinations, i.e. near-far, far-near,

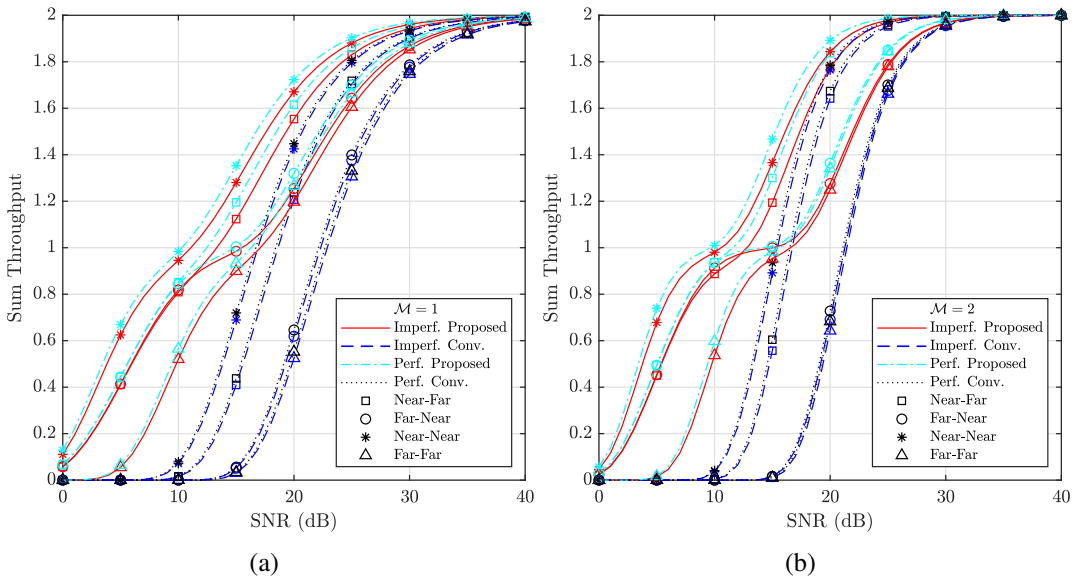


Figure 6.12: Sum throughput versus SNR for BPSK case considering various PU and SU distance combinations, where $K = 7/8$: a) $\mathcal{M} = 1$. b) $\mathcal{M} = 2$.

near-near and far-far. It can be seen that at moderate SNRs, the near-far pair performs better than the far-near pair because the SU in the latter has limited access to the PU spectrum due to its low priority and poor channel conditions. In addition, the near-near pair outperforms the other cases, but such a pairing policy might degrade the overall cell throughput because we might end up pairing two far users.

6.12 Conclusions

To conclude, this chapter analysed the throughput of uncoordinated hybrid interweave-underlay CR-NOMA in Rayleigh block fading channels. The TMs probabilities and average PER performance were derived analytically. Also, the classifier and counting rules performances were presented in closed-form. When the proposed system was benchmarked against a conventional NOMA system, the former revealed up to 10 dB performance gain under some scenarios. It is noted that maximizing the PU/SU-NOMA mode over the PU-OMA mode would maximise the system's throughput. Moreover, the SU performance is limited by the classifier and counting rule performance where the latter has to be finely tuned to get an optimal throughput performance. As a rule of thumb, AND rule is optimal at low SNRs, whereas OR rule is optimal at high SNRs. For moderate SNR values, the $\bar{K}$-out-of-L rule should be tuned to provide optimal performance. However, the tuning depends only on the average SNR, which can be generally acquired with low complexity.

Chapter 7

Design and Analysis of NOMA with Adaptive Modulation and Power

7.1 Chapter Introduction

In this chapter, we derive closed-form expressions for the throughput of non-orthogonal multiple access (NOMA) and use the expressions to maximise the throughput. The design considers a packet-based transmission where the base station optimises the modulation orders and power coefficients while satisfying the block error rate (BLER) requirement for each user. The optimisation problem is solved using integer and mixed-integer programming, where the computational complexity is reduced using an efficient stopping criterion, segment-line search, and quantised signal to noise ratios (SNRs). Furthermore, the analytical and simulation results show that selecting the appropriate power and modulation orders can improve the throughput by about 3 dB at high SNRs. The impact of the SNR quantisation is evaluated, and the obtained results show that the proposed system can tolerate large quantisation steps, which enables reducing the complexity of the optimisation and adaptation processes. Furthermore, the optimisation for the fixed power case is simplified, where the objective is transformed to maximising the throughput by minimising the maximum SNR requirements of the users. The optimisation problem is formulated as a min-max problem to jointly optimise the modulation orders and SNR thresholds, where the original mixed-integer programming is simplified to integer programming by introducing an auxiliary variable. Compared to the grid search approach, the analytical and simulation results show that the min-max approach can improve the throughput with a significant reduction in the number of transmission modes, the throughput can be improved by up to 2.5 dB at

moderate SNRs and by 1 bit/symbol at extremely high SNRs.

7.1.1 Chapter Organisation

The remainder of the chapter is organised as follows. Section 7.2 is the related work and motivations and contributions. Section 7.3 demonstrates the system and channel models. Section 7.4 illustrates the adaptation process at the BS. Section 7.5 derives the NOMA throughput while the analytical and simulation results are shown and discussed in Section 7.6. Finally, Section 7.7 is the conclusion.

7.2 Related Work

Recently NOMA has attracted much attention as it can achieve improved spectral efficiency, support massive connectivity and serve users with diverse quality of service (QoS) requirements [24]. To augment NOMA advantages, adaptive modulation and coding (AMC) have been widely considered [70, 71, 73, 74, 187–189]. For example, Wang *et al.* [74] propose an asymmetric adaptive modulation algorithm based on BLER requirements for uplink (UL) two-user NOMA using only binary phase-shift keying (BPSK) and quadrature phase-shift keying (QPSK). The authors of [187] propose constructing an AMC and power coefficients map for the downlink (DL) via an offline link-level NOMA simulator. Similar to the link-level simulator approach, [188] designs an algorithm aiming at maximising the minimum user rate while achieving user fairness. The authors of [71] consider a two-user DL multi-carrier NOMA and propose an efficient adaptive modulation algorithm to maximise the throughput while satisfying individual users' bit error rate (BER) constraints. However, the throughput is evaluated only using simulation and the system adopts fixed power assignment (FPA). Similarly, Cejudo *et al.* [70] design an efficient resource allocation algorithm to maximise the sum rate while satisfying the BER constraints.

Jia *et al.* [189] derived the average spectral efficiency (SE) for the UL two-user NOMA while assuming FPA and adaptive square M-ary quadrature amplitude modulation (M-QAM). Yu *et al.* [73] proposed a joint adaptive modulation and power assignment (PA) algorithm for the DL to achieve a minimum SE target for one of the users and maximise the SE of the other user without sacrificing the BER performance. It is worth mentioning that [73, 189] used the single-user square M-QAM BER expression rather than the NOMA exact BER expressions. The authors of [190] proposed

a joint PA, channel assignment, and modulation order selection to maximise the effective throughput of a DL multi-user multi-channel NOMA system with imperfect successive interference cancellation (SIC) and square M-QAM. In [27], the necessary and sufficient conditions are derived for the power ranges of NOMA with arbitrary M-QAM modulation orders and number of users to prevent error floors when SIC is used. However, the derived power ranges do not specify the optimal power required to achieve a certain BER. Choi [191] studied the PA problem for two-user DL NOMA considering square M-QAM to maximise the mutual information. Dong *et al.* [79] found the optimal power and phase for two-user UL NOMA with square M-QAM to maximise the minimum Euclidian distance of the received sum-constellation at the receiver side. Han *et al.* [134] derived BER and symbol error rate (SER) expressions for DL NOMA with joint-Gray labelling considering QPSK and arbitrary number of users. Moreover, the BER of DL NOMA system is derived in [24] for an arbitrary number of users and M-QAM modulation orders, where the optimal PA is found considering statistical channel state information (CSI).

In [192], adaptive Bose–Chaudhuri–Hocquenghem (BCH) coding with constellation rotation is proposed to improve the NOMA BER. Qiu *et al.* [110] propose an algebraic constellation rotation design for a two-user DL NOMA in block fading channels. However, BLER is evaluated via computer simulation. Approximate BLER expressions for short-packet communication NOMA are derived in [193–198]. In [193], a two-user DL NOMA is considered with fountain-coding, [194] considered multiple-input multiple-output (MIMO)-NOMA system, [195, 196] studied Internet of Things (IoT) networks, [197] studied intelligent reflecting surface (IRS)-assisted NOMA, and [198] studied residual hardware impairments and channel estimation errors.

7.2.1 Motivations and Contributions

Despite the advantages of NOMA, its BER remains generally the main limitation due to the inter-user interference (IUI). Therefore, the power and modulation order for each user has to be carefully selected to ensure that all users can jointly achieve their QoS requirements. Consequently, the power and modulation order should be selected based on the channel conditions and QoS for each user. Therefore, this chapter considers:

- The design of a NOMA system with adaptive power and M-QAM modulation orders.

- An efficient low-complexity algorithm is proposed to realize the modulation and power adaptation process.
- An efficient min-max optimisation is proposed to realize the modulation adaptation process with the least number of transmission modes (TMs).
- To enable efficient design, we derive exact closed-form expressions for the instantaneous throughput (iTp) and closed-form approximation for the average throughput (aTp).
- A novel approach is developed to correlate the selected modulation orders and power, and thus, no side information about the power is required at the receiver side.

It is worth highlighting that [70] considered resource allocation for multicarrier NOMA under BER constraints. Also, the throughput is evaluated through Monte-Carlo simulation. While we consider the DL NOMA, [189] considers the UL NOMA, and derive the information-theoretic SE under the single-user uncoded BER constraint, FPA, and adaptive square M-QAM.

7.3 Systems and Channel Models

This work considers a packet-switched single cell that consists of a base station (BS) that serves two-user NOMA using a single resource block. The BS adapts the modulation orders and power coefficients to maximise the iTp while satisfying the users' instantaneous BLER. The adaptation is applied on a packet-by-packet basis while assuming a block fading channel that is static for L symbols, which is the packet duration [110, 193]. Without loss of generality, the near and far users are denoted as U_1 and U_2, and the allocated power coefficients are α_1 and α_2, where $\alpha_1 + \alpha_2 = 1$. The BS informs the users about the transmission parameters and the users perfectly report their CSI and QoS requirements to the BS through an error-free feedback channel. The received signal for $U_n \forall n \in \{1,2\}$ is

$$\mathbf{y}_n = \mathbf{H}_n(\sqrt{\alpha_1}\mathbf{x}_1 + \sqrt{\alpha_2}\mathbf{x}_2) + \mathbf{w}_n \tag{7.1}$$

where $\mathbf{y}_n \in \mathbb{C}^{L\times 1}$, the data packet $\mathbf{x}_n = [x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(L)}]^T$, the symbols are drawn uniformly from a BPSK constellation or a rectangular M-QAM. For each data symbol,

$\mathbb{E}[|x_n^{(l)}|^2] = 1\ \forall l$, and $\mathcal{M}_n = \log_2 M_n$, where M_n is the modulation order, $\mathbb{E}[\cdot]$ denotes the statistical expectation, and $(\cdot)^T$ denotes the transpose operation. The additive white Gaussian noise (AWGN) vector $\mathbf{w}_n \sim \mathcal{CN}(0, N_0\mathbf{I}_L)$ where $\sigma^2_{w_n} = N_0/2$ and $\mathbf{I}_L$ is the identity matrix with a size of $L \times L$. The complex channel gain for single-input-single-output (SISO) channel is $\mathbf{H}_n \in \mathbb{C}^{L\times L}$, which corresponds to block Rayleigh fading over a one packet interval, that is $\mathbf{H}_n = h_n\mathbf{I}_L$, $h_n = \check{h}_n \times d_n^{-\lambda/2}$, $\check{h}_n \sim \mathcal{CN}(0,1)$, d_n is the distance between the BS and the nth user, and λ is the pathloss exponent. The small-scale channel gains are assumed to be mutually independent and identically distributed (i.i.d). In this work, we consider using a symbol-by-symbol joint-multiuser maximum likelihood detector (JMLD) with perfect CSI [24, Eq. (5)], which is generally equivalent to SIC where both account for the imperfections caused by the IUI [23, 27]. Nonetheless, the SIC is limited to a certain PA range for reliable detection [27]. Therefore, by dropping the symbol index, the jointly detected symbols can be written as

$$\{\widehat{x}_1, \widehat{x}_2\} = \arg\min_{x_i \in \chi_i} \left| y_n - h_n \sum\nolimits_{i=1}^{2} \sqrt{\alpha_i} x_i \right|^2 \tag{7.2}$$

where x_i are the trail symbols taken from the user's reference constellation χ_i.

7.4 Adaptation Process at the Base Station

7.4.1 Preliminaries

Considering FPA and adaptive power assignment (APA), the instantaneous BLER can be expressed as [199, Eq. 6.21],

$$P_{e_n}|\gamma_n = 1 - (1 - P_{B_n}|\gamma_n)^{\mathcal{M}_n L} \tag{7.3}$$

where $P_{B_n}|\gamma_n$ is the instantaneous BER which is a function of α_1, $\mathcal{M}_1$, $\mathcal{M}_2$ and γ_n. In practice, IUI causes imperfections in the detection and affects the BER. Therefore, the instantaneous BER expressions accounting for these imperfections can be written in general as [24, Eq. (23)],

$$P_{B_n}|\gamma_n = \frac{1}{2^{\mathcal{M}-2}\mathcal{M}_n} \sum\nolimits_q c_q Q\left(\sqrt{|\Delta_q|^2 \gamma_n}\right) \tag{7.4}$$

where $Q(\cdot)$ is the Gaussian complementary cumulative distribution function (CDF), $\mathcal{M} = \sum_n \mathcal{M}_n$, c_q is a weighting coefficient and Δ_q is the distance between the constellation points and the decision regions' boundaries [24]. The MATLAB code to generate (7.4) is available in [200]. Note that the factor $2^{\mathcal{M}-2}$ becomes $2^{\mathcal{M}-1}$ if $\mathcal{M}_n = 1, \forall n$. Furthermore, the iTp for each user is defined as the probability of receiving a certain block successfully multiplied by the number of bits per symbol,

$$R_n|\gamma_n = (1 - P_{e_n}|\gamma_n)\,\mathcal{M}_n. \tag{7.5}$$

Therefore, the iTp is defined as the sum of the instantaneous users' throughputs, i.e. $\sum_{n=1}^{2} R_n|\gamma_n \triangleq R$. It is worth noting that the mutual information can be alternatively used. Nonetheless, it is derivation is tedious for arbitrary modulation orders. Therefore, it is outside the scope of this work.

7.4.2 Grid Search Approach

Inspired by the AMC systems in [187], a tabulated adaptation is proposed based on quantising the instantaneous SNR of both users, which is given as $\gamma_n = \frac{2|h_n|^2}{N_0}$. The quantisation levels for both users are identical, and the number of quantisation levels is given by $k = \frac{\gamma_{\max} - \gamma_{\min}}{\Delta\gamma} + 1$, where $\gamma_{\min}$ is the cut-off outage level that satisfies both users with minimum modulation orders, $\gamma_{\max}$ is the maximum SNR to be considered, and $\Delta\gamma$ determines the resolution of the quantisation. Therefore, $k \times k$ TMs are formed in a grid such that each $\mathrm{TM}_{i,j}$, $\forall i, j \in \{1, \ldots, k\}$, is characterized by the $(\alpha_1, \mathcal{M}_1, \mathcal{M}_2)$ tuple. The conditions to select $\mathrm{TM}_{i,j}$ are $\rho^1_{i,j} \leq \gamma_1 < \rho^1_{i+1,j}$ and $\rho^2_{i,j} \leq \gamma_2 < \rho^2_{i,j+1}$, where $\rho^1_{i,j} = \gamma_{\min} + (i-1)\Delta\gamma$ and $\rho^2_{i,j} = \gamma_{\min} + (j-1)\Delta\gamma$. It is worth noting that the boundary conditions are $\rho^1_{k+1,j} = \rho^2_{i,k+1} = \infty$. A look-up table (LUT) is used to store the TMs conditions and the corresponding tuple $(\alpha_1, \mathcal{M}_1, \mathcal{M}_2)$. It is noted that as $\Delta\gamma$ gets smaller, a higher resolution grid with a larger number of TMs will be created, which provides a better approximation of the continuous adaptation scenario. The LUT generation has an objective to maximise R by optimising the TM parameters. For the FPA scheme, the LUT generation is an integer optimisation problem,

$$(\mathcal{M}_1^*, \mathcal{M}_2^*) = \arg\max_{\mathcal{M}_1, \mathcal{M}_2} R \tag{7.6}$$

while the LUT generation for the APA scheme is a mixed-integer optimisation problem,

$$(\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*) = \arg \max_{\alpha_1, \mathcal{M}_1, \mathcal{M}_2} R. \tag{7.7}$$

Both optimisation problems are subject to satisfaction of a predefined BLER threshold, τ_n, such that

$$P_{e_n}|\gamma_n \leq \tau_n, \forall n. \tag{7.8}$$

7.4.2.1 Adaptive modulation with FPA

A possible FPA scheme is the equally spaced constellation points PA which depends on the selected modulation orders. Hence, informing the users about the modulation orders is sufficient to identify the selected power coefficients at the BS because such PA results in a constellation diagram that is similar to the single-user case, has equally spaced constellation points, with a modulation order of $M_1 \times M_2 \times \cdots \times M_N$. Moreover, the bit mapping of the constellation points still follows the NOMA conventional mapping. The equally spaced constellation points PA condition for $N = 2$ case can be found considering the real part of the nearest constellation point to the origin, $s_0 = -\Lambda_1 \sqrt{\alpha_1 \kappa_1^{-1}} + \sqrt{\alpha_2 \kappa_2^{-1}}$, and the point next to it, $s_1 = -(\Lambda_1 - 2)\sqrt{\alpha_1 \kappa_1^{-1}} + \sqrt{\alpha_2 \kappa_2^{-1}}$, where κ_n and Λ_n are respectively the normalization factor and width for U_n constellation [24]. Hence, $2s_0 = s_1 - s_0$ ensures equally spaced constellation points. After some mathematical manipulation and generalization, the condition can be expressed as follows,

$$\frac{\alpha_n}{\alpha_{n+1}} = \frac{\kappa_n}{(\Lambda_n + 1)^2 \kappa_{n+1}}, \ \forall n \in \{1, 2, \ldots, N-1\} \tag{7.9}$$

which results in a system of equations that need to be solved to find $\alpha_1, \alpha_2, \ldots, \alpha_{N-1}$. By noting that $\alpha_2 = 1 - \alpha_1$ for $N = 2$, the closed-form PA can be expressed as

$$\alpha_1 = \frac{\kappa_1}{\Lambda_1 \kappa_2 (\Lambda_1 + 2) + \kappa_1 + \kappa_2} \tag{7.10}$$

which is more general than [70, Eq. (10)] because the latter is applicable for square M-QAM only, while (7.10) is applicable for the general case of rectangular M-QAM. The objective function in (7.6) has a low complexity as the solver only needs to compute $P_{e_n}|\gamma_n$ and $R_n|\gamma_n$ $\forall n$ using (7.3)–(7.5) at the FPA found in (7.10) for a limited number of cases that correspond to all possible combinations of $\mathcal{M}_1$ and $\mathcal{M}_2$, and select the pair

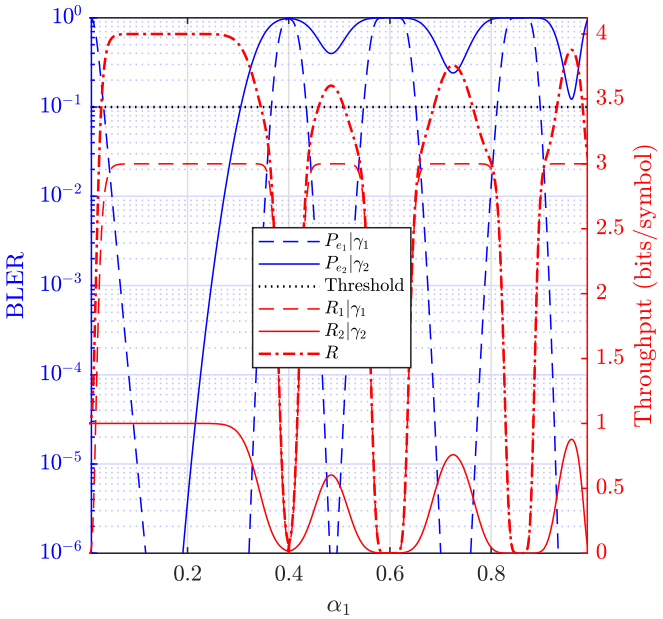


Figure 7.1: Instantaneous BLER and throughput, where $\mathcal{M}_1 = 3$, $\mathcal{M}_2 = 1$, $\gamma_1 = 32$ dB, $\gamma_2 = 29$ dB.

that maximises R while satisfying (7.8). While noting that the computational complexity to compute (7.4) increases as $\mathcal{M}_n$ becomes larger [24], using brute-force search will have a time complexity of $O\left(k^2\mathcal{M}_{\max}^2\right)$, where $\mathcal{M}_{\max}$ is the maximum $\mathcal{M}_n$ allocated $\forall n$. Considering the fact that increasing the modulation order would increase the BLER for a fixed SNR can be useful to define search heuristics. Hence, reducing the optimisation complexity. Therefore, the search for $(\mathcal{M}_1^*, \mathcal{M}_2^*)$ can be performed in ascending order of $\mathcal{M}$, and the search would stop once the BLER constraints are violated. The function for solving (7.6) can be described as $f : \left(\rho_{i,j}^1, \rho_{i,j}^2, \tau_1, \tau_2\right) \longmapsto (\mathcal{M}_1^*, \mathcal{M}_2^*)$, where f finds $(\mathcal{M}_1^*, \mathcal{M}_2^*)$ that maximises R under $(\rho_{i,j}^1, \rho_{i,j}^2, \tau_1, \tau_2)$ conditions. Note that $(\mathcal{M}_1^*, \mathcal{M}_2^*) \longmapsto \alpha_1^*$ using (7.10).

7.4.2.2 Adaptive modulation with APA

Solving the optimisation problem in (7.7) is generally not straightforward because the objective function exhibits multiple local maximum points for R, as depicted in Figure 7.1. Such behaviour is due to the fact that certain α_1 values cause constellation points overlap, and hence detection ambiguity. These points can be found by considering the following condition,

$$(a_{\mathrm{R}} - a_{\mathrm{L}})\sqrt{\alpha_1\kappa_1^{-1}} + (b_{\mathrm{R}} + b_{\mathrm{L}})\sqrt{(1-\alpha_1)\kappa_2^{-1}} = 0 \tag{7.11}$$

where $a_{\mathrm{R}}, a_{\mathrm{L}} \in \{\pm 1, 3, \ldots, \Lambda_1\}$, $b_{\mathrm{R}}, b_{\mathrm{L}} \in \{\pm 1, 3, \ldots, \Lambda_2\}$. Solving (7.11) for all possible combinations of $a_{\mathrm{R}}, a_{\mathrm{L}}, b_{\mathrm{R}}, b_{\mathrm{L}}$ results in a set of $\alpha_1 > 0$ and $\alpha_1 < 1$ values that impose the constellation points overlap, which can be expressed as

$$\mathbb{A} = \frac{\kappa_1 \left(b_{\mathrm{R}} + b_{\mathrm{L}}\right)^2}{\kappa_1 \left(b_{\mathrm{R}}^2 + b_{\mathrm{L}}^2 + 2 b_{\mathrm{R}} b_{\mathrm{L}}\right) + \kappa_2 \left(a_{\mathrm{R}}^2 + a_{\mathrm{L}}^2 - 2 a_{\mathrm{R}} a_{\mathrm{L}}\right)}. \tag{7.12}$$

Therefore, to avoid the local maximum problem, it is important to find all local maximum points by segmenting the search range and eventually finding the global maximum. It is worth noting that after the first overlap with respect to α_1, the constellation points become very close to each other, which degrades the performance severely. Nonetheless, the Euclidean distance between the constellation points starts increasing after the last overlap with respect to α_1. Therefore, investigating the two extreme ranges with respect to α_1, i.e., $[0, \min(\mathbb{A})]$ and $[\max(\mathbb{A}), 1]$, is sufficient to find the global maximum. Hence, the optimisation complexity is reduced compared to the full-line search.

Furthermore, since brute-force search for $(\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*)$ is computationally expensive, a stopping criterion similar to the FPA case is considered to reduce complexity. The pseudocode for the mixed-integer optimisation problem is given in Algorithm 7.1. The function in line 12, fconopt$(\cdot)$, is based on line-search and used to solve constrained maximisation problems where the arguments are as follows, (7.7) is the objective function, x_0 is the initial guess, lb and ub are the lower and upper bounds of the optimisation variable, (7.8) are the non-linear constraints. The algorithm solving (7.7) can be described as $g : \left(\rho_{i,j}^1, \rho_{i,j}^2, \tau_1, \tau_2\right) \longmapsto (\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*)$, where g finds $(\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*)$ that maximises R under $(\rho_{i,j}^1, \rho_{i,j}^2, \tau_1, \tau_2)$.

7.4.2.3 Practical considerations

Since the BS informs the users about $(\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*)$, the adaptation signaling overhead should be considered for $\Delta\gamma \to 0$ cases as the LUT size becomes massive. Assuming that the maximum and minimum modulation orders are 256 and 2, respectively, i.e. $1 \leq \mathcal{M}_n \leq 8$. This represents 8 possibilities, which requires 3 bits to represent $\mathcal{M}_n, n \in (1,2)$, or $3N$ bits for N users. Moreover, since $0 < \alpha_1 < 1$ then using 8 to 16 bits provides negligible quantisation error. Therefore, the total required bits to convey the power information can be $16(N-1)$, which is a significant signalling overhead when compared to the modulation orders. Therefore, to limit the signalling overhead, $\Delta\gamma$ should be optimised to limit the LUT size, or FPA should be considered.

Algorithm 7.1: Adaptive modulation and APA

Input: $\left(\rho_{i,j}^{1}, \rho_{i,j}^{2}, \tau_1, \tau_2\right)$

Output: $(\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*)$

$\gamma_1 = \rho_{i,j}^{1}$, $\gamma_2 = \rho_{i,j}^{2}$, $R^* = \mathcal{M}_1^* = \mathcal{M}_2^* = 0$

for $\mathcal{M}_1 = 1 : \mathcal{M}_{\max}$ **do**

  for $\mathcal{M}_2 = 1 : \mathcal{M}_{\max}$ **do**

    Compute $\mathbb{A}$ using (7.12)

    $\alpha_{1,\text{seg}}^{\min} = \min(\mathbb{A})$, $\alpha_{1,\text{seg}}^{\max} = \max(\mathbb{A})$

    $\mathbf{key}_l = [0, \alpha_{1,\text{seg}}^{\max}]$, $\mathbf{key}_u = [\alpha_{1,\text{seg}}^{\min}, 1]$

    for $k = 1 : 2$ **do**

      $\text{lb} = \text{key}_l^{(k)}$, $\text{ub} = \text{key}_u^{(k)}$

      $x_0 = 0.5\,(\text{lb} + \text{ub})$

      $\left[\check{f}_{\text{val}}^{(k)}, x_{\text{seg}}^{(k)}\right] = \text{fconopt}\left((7.7), x_0, \text{lb}, \text{ub}, (7.8)\right)$

      if $P_{e_n} | \gamma_n > \tau_n, \forall n$ **then**

        $\check{f}_{\text{val}}^{(k)} = \text{NaN}$

    $f_{\text{val}}^{(\mathcal{M}_1, \mathcal{M}_2)} = \max\left(\check{\mathbf{f}}_{\text{val}}\right)$

    $\alpha_1^{(\mathcal{M}_1, \mathcal{M}_2)} = x_{\text{seg}}^{\left(\arg\max\left(\check{\mathbf{f}}_{\text{val}}\right)\right)}$

    if $f_{\text{val}}^{(\mathcal{M}_1, \mathcal{M}_2)} > R^*$ **then**

      $R^* = f_{\text{val}}^{(\mathcal{M}_1, \mathcal{M}_2)}$

      $(\alpha_1^*, \mathcal{M}_1^*, \mathcal{M}_2^*) = \left(\alpha_1^{(\mathcal{M}_1, \mathcal{M}_2)}, \mathcal{M}_1, \mathcal{M}_2\right)$

    if $\mathcal{M}_2 > 1$, $f_{\text{val}}^{(\mathcal{M}_1, \mathcal{M}_2)} = f_{\text{val}}^{(\mathcal{M}_1, \mathcal{M}_2 - 1)} = \text{NaN}$ **then**

      break

  if $\mathcal{M}_1, > 1$, $f_{\text{val}}^{(\mathcal{M}_1, \mathcal{M}_2)} = f_{\text{val}}^{(\mathcal{M}_1 - 1, \mathcal{M}_2)} = \text{NaN}$ **then**

    break

7.4.3 Min-Max Approach

Inspired by the fact that the time complexity and the number of TMs grow significantly by increasing the resolution of the grid, we propose a simpler approach to reduce the number of TMs. While noting that as $\tau_n \to 0$, $P_{e_n}|\gamma_n \approx 0$. Hence, $R \approx \mathcal{M}$. Consequently, minimising the maximum SNR requirement of the users would also maximise R for a given $\mathcal{M}$. Therefore, the objective function can be formulated as a mixed-integer programming optimisation for a given $\mathcal{M} = \mathcal{M}_t \in \{2, \ldots, 2\mathcal{M}_{\max}\}$,

$$(\rho_1^*, \rho_2^*, \mathcal{M}_1^*, \mathcal{M}_2^*) = \arg \min_{\mathcal{M}_1, \mathcal{M}_2, \gamma_1, \gamma_2} \max\{\gamma_1, \gamma_2\} \tag{7.13a}$$

which is subject to

$$\mathcal{M} = \mathcal{M}_t \tag{7.13b}$$

$$\gamma_n > 0, \forall n \tag{7.13c}$$

$$\gamma_1 > \gamma_2, \tag{7.13d}$$

$$P_{e_n}|\gamma_n \leq \tau_n, \forall n. \tag{7.13e}$$

It is important to mention that (7.13b) is an equality constraint that will allow a heuristic search in all possible combinations of $\mathcal{M}_1$ and $\mathcal{M}_2$ that result in $\mathcal{M}_t$ bit/symbol. For instance, if $\mathcal{M}_t = 4$, then there are three $\{\mathcal{M}_1, \mathcal{M}_2\}$ combinations that satisfy this condition which are $\{3,1\}$, $\{1,3\}$ and $\{2,2\}$. On the other hand, the inequality constraint (7.13c) ensures that γ_n is in its allowed space, (7.13d) ensures the near user has better channel conditions compared to the far user, (7.13e) ensures the satisfaction of a predefined BLER threshold, τ_n.

Since the optimisation problem is mixed-integer programming, its complexity is relatively high even when using efficient search algorithms. Thus, a low-complexity integer programming optimisation problem is proposed, which is based on defining an auxiliary variable $\rho_n = g(\mathcal{M}_1, \mathcal{M}_2, \alpha_1, L, \tau_n)$ which represents the γ_n at which the instantaneous BLER for U_n strictly satisfies the BLER threshold τ_n. The function $g(\cdot)$ can be found by solving (7.3) for γ_n and letting $P_{e_n}|\gamma_n = \tau_n$. Therefore,

$$(\rho_1^*, \rho_2^*, \mathcal{M}_1^*, \mathcal{M}_2^*) = \arg \min_{\mathcal{M}_1, \mathcal{M}_2} \max\{\rho_1, \rho_2\} \tag{7.14a}$$

Table 7.1: Summary of grid search and min-max approaches.

	Grid Search	Min-Max
optimisation	Integer programming	Integer programming
Objective	$\max R$	$\min\max\{\rho_1,\rho_2\}$
Variables	$\mathcal{M}_1,\mathcal{M}_2$	$\mathcal{M}_1,\mathcal{M}_2,\mathcal{M}_t$
Solution	$\mathcal{M}_1^*,\mathcal{M}_2^*$	$\rho_1^*,\rho_2^*,\mathcal{M}_1^*,\mathcal{M}_2^*$
TMs	k^N	$N(\mathcal{M}_{\max}-1)+1$
Complexity	$O\left(k^N\mathcal{M}_{\max}^N\right)$	$O\left(\mathcal{M}_{\max}^N\right)$

which is subject to

$$\mathcal{M} = \mathcal{M}_t \tag{7.14b}$$

The function for solving (7.14a) can be described as $\omega:(\mathcal{M}_t,\tau_1,\tau_2)\longmapsto(\rho_1^*,\rho_2^*,\mathcal{M}_1^*,\mathcal{M}_2^*)$, where ω finds $(\rho_1^*,\rho_2^*,\mathcal{M}_1^*,\mathcal{M}_2^*)$ that minimises the maximum SNR requirements of the users under $(\mathcal{M}_t,\tau_1,\tau_2)$ conditions. Note that $(\mathcal{M}_1^*,\mathcal{M}_2^*)\longmapsto\alpha_1^*$ using (7.10). Using brute-force search for the whole combinations of $\mathcal{M}_1$ and $\mathcal{M}_2$ that are constrained by $\mathcal{M}_t$ bit/symbol will have a time complexity of $O(\mathcal{M}_t-1)$, while the time complexity to consider all values of $\mathcal{M}_t$ is $O\left(\mathcal{M}_{\max}^2\right)$. It is worth mentioning that extending this approach to a higher number of users will have an increase in the time complexity such that $O\left(\mathcal{M}_{\max}^N\right)$. Table 7.1 summarizes the main similarities and differences of the grid search and the min-max approaches.

It is worth noting that the number of TMs is $2\mathcal{M}_{\max}-1$ which is less than $k\times k$ TMs as in the grid search approach. The TMs in the min-max approach are denoted by $\mathrm{TM}_{\mathcal{M}_t}$, $\mathcal{M}_t\in\{2,\ldots,2\mathcal{M}_{\max}\}$, is characterized by the $(\alpha_1,\mathcal{M}_1,\mathcal{M}_2)$ tuple. The conditions to select $\mathrm{TM}_{\mathcal{M}_t}$ can be one of the following:

$$\rho_n^{*(\mathcal{M}_t)} \le \gamma_n < \rho_n^{*(\check{\mathcal{M}}_t)}\,\forall n \tag{7.15}$$

$$\{\gamma_1 \ge \rho_1^{*(\check{\mathcal{M}}_t)}\}\cap\{\rho_2^{*(\mathcal{M}_t)} \le \gamma_2 < \rho_2^{*(\check{\mathcal{M}}_t)}\} \tag{7.16}$$

$$\{\gamma_2 \ge \rho_2^{*(\check{\mathcal{M}}_t)}\}\cap\{\rho_1^{*(\mathcal{M}_t)} \le \gamma_1 < \rho_1^{*(\check{\mathcal{M}}_t)}\} \tag{7.17}$$

where $\check{\mathcal{M}}_t$ is the succeeding $\mathcal{M}_t$, $\rho_n^{*(\check{\mathcal{M}}_t)}=\infty$ when $\mathcal{M}_t=2\mathcal{M}_{\max}$. Furthermore, no transmission takes place if $\gamma_n<\rho_n^{*(\mathcal{M}_t=2)}$ for either users and a mutual outage is declared.

7.5 Throughput Analysis

In this section, we derive the probability of each TM, and then the aTp by approximating the average BLER.

7.5.1 Probability of Transmission Modes

Each TM has lower and upper SNR thresholds for the instantaneous SNRs in order to be selected.

7.5.1.1 Grid search TMs

Each TM has lower and upper SNR thresholds for the instantaneous SNRs in order to be selected. The lower and upper γ_n thresholds for $\mathrm{TM}_{i,j}$ will be denoted as $\rho_{i,j}^{l,1} = \rho_{i,j}^{1}$, $\rho_{i,j}^{l,2} = \rho_{i,j}^{2}$, $\rho_{i,j}^{\mathrm{u},1} = \rho_{i+1,j}^{1}$ and $\rho_{i,j}^{\mathrm{u},2} = \rho_{i,j+1}^{2}$. Hence, the probability of $\mathrm{TM}_{i,j}$ can be calculated as follows,

$$\Pr\left(\mathrm{TM}_{i,j}\right) = \prod_{n=1}^{2} \Pr\left(\rho_{i,j}^{l,n} \leq \gamma_n < \rho_{i,j}^{\mathrm{u},n}\right) = \prod_{n=1}^{2} F_{\gamma_n}\left(\rho_{i,j}^{\mathrm{u},n}\right) - F_{\gamma_n}\left(\rho_{i,j}^{l,n}\right) \tag{7.18}$$

where $F_{\gamma_n}(\cdot)$ is the CDF of γ_n. Furthermore, the probability of transmission, i.e., $\Pr(\mathcal{M}_1, \mathcal{M}_2) > 0$, is $P_X = \sum_{i,j} \Pr\left(\mathrm{TM}_{i,j}\right)$. Moreover, the average number of transmitted bits for U_n is $S_n = \sum_{i,j} \Pr\left(\mathrm{TM}_{i,j}\right) \mathcal{M}_n^{*(i,j)}$.

7.5.1.2 Min-Max TMs

The probability of $\mathrm{TM}_{\mathcal{M}_t}$ can be calculated considering the conditions in (7.15)–(7.17) and it is expressed as follows,

$$\begin{aligned}\Pr\left(\mathrm{TM}_{\mathcal{M}_t}\right) = \prod_{n=1}^{2} F_{\gamma_n}\left(\rho_n^{*(\mathcal{M}_t)}\right) - F_{\gamma_n}\left(\rho_n^{*(\check{\mathcal{M}}_t)}\right) + \\ \left(F_{\gamma_1}\left(\rho_1^{*(\mathcal{M}_t)}\right) - F_{\gamma_1}\left(\rho_1^{*(\check{\mathcal{M}}_t)}\right)\right)\left(1 - F_{\gamma_2}\left(\rho_2^{*(\check{\mathcal{M}}_t)}\right)\right) + \\ \left(F_{\gamma_2}\left(\rho_2^{*(\mathcal{M}_t)}\right) - F_{\gamma_2}\left(\rho_2^{*(\check{\mathcal{M}}_t)}\right)\right)\left(1 - F_{\gamma_1}\left(\rho_1^{*(\check{\mathcal{M}}_t)}\right)\right).\end{aligned} \tag{7.19}$$

Moreover, the average number of transmitted bits for U_n is $S_n = \sum_{\mathcal{M}_t} \Pr\left(\mathrm{TM}_{\mathcal{M}_t}\right) \mathcal{M}_n^{*(\mathcal{M}_t)}$.

7.5.2 Average BLER

The average BLER of U_n can be found by considering the average BLER for all TMs. For instance, when considering the grid search, the average BLER can be expressed as

$$P_{e_n} = \frac{1}{P_X} \sum\nolimits_{i,j} \Pr\left(\mathrm{TM}_{i,j}\right) P_{e_n}^{(i,j)} \tag{7.20}$$

where

$$P_{e_n}^{(i,j)} = \frac{1}{\varpi_n} \int_{\rho_{i,j}^{l,n}}^{\rho_{i,j}^{\mathrm{u},n}} \left(1-(1-P_{B_n}|\gamma_n)^{b_n}\right) f_{\gamma_n}(\gamma_n)\, \mathrm{d}\gamma_n \tag{7.21}$$

$$= 1 - \sum_{r=0}^{b_n} \binom{b_n}{r} \frac{(-1)^r}{\varpi_n} \underbrace{\int_{\rho_{i,j}^{l,n}}^{\rho_{i,j}^{\mathrm{u},n}} (P_{B_n}|\gamma_n)^r f_{\gamma_n}(\gamma_n)\, \mathrm{d}\gamma_n}_{I_n} \tag{7.22}$$

and $b_n = \mathcal{M}_n^{*(i,j)} L$, $\binom{\cdot}{\cdot}$ is the binomial coefficient, $f_{\gamma_n}(\gamma_n)$ is the probability density function of γ_n. It is worth noting that the Binomial expansion theorem is used to simplify (7.21) such that $(1-z)^{b_n} = \sum_{r=0}^{b_n} \binom{b_n}{r} (-z)^r$. In addition, the integral, I_n, involves a multinomial expression since $P_{B_n}|\gamma_n$, (7.4), is a sum of Q-function terms, i.e. $Q(\cdot) + \cdots + Q(\cdot)$. Consequently, evaluating I_n in its current form leads to intractable analysis. Therefore, $P_{B_n}|\gamma_n$ is relaxed by considering a tight approximation that accounts for the most dominant $Q(\cdot)$ term only [180],

$$P_{B_n}|\gamma_n \approx \beta Q\left(\sqrt{\xi \gamma_n}\right) \tag{7.23}$$

where $\xi = \min_q \left|\Delta_q\right|^2$ and $\beta = \frac{1}{2^{\mathcal{M}-2\mathcal{M}_n}} \sum_q c_q$, $q = \arg\min_q \left|\Delta_q\right|^2$. Moreover, since the integral involves $Q^r(\cdot)$, its evaluation is still an open research problem for any integer. Hence, $Q(z) \approx \frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi(z^2+1)}}$ is used which is a tight approximation for a wide range of argument values [201]. Therefore,

$$I_n \approx \frac{\exp(\Omega)\, \Omega^{\frac{r-2}{2}} \beta^r}{\xi \bar{\gamma}_n (2\pi)^{r/2}} \left[\Gamma\left(\frac{2-r}{2}, \Omega(\xi \rho_{i,j}^{l,n} + 1)\right) - \Gamma\left(\frac{2-r}{2}, \Omega(\xi \rho_{i,j}^{\mathrm{u},n} + 1)\right)\right] \tag{7.24}$$

and $\Omega = \frac{r}{2} + \frac{1}{\xi \bar{\gamma}_n}$, the incomplete upper Gamma function is denoted by $\Gamma(a,x) = \int_x^\infty \exp(-t) t^{a-1}\, \mathrm{d}t$, while $\bar{\gamma}_n = \frac{2\mathbb{E}\left[|h_n|^2\right]}{N_0}$.

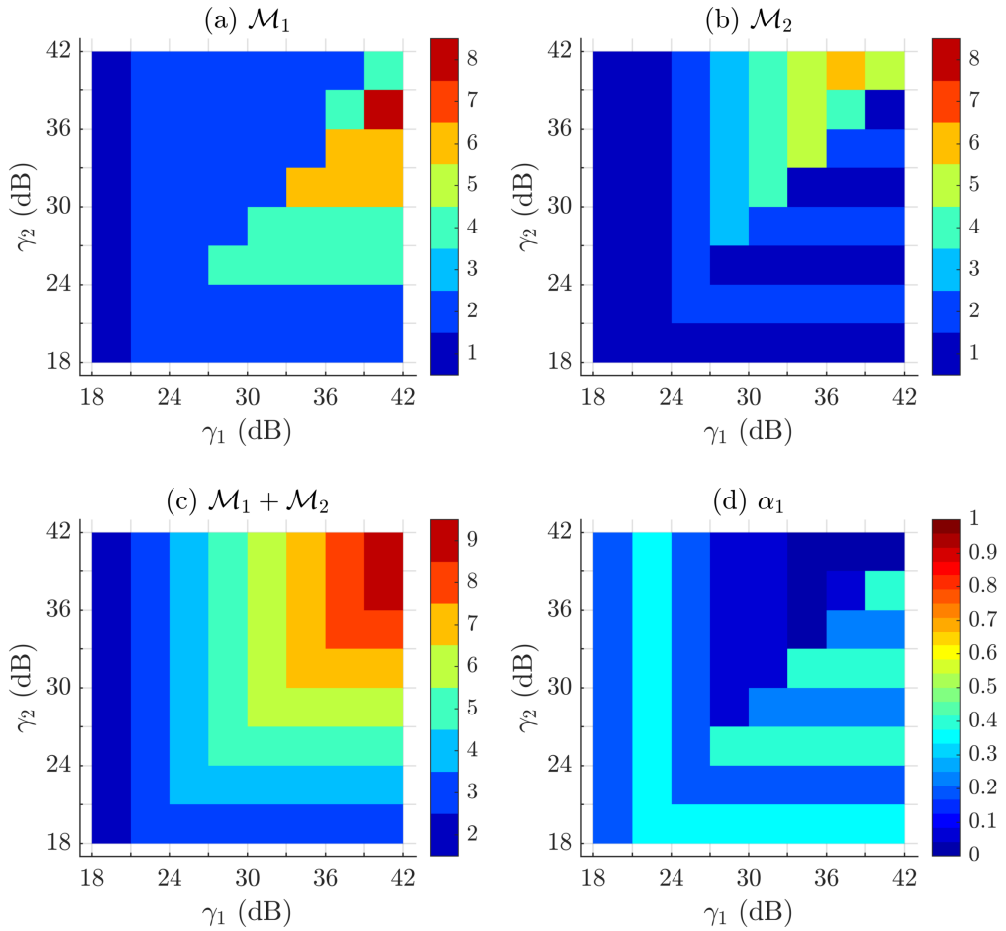


Figure 7.2: Transmission modes parameters for the FPA scheme. a) $\mathcal{M}_1$. b) $\mathcal{M}_2$. c) $\mathcal{M}_1 + \mathcal{M}_2$. d) α_1.

7.5.3 Throughput Expression

The aTp for U_n depends on the average BLER and the average number of transmitted bits. Hence, R_n is given as [202, Eq. (18)] $R_n = (1 - P_{e_n}) S_n$, while the aTp of the system is $\overline{R} = R_1 + R_2$. Furthermore, while $\tau_n \to 0$, $P_n|\gamma_n \approx 0$, R_n can be given as $R_n = (1 - P_{e_n}) S_n \approx S_n$.

7.6 Results and Discussions

This section presents the aTp analytical and numerical results for the grid search and min-max approaches, as well as the optimal TM parameters. It is assumed that U_1 and U_2 are at distances d_1 and $1.67d_1$ from the BS, and the pathloss exponent is $\lambda = 2.7$. Unless stated otherwise, the packet consists of $L = 32$ symbols and the BLER requirement is $\tau_n = 0.01$, $\forall n$, $\mathcal{M}_{\max} = 8$, while SNR $\triangleq 1/2\sigma_w^2$.

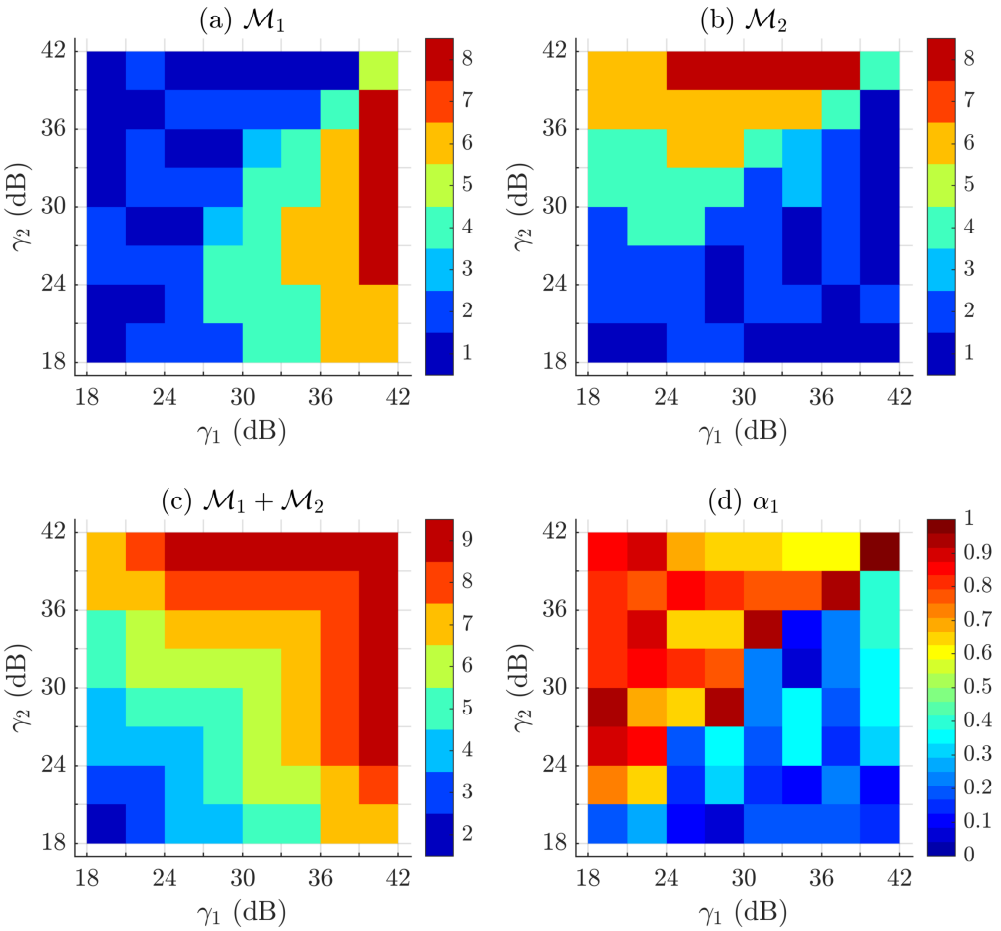


Figure 7.3: Transmission modes parameters for the APA scheme. a) $\mathcal{M}_1$. b) $\mathcal{M}_2$. c) $\mathcal{M}_1 + \mathcal{M}_2$. d) α_1.

7.6.1 Grid Search Validation

Figures 7.2 and 7.3 show the selected TMs parameters, $(\alpha_1, \mathcal{M}_1, \mathcal{M}_2)$ using (7.6)–(7.8). The x and y-axis represent γ_1 and γ_2, respectively. Color coding is used to represent $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_1 + \mathcal{M}_2$, and α_1. The figures are generated using $\Delta\gamma = 3$ dB. The selection of $\gamma_{\min}$ is based on the minimum γ_n requirements for both users to satisfy the QoS requirements with the minimum possible modulation orders which is $\gamma_{\min} = 18$ dB. Additionally, $\gamma_{\max} = 42$ dB, and $\mathcal{M}_{\max} = 8$. As can be noted from the figures, while $\mathcal{M}_n$ of an individual user using FPA may outperform an individual user with the APA in certain scenarios, $\sum_n \mathcal{M}_n$ for the APA consistently outperforms the FPA for $\gamma_1 > \gamma_2$ or $\gamma_1 < \gamma_2$. However, when $\gamma_1 = \gamma_2$ both schemes offer similar $\sum_n \mathcal{M}_n$ which, indicates that the FPA is optimal only for $\gamma_1 = \gamma_2$. Moreover, $\mathcal{M}_1$ and $\mathcal{M}_2$ for the APA in Figures 7.3a and 7.3b are symmetrical about the right diagonal, i.e., $\mathcal{M}_1|(\gamma_1, \gamma_2) = \mathcal{M}_2|(\gamma_2, \gamma_1)$. For example, $\mathcal{M}_1|(21\text{ dB}, 24\text{ dB}) = \mathcal{M}_2|(24\text{ dB}, 21\text{ dB}) = 2$.

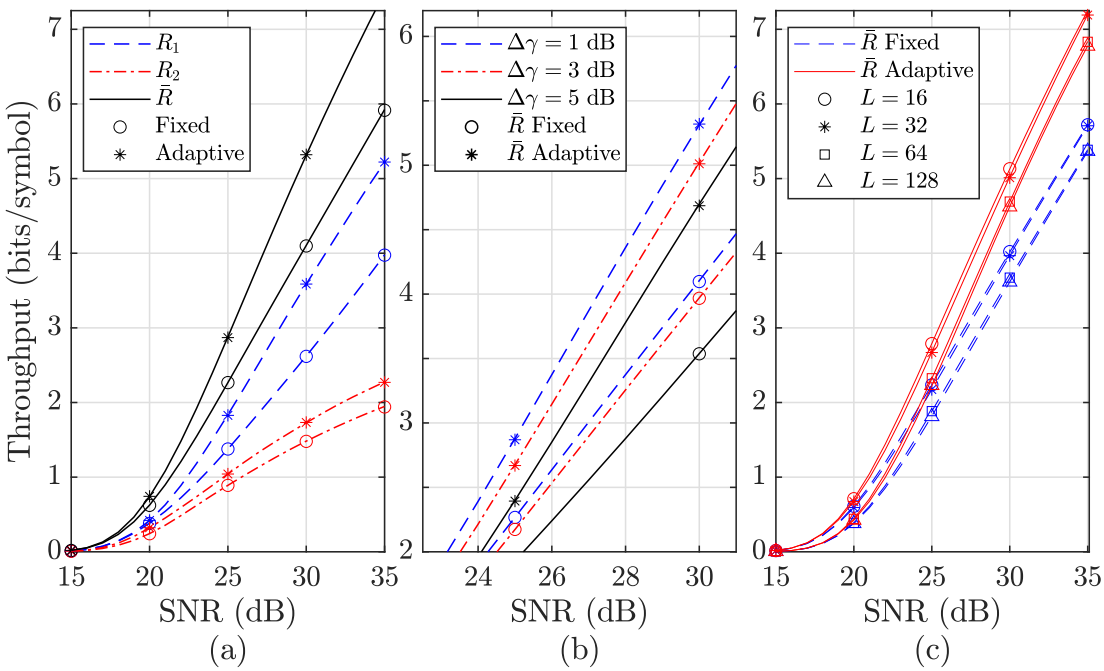


Figure 7.4: Analytical (lines) and simulated (markers) aTp for FPA and APA schemes. a) $L = 32$, $\Delta\gamma = 1$ dB. b) $L = 32$, $\Delta\gamma = \{1,3,5\}$ dB. c) $L = \{16,32,64,128\}$, $\Delta\gamma = 3$ dB.

Figure 7.4 shows that the analytical and simulation aTp results match closely for various parameters, which confirms the accuracy of the derived approximation. Fig 7.4a considers $\Delta\gamma = 1$ dB and $L = 32$, and it shows that the APA performance gain is proportional to SNR. For example, the APA outperforms the FPA by about 3 dB at 6 bits/symbol. Furthermore, R_1 is higher than R_2 because U_1 has better channel conditions. Figure 7.4b evaluates the effect of increasing $\Delta\gamma$ at a narrower SNR span, where the throughput degrades slightly when $\Delta\gamma$ increases from 1 to 3 dB. However, the degradation becomes significant for $\Delta\gamma = 5$ dB. Based on the observed trend, using $\Delta\gamma < 1$ will not have a tangible impact on the throughput, while the complexity will increase significantly.

Furthermore, Figure 7.4c analyzes the effect of changing L for $\Delta\gamma = 3$ dB. The figure shows that increasing L degrades aTp, which is due to the need for lower BER to achieve the same BLER. The proposed algorithm can be extended for $N > 2$, however, at the expense of additional complexity. Moreover, increasing the number of users beyond 2 degrades the BER, and thus, the BLER QoS requirements will be mostly violated causing a throughput degradation at low and moderate SNRs. For example, increasing the number of users from 2 to 3 causes a BER degradation of 10 dB at BER of 10^{-3} [24, Figure 14]. It is also worth noting that increasing L caused throughput degradation because the BLER is proportional to L in uncoded systems, and in coded

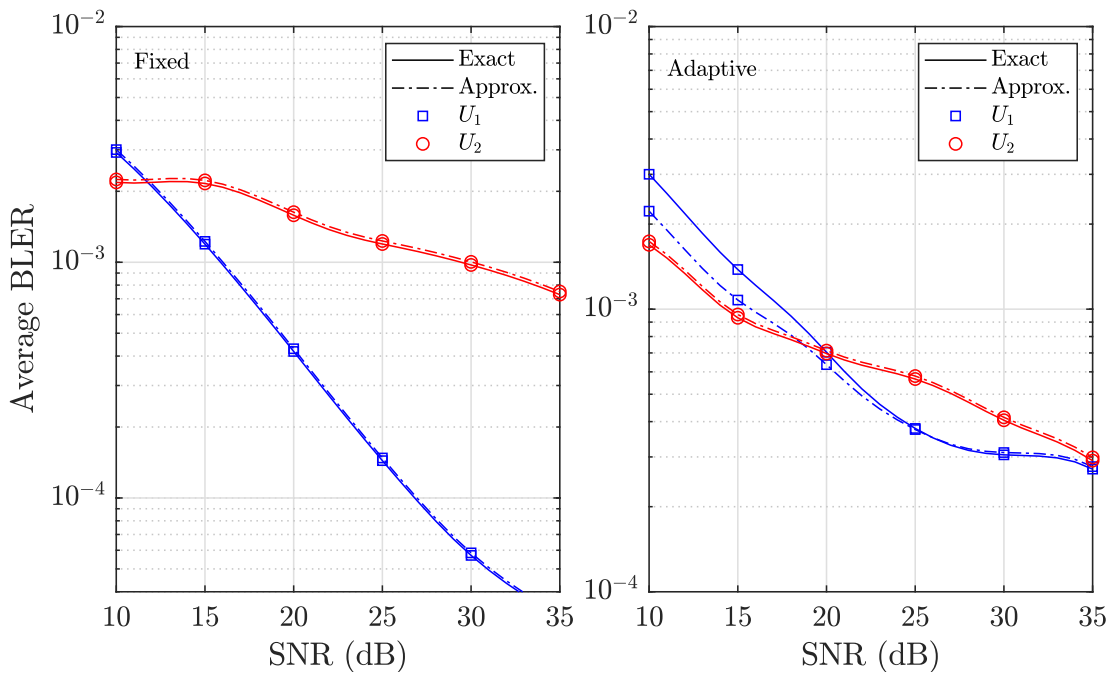


Figure 7.5: Exact and approximated average BLER for fixed and adaptive PA schemes, where $L = 32$ and $\Delta\gamma = 3$ dB.

systems when the packet length is increased while the codeword length is fixed, i.e., when multiple codewords are grouped to increase the packet length. If the packet length is equal to the codeword length, then increasing the packet length improves the BLER because the error correction capability in this case improves.

Figure 7.5 shows that the exact and approximated average BLER results exhibit a close match for wide range of SNRs,and the average BLER does not exceed 0.01, which is the threshold.

Figure 7.6 verifies the time complexity of the proposed stopping criteria at different pairs of γ_1 and γ_2. For the FPA, the stopping criteria reduce the complexity by 10 to 1400 fold, as compared to brute-force search. When the stopping criteria was adopted for the APA, the proposed segment-line search reduces the time complexity by about 30 to 116 fold when compared to full line search. Moreover, Figure 7.7 presents the convergence and time complexity for various pairs of γ_1 and γ_2. It is observed that APA requires more iterations to converge as compared to FPA. The reason is that the former searches in two power ranges, while the latter only checks the feasibility at a fixed power. In addition, it is noted that the time complexity increases exponentially with the number of iterations, which is due to the increase in the BLER expressions size at larger iteration numbers and higher modulation orders.

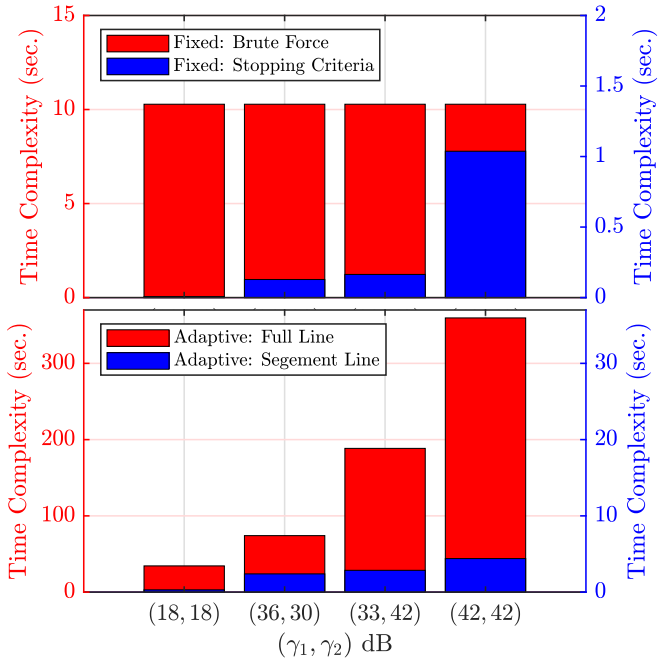


Figure 7.6: The time complexity of the proposed algorithm compared with different benchmarks at various instantaneous SNRs.

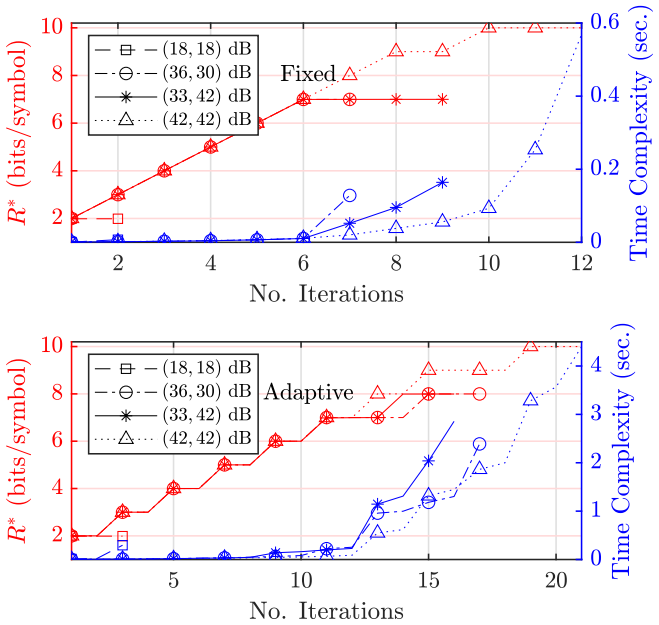


Figure 7.7: The convergence and time complexity of the proposed algorithm at various instantaneous SNRs.

7.6.2 Min-Max Approach Validation

Figures 7.8 and 7.9 show the selected TMs parameters, $(\alpha_1, \mathcal{M}_1, \mathcal{M}_2)$ for min-max and the grid search approaches. Similarly, the x and y-axis represent γ_1 and γ_2, respectively. Color coding is used to represent $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_1 + \mathcal{M}_2$, and α_1. Figure 7.9 is generated

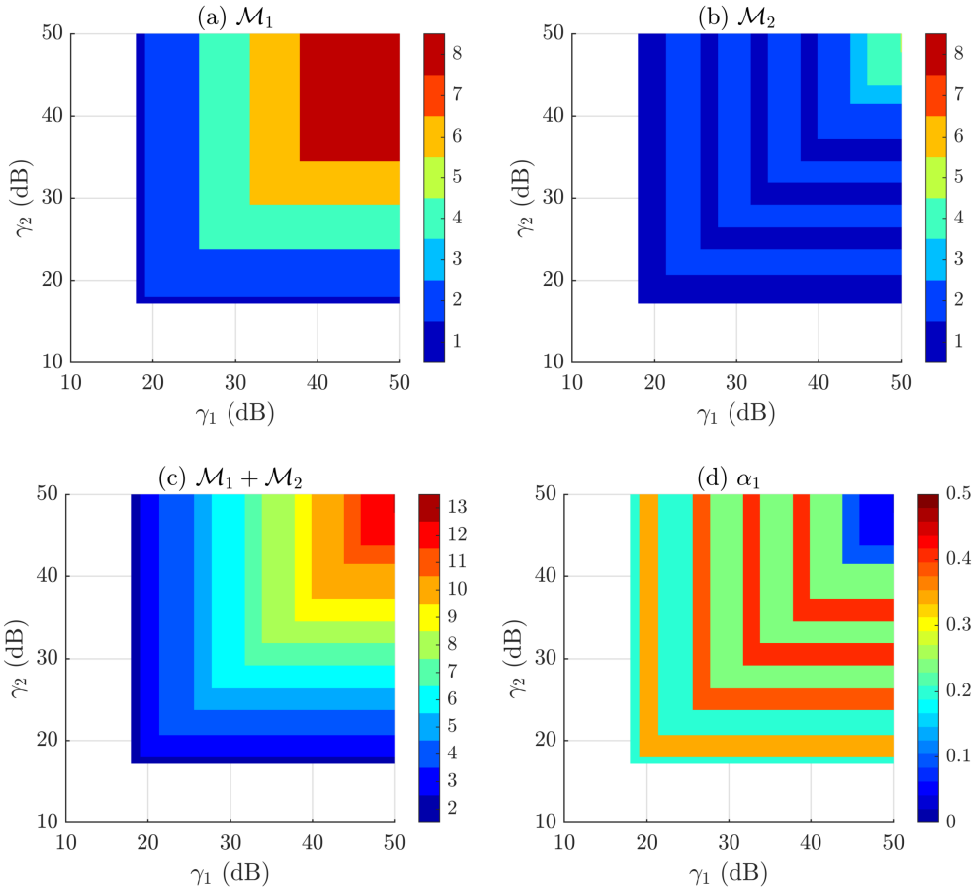


Figure 7.8: Transmission modes parameters for the min-max approach. a) $\mathcal{M}_1$. b) $\mathcal{M}_2$. c) $\mathcal{M}_1 + \mathcal{M}_2$. d) α_1.

using $\Delta\gamma = 3$ dB, $\gamma_{\text{min}} = 18$ dB, $\gamma_{\text{max}} = 54$ dB. Nonetheless, the region of interest is shown to be $10 < \gamma_n < 50$, $\forall n$. As can be seen in Figures 7.8a–b and 7.9a–b is that $\mathcal{M}_1$ ($\mathcal{M}_2$) is consistently higher (lower) for the min-max approach as U_1 tends to have higher SNR requirements compared to U_2, and hence, the case that satisfies the min-max criterion is when $\mathcal{M}_1 > \mathcal{M}_2$. In addition, it is noted that the quantisation results in sub-optimal solutions for $\mathcal{M}_1 + \mathcal{M}_2$. This can be seen by computing the sum weighted relative area for all $\mathcal{M}_1 + \mathcal{M}_2$, which is denoted as $\Theta = \sum_{\mathcal{M}_1+\mathcal{M}_2} \frac{A(\mathcal{M}_1+\mathcal{M}_2)}{A(\text{grid})} \times (\mathcal{M}_1 + \mathcal{M}_2)$, where $A(\cdot)$ is the area. It is found that $\Theta = \{3.73, 3.53, 3.34\}$ for the grid search with $\Delta\gamma = \{1, 3, 5\}$ dB, while $\Theta = 3.76$ for the min-max approach. Consequently, the relative improvement gain for the min-max approach with respect to grid search given $\Delta\gamma$ is denoted as $\Phi|\Delta\gamma$ and computed to be $\Phi|\Delta\gamma = \{0.8\%, 6.5\%, 12.6\%\}$ for $\Delta\gamma = \{1, 3, 5\}$ dB.

Figure 7.8 shows that the analytical and simulation aTp results match closely for various parameters, which confirms the accuracy of the derived approximation. It is

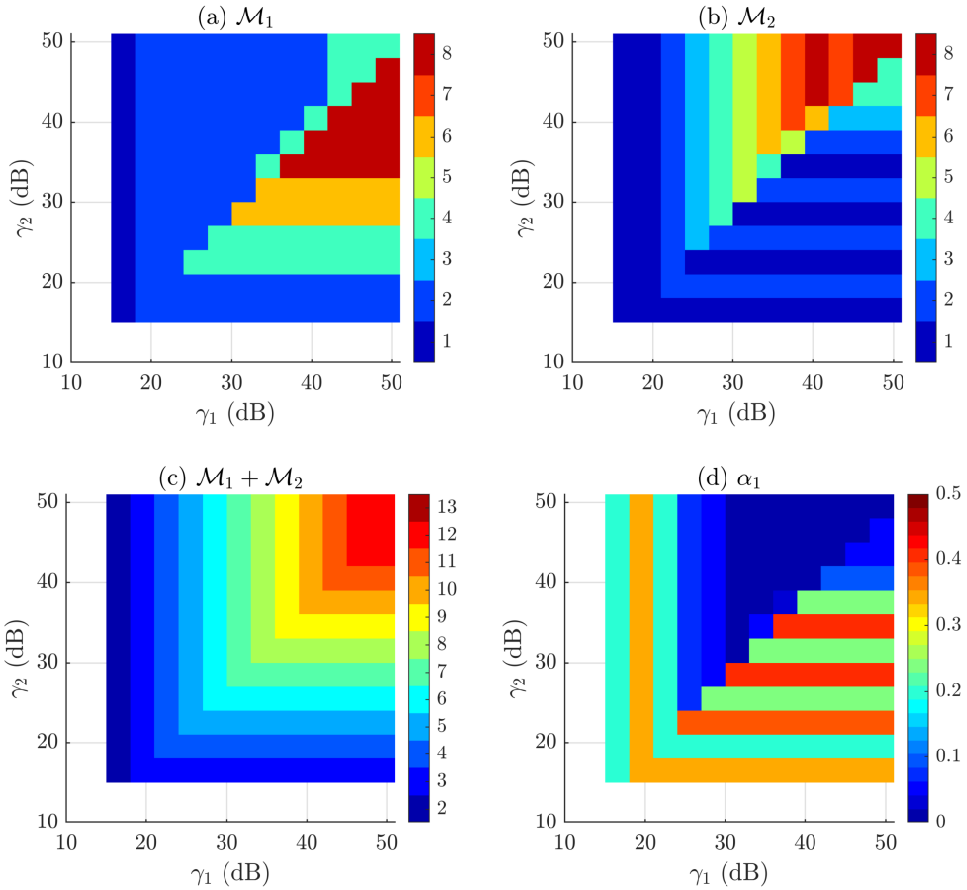


Figure 7.9: Transmission modes parameters for the min-max approach. a) $\mathcal{M}_1$. b) $\mathcal{M}_2$. c) $\mathcal{M}_1 + \mathcal{M}_2$. d) α_1.

noted min-max approach performance gain for $\bar{R}$ at 6 bits/symbol is 2.5 dB when compared to the grid search with $\Delta\gamma = 5$ dB. Furthermore, $\bar{R}$ for the grid search approach converges as $\Delta\gamma \to 0$ at the expense of increased complexity and number of TMs indicating that larger memory requirements are needed to store the LUT. Moreover, $\bar{R}$ asymptotically saturates at 13 bits/symbol and 12 bits/symbol for the min-max and grid search approaches.

7.7 Conclusions

This chapter derived closed-form expressions for the throughput over Rayleigh block fading channels for two-user NOMA systems using LUT-based adaptive modulation with FPA and APA. Two adaptation processes are considered at the BS which are the grid search and min-max approaches. Reduced complexity integer and mixed-integer optimisation problems were formulated and solved. The results show that the FPA

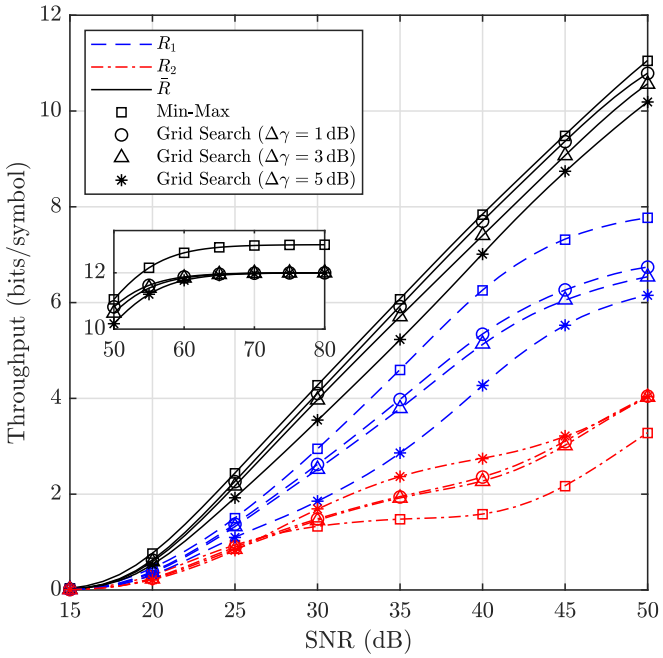


Figure 7.10: The analytical (lines) and simulated (markers) aTp results vs. SNR for the min-max approach and grid search approach [1] .

scheme provides similar performance to the APA scheme when both users experience identical channel conditions, and that APA outperforms FPA in other channel conditions. Moreover, the system throughput demonstrated high tolerance to the granularity of the quantisation levels. Furthermore, the results show the min-max approach can improve the throughput by up to 2.5 dB at moderate SNRs.

Chapter 8

Generalized NOM for ARQ Systems

8.1 Chapter Introduction

Non-orthogonal multiplexing (NOM) is a novel superposition coding scheme that has been recently proposed to improve the throughput of wireless systems. However, restricting the number of multiplexed packets to two limits the throughput improvement of NOM to 100% in best case scenarios. Therefore, this work presents a generalized non-orthogonal multiplexing (GNOM) design with arbitrary number of multiplexed packets. In the multiplexing process, new and retransmitted packets due to automatic repeat request (ARQ) are combined while considering the impact of channel conditions on the power assigned per packet. The proposed GNOM employs an efficient heuristic algorithm to perform the power assignment and multiplexing decisions. Moreover, the complexity can be controlled by enforcing a limit on the maximum number of multiplexed packets per transmission, making it suitable for Internet of Things (IoT) nodes with diverse computational capabilities and quality of service requirements. The obtained results demonstrate the effectiveness of proposed scheme, which offers up to 200% throughput improvement at moderate signal to noise ratios (SNRs), and up to 700% at high SNRs. Furthermore, the new scheme can reduce the transmission power consumption by up to 6 dB in the high SNR region.

8.1.1 Chapter Organisation

The rest of the paper is organised as follows. Related work and motivations and contributions are mentioned in Section 8.2. The conventional ARQ system and channel models are described in Section 8.3. The proposed GNOM system with adaptive power

allocation is explained in Section 8.4. The proposed GNOM transmission protocol is presented in Section 8.5. Simulation and numerical results are discussed in Section 8.6. The conclusion is drawn in Section 8.7.

8.2 Related Work

The integration of IoT in various applications has been growing vastly in the last decade, which triggered the introduction of more specific applications such as Internet of Drones (IoD) [203–206] and Internet of Vehicles (IoV) [207–211], or even more advanced configurations that include both IoD and IoV. Human and machine-centric IoT applications include transportation, smart cities, health care, agriculture, environment, retail, and smart homes. These applications tend to have a broad range of quality of service (QoS) requirements in terms of latency, power and data rates, and can be supported using various networks infrastructure including satellite-terrestrial networks [212,213]. For example, the Sigfox, LoRaWAN, narrowband (NB)-IoT and Long Term Evolution (LTE)-machine-type communication (MTC) are optimised to provide bit rates of 100 bps, 50 kbps, 250 kbps and 1 Mbps, respectively [214]. Therefore, wireless protocols that will be used to serve IoT applications should have sufficient flexibility to accommodate the diverse data rate requirements. Moreover, they should consider the power and computational capability constraints of the IoT nodes.

For certain applications, ultra-reliable low latency communication (uRLLC) transmission is crucial because erroneous or delayed information can cause fatal consequences. For example, vehicles in IoV environment should perform vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) connectivity to enable vehicles to make certain critical decisions [99]. IoD applications have similar requirements because the drones typically have to collect and relay location-specific critical information [203–206]. Therefore, the received data should be verified for correctness before being used to make decisions. In the context of the decision making processes, decision fusion has been recognized as one of the most efficient approaches to improve the reliability of the decisions made about a certain phenomenon [103, 215]. Nevertheless, having redundant information from multiple sources to perform fusion is infeasible in several applications.

In addition to IoT applications, having reliable data is required for most current applications that rely on wireless communications. The International Telecommunication Union (ITU) has specified that the packet error rate (PER) for any class

of service should not be greater than 10^{-4} [216, Table I]. In this context, ARQ is considered the prominent technique that can be used to achieve this goal. Consequently, ARQ is adopted for the fifth generation (5G) [217] and for several other wireless communications standards such as LTE, wireless fidelity (WiFi), wireless personal area networks (WPANs) and worldwide interoperability for microwave access (WiMAX) [216]. Moreover, ARQ has been considered for several technologies such as non-orthogonal multiple access (NOMA) [218–220], multiple-input multiple-output (MIMO) [221], massive MIMO (mMIMO) [222] and cognitive radio (CR) [223].

8.2.1 ARQ Overview

ARQ allows nearly error-free transmission by checking all received packets for errors and instructing the transmitter to retransmit the packets that fail the error check. Packets that never pass the error check will be eventually dropped [216]. Despite its advantages, ARQ may cause severe throughput and power degradation, particularly at low and moderate SNRs [224] due to the repeatedly transmitted data and the need for frequent feedback messages from the receiver to the transmitter. Consequently, extensive research in the literature is devoted to mitigating the adverse effects of ARQ. For example, integrating forward error correcting (FEC) and ARQ, denoted as hybrid automatic repeat request (HARQ), can reduce the number of retransmissions by reducing the PER [225]. Moreover, HARQ provides more flexibility to control the type and amount of retransmitted data, which can be optimised to improve the system efficiency [216]. However, FEC may also reduce the system throughput and power due to the redundant parity bits, and the additional encoding and decoding complexity and delay. The packet length is a critical parameter that affects the PER, and thus, optimising the packet length can significantly improve the system throughput [226].

Utilising time diversity by combining the multiple retransmitted packets, known as chase combining (ChC), can significantly improve ARQ system efficiency. In ARQ-ChC systems, maximum ratio combining (MRC) can be used to combine all the transmissions that correspond to a particular packet to reduce the PER and improve the throughput. ChC also enables a remarkable power saving because the retransmitted packets usually do not require the same power used in the initial packet transmission [224].

8.2.2 ARQ-based NOMA/NOM

The fact that full power is not required for retransmitted data enables multiplexing new and retransmitted bits in the power domain [2,227,228]. A comprehensive survey on power-domain NOMA is reported in [229], while the integration of power-domain NOMA and the enabling communication techniques and technologies are discussed in [81]. In [227], the focus is to reduce the delay using an early retransmission technique. The paper also considers combining new and retransmitted packets using superposition coding where new packets are allocated fixed high power as compared to retransmitted packets. In each transmission session, a maximum of one new and one retransmitted packets can be combined and transmitted simultaneously. However, the fixed power assignment and false negative acknowledgments (NACKs) caused by early retransmission decisions eliminate the throughput gain that could have been achieved using the superposition coding. Khreis *et al.* [228] proposed a multi-layer HARQ that utilises the idle time that the transmitter usually experiences while waiting for a feedback message (F_M). The main concept is to retransmit certain packets multiple times without waiting for a F_M. To avoid throughput losses, the redundant data is combined with new transmission using superposition coding. The results presented in [228] show that the throughput improvement is about 10%. NOM is introduced in [2] to utilise the power and spectrum efficiently where a low complexity heuristic algorithm is proposed to combine new and retransmitted data. Although the proposed algorithm provides up to 100% throughput improvement at high SNRs, the work considers fixed power assignment and the combining is limited only to two packets. Thus, the throughput improvement is limited because the power allocation can significantly affect the system performance in both NOMA [230] and NOM [2]. In [231], a power-domain uplink (UL) NOMA scheme for NB-IoT is presented to handle the challenge of serving massive devices with available resources. Multiple massive machine-type communication (mMTC) devices can share a single sub-carrier in the proposed system, allowing for increased device admission in the overall system. Khan *et al.* [232] propose an improvement in the spectral efficiency of IoT networks under power-domain NOMA. While taking practical constraints into account, the proposed framework takes the limited number of frequency blocks in the IoT network into account and provides an optimal technique for power and frequency block allocation.

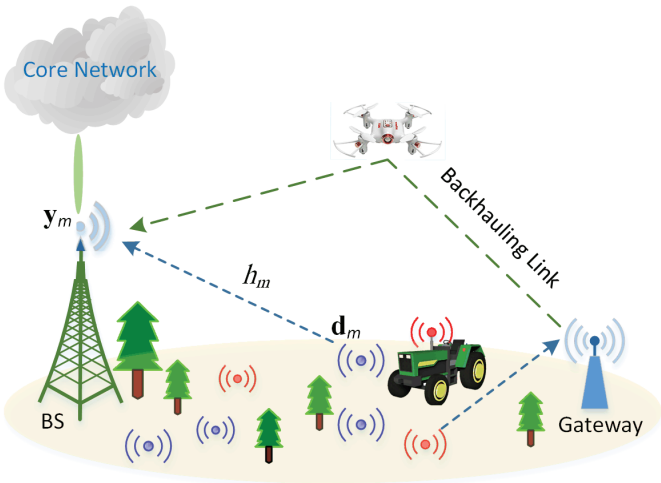


Figure 8.1: Simplified system architecture with application to precision agriculture.

8.2.3 Motivation and Contributions

As can be noted from the aforementioned discussions and the cited literature, IoT applications are increasing the demand for substantial additional transmission resources due to the rising traffic volume and emerging high data rate requirements. An example of precision agriculture using IoT is depicted in Figure 8.1 where several sensors need to communicate data about the field, which may require low data rates. However, a remotely operated tractor needs to communicate large data volumes at high data rates. Overall, aggregating the data generated by a large number of sensors of various types produces large data volumes that have to be transmitted with constraints on reliability, delay, complexity, power and spectrum. Therefore, conventional solutions will mostly fail to simultaneously satisfy such conflicting requirements.

Although NOM has the potential to overcome the limitations inherent in ARQ, its throughput improvement is limited to only two packets, and the packets are multiplexed with fixed power. Therefore, this article:

1. Proposes a GNOM for ARQ-based wireless systems to improve their power and spectral efficiencies, and as a consequence, the throughput will be improved as well.

2. In the proposed GNOM, the transmitter computes the power coefficients of the multiplexed GNOM packets to stack the maximum possible number of packets in a single transmission while satisfying a predetermined PER constraint.

3. The impact of changing the maximum number of multiplexed packets to a specific value N_{cap} packets is considered to provide the system designer with an

additional degree of freedom to optimise the spectral efficiency and complexity. As a consequence, the system design can be tailored to suit several IoT applications based on their QoS requirements and computational capabilities.

4. The GNOM system complexity is evaluated and compared to a NOM based system.

The obtained numerical results show that a significant throughput gain can be achieved by GNOM when compared to conventional ARQ and NOM systems. The power efficiency is also evaluated in terms of average power per packet.

8.3 ARQ System and Channel Models

In this work, multiplexing data packets that belong to a single user in the power domain is denoted as NOM, while if the data packets belong to multiple users the system is denoted as NOMA.

The ARQ system considered in this work consists of an IoT node that needs to send M packets over a wireless channel. Therefore, the system configuration corresponds to the uplink transmission in IoT networks as shown in Figure 8.1. The IoT nodes can be connected to an IoT gateway, or any type of base station (BS). The transmitter and receiver utilise truncated ARQ with stop-and-wait (SW) flow control protocol where the maximum number of allowed transmissions is $\mathcal{M}$, which includes the initial transmission. A packet that is transmitted $\mathcal{M}$ times and still has errors will be dropped. Without loss of generality, it is assumed that binary phase-shift keying (BPSK) modulation is adopted at the transmitter to ensure reliable transmission. BPSK modulation is adopted in various standards such as the NB-IoT standard [102]. Nevertheless, it is straightforward to extend the work to other modulation schemes. The channel between the transmitter and receiver is considered as a block fading where the channel remains fixed for the period of one packet but changes randomly over consecutive packets. Therefore, the received packet that corresponds to the kth transmitted packet can be written as

$$\mathbf{y}_{k,l}^{(u)} = h_{k,l}^{(u)} \sqrt{\alpha_{k,l}^{(u)}}\, \mathbf{x}_{k,l}^{(u)} + \mathbf{w}_{k,l}^{(u)} \tag{8.1}$$

where l is the transmission slot, u is the transmission/retransmission counter for each packet, α is the transmit power which is normalized to unity, $h \sim \mathcal{CN}(0, \sigma_h^2)$ is the channel frequency response, $\mathbf{w}$ is the additive white Gaussian noise (AWGN) vector,

$\mathbf{w} = [w_1, w_2, \ldots, w_L]$, $w_i \sim \mathcal{CN}(0, N_0)$, and $\mathbf{x} = [x_1, x_2, \ldots, x_L]$ is the transmitted data packet, $x_i \in \{-1, 1\}$ $\forall i$, and L is the packet length in symbols.

At the receiver, the received symbols in each packet are demodulated, and then the packet undergoes an error detection process, which is typically considered to be perfect, i.e., false alarm and misdetection probabilities are equal to zero. After error detection, an F_M is sent back to the transmitter to instruct it to either retransmit the previously transmitted packet when the F_M is a NACK, or send a new packet when F_M is an acknowledgment (ACK). In this work, we consider Type-I ARQ, and thus, the same packet can be transmitted up to $\mathcal{M}$ times, $u \in \{1, 2, \ldots, \mathcal{M}\}$. The receiver can exploit the channel temporal variations and combine the multiple received versions of the same packet using ChC. Therefore, the PER for the kth packet after the uth transmission can be written as

$$\bar{P}_{k,l}^{(u)} = 1 - (1 - P_{k,l}^{(u)})^L \tag{8.2}$$

where

$$P_{k,l}^{(u)} = Q\left(\sqrt{\sum_{i=1}^{u} \alpha_{k,\ell+i-1}^{(i)} \gamma_{k,\ell+i-1}^{(i)}}\right) \tag{8.3}$$

and $\gamma_{k,\ell+i-1}^{(i)} = \frac{2\left|h_{k,\ell+i-1}^{(i)}\right|^2}{\sigma_w^2}$, $\sigma_w^2 = N_0/2$, ℓ corresponds to the slot in which packet k was transmitted for the first time such that $l = \ell + u - 1$.

8.4 GNOM-ARQ System

In conventional ARQ, the power per packet can be considered fixed, i.e., $\alpha_{k,l}^{(u)} = \alpha_{k,\ell}$ $\forall\{k, u, l\}$. Alternatively, $\alpha_{k,l}^{(u)}$ can be optimized to save power [224], which is necessary in the case that a packet is transmitted more than once. In such scenarios, the transmitter may frequently experience cases with $\alpha_{k,l}^{(u)} \leq p_{\max}$, where $p_{\max}$ is the transmitter maximum transmit power. Consequently, the power amplifier efficiency at the transmitter could be significantly harmed. Therefore, we propose in this work to opportunistically multiplex multiple packets at the transmitter in the power domain such that the total transmit power is reasonably large. In addition to resolving the power amplifier efficiency problem, a significant spectral efficiency improvement can be gained because multiple packets can be transmitted using the same transmission time/frequency resources that are used in conventional ARQ systems. As a consequence of the spectral

efficiency improvement, the delay can be reduced since it is inversely proportional to the transmission rate. Moreover, the average power per packet will be reduced because several packets with small power values will be successfully received.

8.4.1 GNOM

Similar to conventional ARQ, consider that the transmitting IoT node will transmit M packets over a wireless channel to a particular receiver. We assume that the channel frequency response of that transmission session is known at the transmitter, which is a widely adopted assumption in stationary, low, and moderate mobility systems. If the channel conditions allow multiplexing more than one packet into a particular transmission session, then the transmitted GNOM packets during the first transmission slot can be written as

$$\mathbf{d}_1 = \sum_{i=1}^{N_1} \sqrt{\alpha_{i,1}^{(1)}} \mathbf{x}_{i,1}^{(1)} \tag{8.4}$$

where N_1 is the number of packets that are multiplexed in the current transmission slot $N_1 \leq N_{\text{cap}}$, N_{cap} is the maximum number of GNOM packets that can be multiplexed, $\alpha_{i,1}^{(1)}$ is the power coefficient of the ith GNOM packet, the total power in the transmission slot is normalized to unity, and thus, $\sum_{i=1}^{N_1} \alpha_{i,1}^{(1)} = 1$. It is worth noting that the case where $N_{\text{cap}} = 1$ corresponds to the conventional ARQ system, while $N_{\text{cap}} = 2$ represents the NOM [2].

Although the power per packet in (8.4) can be assigned in an arbitrary manner, the selected powers will affect the receiver design in terms of computational complexity, buffering requirements and packet delay. Therefore, to enable using the low complexity successive interference cancellation (SIC) receiver, minimize the buffering requirements, and avoid jitter, the power selection should be performed following the general rules used for NOMA. To achieve these goals, the system design should adhere to the following general design rules:

1. The packets' sequence numbers are used as a priority indicators, that is, a packet with the lowest index has the highest priority, and so forth. Packets with higher priority should be assigned more power than packets with lower priority, $\alpha_{1,l}^{(u)} > \alpha_{2,l}^{(u)} > \cdots > \alpha_{N_l,l}^{(u)}$. More specific rules should be specified based on the adopted modulation scheme and the number of multiplexed packets [27].

2. Because the packet with the highest priority will be allocated the highest power, it will be highly likely that lower priority packets will not pass the error check if

the primary packet fails the error check. Consequently, all multiplexed packets will be simultaneously dropped after $\mathcal{M}$ unsuccessful transmissions. To prevent such scenarios, the packet with the minimum index, denoted as the primary packet, is the only packet whose transmission counter is incremented when it is received unsuccessfully. Consequently, only the primary packet will be dropped after $\mathcal{M}$ unsuccessful transmissions. For example, given that the primary packet $\mathbf{x}_1$ has failed $m-1$ times, then the composite packet in the mth transmission time slot can be written as

$$\mathbf{d}_m = \sqrt{\alpha_{1,m}^{(m)}}\mathbf{x}_{1,m}^{(m)} + \sum_{i=2}^{N_m} \sqrt{\alpha_{i,m}^{(1)}}\mathbf{x}_{i,m}^{(1)}. \tag{8.5}$$

As can be noted from (8.5), the transmission counter u is incremented only for $\mathbf{x}_1$.

At the receiver, the received composite packet that corresponds to $\mathbf{d}_m$ can be expressed as

$$\mathbf{y}_m = h_m \left(\sqrt{\alpha_{1,m}^{(m)}}\mathbf{x}_{1,m}^{(m)} + \sum_{i=2}^{N_m} \sqrt{\alpha_{i,m}^{(1)}}\mathbf{x}_{i,m}^{(1)} \right) + \mathbf{w}_m. \tag{8.6}$$

For notational convenience, the channel frequency response h for the composite packet $\mathbf{d}_m$ will be written as $h_{k,l}^{(u)}$. Therefore,

$$\mathbf{y}_m = h_{1,m}^{(m)}\sqrt{\alpha_{1,m}^{(m)}}\mathbf{x}_{1,m}^{(m)} + \sum_{i=2}^{N_m} h_{i,m}^{(1)}\sqrt{\alpha_{i,m}^{(1)}}\mathbf{x}_{i,m}^{(1)} + \mathbf{w}_m \tag{8.7}$$

where $h_{k,l}^{(u)} = h_l \; \forall\{k,u\}$. Moreover, the transmission time slot index, l, will be dropped unless it is necessary to include it. Following the SIC detection principle, the primary packet is detected first, and then the secondary packets are detected according to their index values. However, if the primary packet is not detected successfully, then the interference cancellation will fail and most likely all secondary packets will not be detected successfully as well. Consequently, after the detection stage, the primary packet will be tested for errors, and if it fails, the detection process for all secondary packets will be aborted, and these packets will be considered erroneous. The same procedure is applied to the secondary packets, i.e., if $\mathbf{x}_i$ fails the error check, then $\mathbf{x}_l \; \forall l > i$ will be automatically considered erroneous. Consequently, F_M will have the form $\mathcal{A} = [1,1,\ldots,1,0,0,\ldots,0]$ where 1 stands for ACK and 0 stands for NACK.

Algorithm 8.1: Power allocation.

Input: l, k, $\gamma_{k,l}^{(1)}$, τ, L, N_{cap}
Output: N_l, α_l

```
α_l = [1,0,0,...,0] ∈ B^{N_cap × 1}, N_l = 1
Compute the PER P̄^(1)_{1,l} using (8.2) and (8.3)
if P̄^(1)_{1,l} < τ then
    for N_l = 2 : N_cap do
        Find α_l by solving (8.8a)–(8.8e)
        Compute the PER P̄^(1)_{N_l,l} using (8.2) and (8.9)
        if P̄^(1)_{N_l,l} ≥ τ then
            break
```

8.4.2 Power Allocation

The number of multiplexed GNOM packets per transmission slot mostly depends on the channel frequency response and noise power. Before the commencement of the packet transmission process, a power allocation routine is performed to allocate the available power to maximize the possible number of multiplexed packets N_l during each transmission slot to maximize the spectral efficiency while satisfying the PER constraint.

To reduce the complexity and the signalling overhead, the routine is called only when the counter of the primary packet $u = 1$, and all transmission parameters are kept fixed during any retransmission process, if there is any. The maximum number of packets that can be multiplexed in a particular transmission session is obtained using Algorithm 8.1, which can be described as follows:

1. Compute the PER for a single packet using (8.2) and (8.3), and compare it to the PER threshold τ. If $\bar{P}_{1,l}^{(1)} < \tau$, go to next step, otherwise $N_l = 1$ and thus, only one packet will be transmitted.

2. Compute the power coefficients to transmit a GNOM packet. The process starts with two multiplexed packets and may continue up to N_{cap}. The optimization

problem for a given N_l can be formulated as follows

$$\arg \min_{\alpha_l \in \mathbb{R}^{N_l \times 1}} P_{N_l,l}^{(1)} \tag{8.8a}$$

subject to:

$$\bar{P}_{j,l} \leq \tau, \ \forall j \in \{1, \cdots, N_l - 1\} \tag{8.8b}$$

$$\sqrt{\alpha_{j,l}^{(1)}} > \sum_{r=j+1}^{N_l} \sqrt{\alpha_{r,l}^{(1)}}, \ \forall j \in \{1, \cdots, N_l - 1\} \tag{8.8c}$$

$$\sqrt{\alpha_{N_l,l}^{(1)}} > 0 \tag{8.8d}$$

$$\sum_{r=1}^{N_l} \sqrt{\alpha_{r,l}^{(1)}} = 1 \tag{8.8e}$$

where $\mathbb{R}$ corresponds to the set of real numbers between 0 and 1, the objective in (8.8a) is to find power coefficient α_l for which the bit error rate (BER) of the last secondary packet, $\mathbf{x}_{N_l,l}$, is minimized while satisfying the constraints (8.8b)–(8.8e). The constraint in (8.8b) is non-linear, and it is used to satisfy the PER threshold for the primary and all secondary packets except $\mathbf{x}_{N_l,l}$ which will be transmitted regardless of the PER constraint. Constraints (8.8c)–(8.8d) are required to satisfy power coefficient bounds [27, Eq. (21)], (8.8e) ensures a unity sum of power coefficients during a transmission slot. By noting that (8.8a) and (8.8b) are highly non-linear and non-convex at the same time, search methods such as gradient descent and interior point algorithm can provide near-optimum solutions [24]. The BER for GNOM can be written as [24, Eq. (23)]

$$P_{k,l}^{(u)} = \sum_j \frac{1}{c_j} Q\left(|\Delta_j| \sqrt{\sum_{i=1}^{u} \gamma_{k,\ell+i-1}^{(i)}} \right) \tag{8.9}$$

where ℓ corresponds to the slot in which packet k was transmitted for the first time such that $l = \ell + u - 1$, c_j and Δ_j values can be found in Appendix D. Similarly, the PER can be calculated according to (8.2).

3. Once α_l is calculated for $N_l = 2$, a feasibility check is done where the PER constraint satisfaction is checked for the secondary packet. If it is strictly satisfied or not satisfied, then $N_l = 2$ and the computed α_l are returned. Otherwise, α_l for $N_l = 3$ will be computed. This loop will continue until $N_l = N_{\text{cap}}$ unless the PER constraint of the N_l^{th} packet is strictly satisfied or not satisfied.

It is worth mentioning that if a primary packet is not successfully detected, the N_l superimposed packets are buffered, and a NACK is sent to the transmitter. The transmitter retransmits the same data packets using the same power allocation coefficients that were used in the first transmission of these packets. The reason for using the same packets and power allocation is to simplify the MRC process. If a new power allocation process is performed, the retransmitted packets may include new packets that were not transmitted in the first transmission, and with different power coefficients. Consequently, the combining process becomes infeasible [218]. Upon receiving the retransmitted packets, the receiver applies MRC to combine the retransmitted and buffered packets. If the primary packet transmission counter exceeds $\mathcal{M}$ without being successfully detected, then the composite packet and all its buffered versions are dropped and the receiving buffer is cleared. Otherwise, the primary packet can be subtracted from the combined packet to allow combining and detecting the secondary packets. It is worth noting that when the primary packet detection fails, the interference to all other superimposed packets will be very high, and mostly none of them can be detected successfully. Therefore, the detection process is aborted to save time and computational power.

8.5 GNOM Transmission Protocol

As can be noted from the power assignment process, the transmitter may send one packet, or up to N_{cap} multiplexed packets in each transmission slot. According to the SW protocol, the transmitter will be expecting an F_M for each transmitted packet. The F_M will contain the feedback for all multiplexed packets in a vector $\mathcal{A}$, which can be $\mathcal{A} = [\mathcal{A}_1]$, $\mathcal{A} = [\mathcal{A}_1, \mathcal{A}_2]$, or $\mathcal{A} = [\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_{N_{\text{cap}}}]$, $\mathcal{A}_i \in \{0,1\}$ $\forall i$, where 1 represents an ACK and 0 is for a NACK. However, in the case that a packet is not detected correctly, it will cause severe interference to the low-power packets, and thus, most likely they will not be detected correctly as well. Therefore, to save time and computational complexity, if packet $\mathbf{x}_i$ fails, then the detection of $\mathbf{x}_{i+1}$ or $\{\mathbf{x}_{i+1}, \mathbf{x}_{i+2}\}$ is aborted and all lower power packets will be considered incorrect. Consequently, the F_M for the $N_{\text{cap}} = 3$ case, can only be $\mathcal{A} = [0,0,0]$, $\mathcal{A} = [1,0,0]$, $\mathcal{A} = [1,1,0]$, or $\mathcal{A} = [1,1,1]$. In the case that $\mathcal{A} = [0,0,0]$, the transmitter considers that all three packets have failed. Therefore, the packet with index i remains the primary packet during the following transmission, but the transmission index l is increased by one. If the primary packet is detected correctly during its initial transmission but the secondary

packets are unable to pass the error check, the acknowledgement $\mathcal{A} = [1,0,0]$ is sent back to the transmitter. In the next transmission slot, the primary packet becomes the packet with index $i+1$. In addition, the buffer containing the observations and channel realization of the previous transmission slot will be cleared. Moreover, the primary packet's transmission counter u is reset to 1 for the new primary packet. It is worth noting that the primary packet will be dropped if the primary packet transmission counter exceeds the maximum number of allowed transmissions $\mathcal{M}$.

8.5.1 GNOM Receiver for Fixed N_l

To simplify the GNOM receiver description, we consider an example where the power assignment algorithm during the first transmission slot, $l = 1$, returns $N_1 = 3$ and $\alpha_1 = [\alpha_1, \alpha_2, \alpha_3]$. Hence, the system transmits three packets during this transmission slot. The modulated packets, simply termed as packets, are denoted as $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, where packet $\mathbf{x}_1$ is the primary packet. The receiver employs SIC to sequentially cancel the inter-packet interference. Hence, the individual packets can be detected from the received composite packet. For the considered example, the received composite packet is given as

$$\mathbf{y}_1 = h_1 \left(\sqrt{\alpha_{1,1}^{(1)}} \mathbf{x}_{1,1}^{(1)} + \sqrt{\alpha_{2,1}^{(1)}} \mathbf{x}_{2,1}^{(1)} + \sqrt{\alpha_{3,1}^{(1)}} \mathbf{x}_{3,1}^{(1)} \right) + \mathbf{w}_1. \tag{8.10}$$

The receiver employs a maximum likelihood detector (MLD) to detect the primary packet $\mathbf{x}_1$ while considering the other secondary packets as unknown additive noise. The detected packet $\hat{\mathbf{x}}_1$ is applied to a cyclic redundancy check (CRC) process for verification, denoted by $\mathcal{C}(\hat{\mathbf{x}}_1)$, where $\mathcal{C}(\hat{\mathbf{x}}_1) = 1$ indicates that the packet has passed the check, and 0 otherwise. Given that $\mathcal{C}(\hat{\mathbf{x}}_1) = 1$, then SIC is applied to subtract the interference caused by $\mathbf{x}_1$, then MLD is used to detect $\mathbf{x}_2$. If $\mathcal{C}(\hat{\mathbf{x}}_2) = 1$, then $\mathbf{x}_3$ will be detected after eliminating the interference caused by $\mathbf{x}_2$ using SIC. Given that $\mathcal{C}(\hat{\mathbf{x}}_i) = 1 \; \forall i$, then $\mathcal{A} = [1,1,1]$.

However, if $\mathcal{C}(\hat{\mathbf{x}}_1) = 0$, then $\mathcal{A} = [0,0,0]$, which implies that the same composite packet will be retransmitted. Moreover, the channel information h_1 and received composite packet $\mathbf{y}_1$ are stored into two buffers, such that $\mathbf{h} = [h_1]$ and $\mathbf{Y}_1 = [\mathbf{y}_1]$. After retransmission, the received sequence $\mathbf{y}_2$ and the channel gain are buffered in $\mathbf{Y}_1$ and $\mathbf{h}$ respectively, and thus, the buffers' contents become $\mathbf{h} = [h_1, h_2]^T$ and $\mathbf{Y}_1 = [\mathbf{y}_1^T, \mathbf{y}_2^T]^T$,

where $(\cdot)^T$ is the transpose. To exploit the temporal diversity, MRC is applied to combine $\mathbf{y}_1$ and $\mathbf{y}_2$, where the combined packet can be written as

$$\mathbf{s}_{1,2} = \frac{\mathbf{h}^H}{\mathbf{h}^H\mathbf{h}}\mathbf{Y}_1 \tag{8.11}$$

where $(\cdot)^H$ is the Hermatian transpose. MLD is then used to extract $\hat{\mathbf{x}}_1$ from $\mathbf{s}_{1,2}$. The same process is repeated as long as $l \leq \mathcal{M}$ where $\mathbf{y}_l$ and h_l $\forall l$ are stored in the receiver buffers. When $l > \mathcal{M}$, the primary packet $\mathbf{x}_1$ is dropped, and the next packet $\mathbf{x}_2$ becomes the primary in the next transmission.

In the case that $\mathcal{C}(\hat{\mathbf{x}}_1) = 1$ in the lth transmission round where $l \leq \mathcal{M}$, the receiver uses $\hat{\mathbf{x}}_1$ and the SIC to eliminate the interference caused by $\mathbf{x}_1$ from all the stored sequences $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_l$, which can be expressed as

$$\mathbf{Y}_2 = \mathbf{Y}_1 - \sqrt{\alpha_{1,l}^{(l)}}\mathbf{h} \otimes \hat{\mathbf{x}}_1 \tag{8.12}$$

where $\otimes$ is a kronecker product. MRC is then used to combine the sequences in $\mathbf{Y}_2$, which now corresponds to the packet $\mathbf{x}_2$. Therefore,

$$\mathbf{s}_{2,l} = \frac{\mathbf{h}^H}{\mathbf{h}^H\mathbf{h}}\mathbf{Y}_2. \tag{8.13}$$

MLD is used to detect $\hat{\mathbf{x}}_2$. If $\mathcal{C}(\hat{\mathbf{x}}_2) = 1$, then SIC is used to eliminate the interference caused by $\mathbf{x}_2$. Thus,

$$\mathbf{Y}_3 = \mathbf{Y}_2 - \sqrt{\alpha_{2,l}^{(l)}}\mathbf{h} \otimes \hat{\mathbf{x}}_2. \tag{8.14}$$

The received sequences that correspond to $\mathbf{x}_3$ are then combined,

$$\mathbf{s}_{3,l} = \frac{\mathbf{h}^H}{\mathbf{h}^H\mathbf{h}}\mathbf{Y}_3. \tag{8.15}$$

Finally, MLD is applied to produce $\hat{\mathbf{x}}_3$.

Algorithm 8.2: Receiver Design.

Input: l, k, N_l, $\mathbf{h}$, $\mathbf{Y}_1$, α_l
Output: $\mathcal{A} \in \mathbb{B}^{1\times N_l}$

1 $\mathcal{A} \leftarrow [0,0,\ldots,0]$
2 **for** $n = 1 : N_l$ **do**
3 $\quad q \leftarrow k+n-1$
4 $\quad \mathbf{s}_{q,l} \leftarrow \frac{\mathbf{h}^H}{\mathbf{h}^H\mathbf{h}}\mathbf{Y}_n$
5 $\quad \hat{\mathbf{x}}_q \leftarrow \mathrm{MLD}(\mathbf{s}_{q,l})$
6 $\quad$ **if** $\mathcal{C}(\hat{\mathbf{x}}_q) = 0$ **then**
7 $\quad\quad$ **break**
8 $\quad \mathbf{Y}_{n+1} \leftarrow \mathbf{Y}_n - \sqrt{\alpha_{q,l}^{(u)}}\mathbf{h} \otimes \hat{\mathbf{x}}_q$
9 $\quad \mathcal{A}(n) \leftarrow 1$

8.5.2 General Receiver Design

By induction on the example in Section 8.5.1, the receiver design for the general case is given in Algorithm 8.2, where

$$\mathbf{h} = [h_\ell, h_{\ell+1}, \ldots, h_{l-1}, h_l]^T \tag{8.16}$$

$$\mathbf{Y}_1 = [\mathbf{y}_{k,\ell}^{(1)}; \mathbf{y}_{k,\ell+1}^{(2)}; \ldots; \mathbf{y}_{k,l-1}^{(u-1)}; \mathbf{y}_{k,l}^{(u)}] \tag{8.17}$$

$$\alpha_l = [\alpha_{k,l}^{(u)}, \alpha_{k+1,l}^{(u)}, \ldots, \alpha_{k+N_l-1,l}^{(u)}] \tag{8.18}$$

where the semicolon indicates the end of the row in matrix $\mathbf{Y}$. The received composite packet during lth transmission slot is given as,

$$\mathbf{y}_{k,l}^{(u)} = h_l\sqrt{\alpha_{k,l}^{(u)}}\mathbf{x}_{k,l}^{(u)} + \sum_{i=k+1}^{k+N_l-1} h_l\sqrt{\alpha_{i,l}^{(1)}}\mathbf{x}_{i,l}^{(1)} + \mathbf{w}_l \tag{8.19}$$

where u denotes the primary packet transmission counter, k denotes the primary packet number, and N_l denotes the total number of stacked packets during the lth transmission slot. The process at the receiver depends on $\mathbf{h}$ and $\mathbf{Y}_1$, which contain the current channel coefficients and received sequence of the lth transmission slot as well as the previous transmissions of the kth primary packet. The elements from previous transmissions can be extracted from the buffers which have elements of past transmissions if $\ell - l > 0$, otherwise the buffers are cleared. The receiver initializes the F_M with NACKs prior to decoding **(Step 1)**. The detection process is performed in a loop that iterates N_l times to sequentially detect the N_l GNOM multiplexed packets unless the

Table 8.1: Time complexity in milliseconds to compute (8.8a) for different SNRs and N_l where $L = 32$ and $\tau = 10^{-1}$.

	SNR (dB)						
	10	20	30	35	40	45	50
$N_l = 2$	1.9	0.6	0.7	0.7	0.7	0.7	0.8
$N_l = 3$	27.7	1.9	0.6	0.6	0.6	0.7	0.8
$N_l = 4$	–	3.4	1.8	0.7	0.8	0.7	0.8
$N_l = 5$	–	24.6	6.0	7.2	0.7	0.8	1.0
$N_l = 6$	–	–	16.1	11.4	8.9	0.9	1.0
$N_l = 7$	–	–	–	26.9	17.3	20.3	1.2

CRC check fails for the currently detected packet **(Step 6)**. The loop starts by defining an auxiliary variable q representing the index of the packet under process **(Step 3)**. MRC is applied to the packet under process **(Step 4)** prior to the detection process **(Step 5)**. **(Steps 6–7)** check the CRC for the detected packet, and if the packet fails the CRC check, the receiver halts the process **(Step 7)** and returns the initialized F_M in **(Step 1)**. If the packet passes the CRC check, SIC is performed to eliminate the interference caused by the detected packets to the other secondary GNOM packets with less power **(Step 8)**. Also, an ACK is set for the primary packet in the F_M **(Step 9)**. This process continues to detect the remaining secondary packets unless a secondary packet fails the CRC check, then an F_M is returned indicating the last correctly detected secondary GNOM packet.

8.5.3 Computational Complexity

The GNOM system design generally follows that of NOMA where the power coefficients of all data symbols should be computed before the superimposing process [24]. Consequently, the computational complexity depends on N_l, N_{cap}, and the SNR. However as depicted in Algorithm 8.1, GNOM should find the maximum number of packets that can be superimposed, which is bounded by N_{cap}. Therefore, the GNOM transmitter complexity can be controlled by limiting the value of N_{cap}, at the expense of some throughput loss. For example, lower N_{cap} values can be used for IoT nodes that have limited computational capabilities. For the case of NOM, the power can be assigned in a fixed or adaptive manner. For the fixed power assignment, the complexity at the transmitter is negligible while the NOM complexity will be equivalent to GNOM when $N_{\text{cap}} = 2$. Therefore, the complexity of NOM with adaptive power will be considered for comparing the complexity of both systems. It is worth noting that the NOM performance with adaptive power slightly outperforms that with fixed power [2].

Table 8.2: Algorithm 1 time complexity in milliseconds for different SNRs and N_{cap} where $L = 32$ and $\tau = 10^{-1}$.

	SNR (dB)						
	10	20	30	35	40	45	50
NOM	2.9	1.3	1.6	1.5	1.6	1.7	1.5
$N_{\text{cap}} = 3$	31.4	3.2	2.2	2.4	2.0	2.4	2.2
$N_{\text{cap}} = 4$	31.3	7.1	4.2	3.0	2.8	3.0	3.0
$N_{\text{cap}} = 5$	31.4	34.9	10.3	10.8	4.3	4.0	3.8
$N_{\text{cap}} = 6$	31.8	33.9	28.6	22.9	13.5	5.1	5.0
$N_{\text{cap}} = 7$	31.7	34.1	27.9	51.4	31.7	26.5	6.0

In this work, the computation time is used to indicate the system complexity because counting the number of floating-point operations is intractable. Hence, the time complexity is evaluated using Matlab on a laptop with an Intel i5 processor and 16-GB RAM. To evaluate the time complexity of the power allocation process, Table 8.1 presents the time required to compute (8.8a) for $N_l = 2, 3, \ldots, 7$. As can be noted from the table, the time complexity is proportional to N_l and inversely proportional to SNR. Increasing SNR makes the BER curves steeper, which enables the search process to converge fast, and vice versa. On the other hand, increasing N_l increases the computational complexity of the BER formulas, and hence, the computations time increases as well. It is worth noting that the search for the power coefficients is guided by boundaries derived in [27], which significantly reduces the search space. Table 8.1 also shows that when N_l would lead to an infeasible solution, the computation time increases significantly. The infeasible values of N_l are given by the shaded cells in Table 8.1, and the empty cells have the same value as the first infeasible cell.

Table 8.2 shows the time complexity in milliseconds for Algorithm 8.1 while considering various instantaneous SNR values. The first row in the table corresponds to the NOM. Generally speaking, the trends observed in Table 8.1 are also applicable to Table 8.2, except that computation times are larger since Algorithm 8.1 also involves finding N_l. Consequently, significant complexity reduction can be achieved if we use efficient approaches to estimate N_l rather than using linear search. A possible approach is to exploit the relation between N_l and SNR shown in Table 8.1 to reduce the search space for N_l, For example, at high SNRs we may start the search process at $N_l = N_{\text{cap}}$. Moreover, we can stop the search process without going through infeasible cases. The infeasible solutions are distinguished using the shaded cells in the table. As compared to NOM, the GNOM complexity is relatively higher. However, the difference for several cases of interest is generally tolerable since it is in the range of a few milliseconds in various cases. As can be noted from the complexity results, increasing N_{cap} for the GNOM would increase the computational complexity, and hence

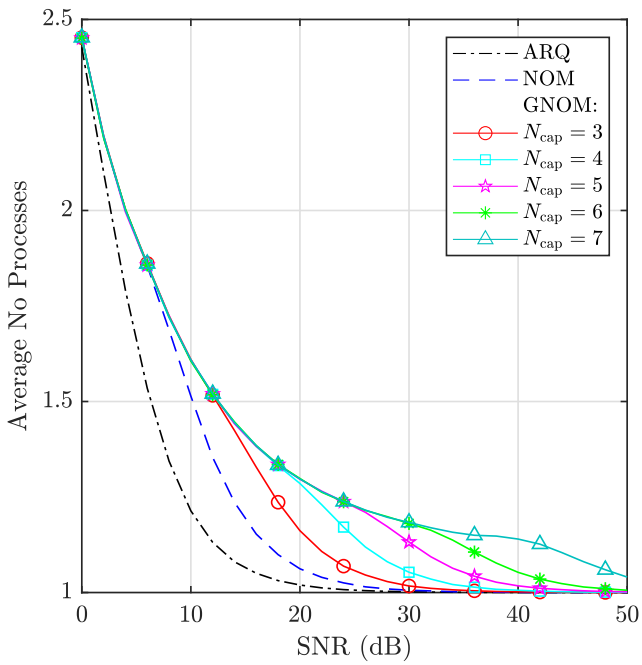


Figure 8.2: Average number of detection processes at the receiver for GNOM with different N_{cap} compared to the ARQ and NOM [2], where $\tau = 10^{-1}$ and $L = 32$.

the computational power. Nevertheless, the GNOM can reduce the average transmission power per packet as shown in Section 8.6. By noting that the transmission power is typically more significant than the computational power [186], then GNOM can reduce the overall power consumption of the transmitting IoT node, which is an essential feature for power-constrained nodes.

At the receiver side, the GNOM detects the packets using SIC, and stops once a packet is declared erroneous. Figure 8.2 shows the average number of detection operations per data packet for $\tau = 10^{-1}$ and $L = 32$. It can be seen that all the considered systems have the same average number of detection operations at low and high SNRs. At low SNRs, the GNOM receiver will mostly stop after detecting the first packet, and thus, the average number of detection operations approaches $\mathcal{M}$. At high SNRs all packets will mostly pass the error check and thus, each packet will be detected only once. At moderate SNRs, the average processing per packet for GNOM is generally higher than ARQ and NOM because NOM usually has one packet with very low power, and hence, it is more likely that it will fail the error detection process, retransmitted, and processed again at the receiver. Nevertheless, the difference between GNOM and NOM is generally less than 10%.

Table 8.3: Simulation parameters summary.

Parameter	Value
No. transmitted packets	$M = 3 \times 10^6$
GNOM design parameter	$N_{\text{cap}} > 2$
Packet length	$L \in \{32, 128\}$
Channel model	Rayleigh block fading
Maximum transmission	$\mathcal{M} = 3$
ARQ type	ARQ type-I
PER threshold	$\tau \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$

8.6 Numerical Results and Discussions

This section presents the performance evaluation of the proposed transmission model. Monte Carlo simulation is performed for $M = 3 \times 10^6$ packets with $N_{\text{cap}} = 7$ packets per transmission slot unless otherwise stated. Two packet lengths are considered, $L = 32$ and 128 symbols, and maximum allowable transmissions $\mathcal{M} = 3$. The performance is evaluated for PER threshold $\tau \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$. Table 8.3 summarises the simulation parameters and assumptions. The channel fading coefficients h_m $\forall m$ are generated as mutually independent and identically distributed (i.i.d) random variables whose envelope follows the Rayleigh distribution. The performance of the proposed system is measured through several metrics and compared to the conventional ARQ system in which a single packet with normalized unity power is transmitted. Also, the performance is compared to GNOM system [2] with $\alpha_l = [0.8, 0.2]$. The SNR shown in the figures is defined as SNR $\triangleq 1/2\sigma_w^2$. Additionally, in the legend, GNOM denotes the proposed GNOM, ARQ denotes the conventional ARQ system and NOM denotes the technique proposed in [2].

The GNOM number of multiplexed packets versus SNR over AWGN channels is shown in Figure 8.3 for $L = 32$, where the benchmarks are the conventional ARQ and NOM systems. It should be noted that N_l, in this context, represents spectral efficiency with a unit of bit per channel use or bits per symbol. As can be noted from the figure, when the PER threshold decreases, higher SNR is required to obtain the same value of N_l for both NOM and GNOM systems. It can be seen that the NOM saturates at $N_l = 2$, whereas the GNOM saturates at $N_l = N_{\text{cap}}$. It is worth mentioning that the results for $L = 128$ follow the same trend but with a shift to the right for a fraction of a dB. Furthermore, the power allocation per packet for the AWGN channel is shown in Figure 8.4. It is noted that power coefficients for the secondary packets decrease significantly when the total number of multiplexed packets increases. Moreover, the more stringent the PER requirements, the smaller the power coefficients

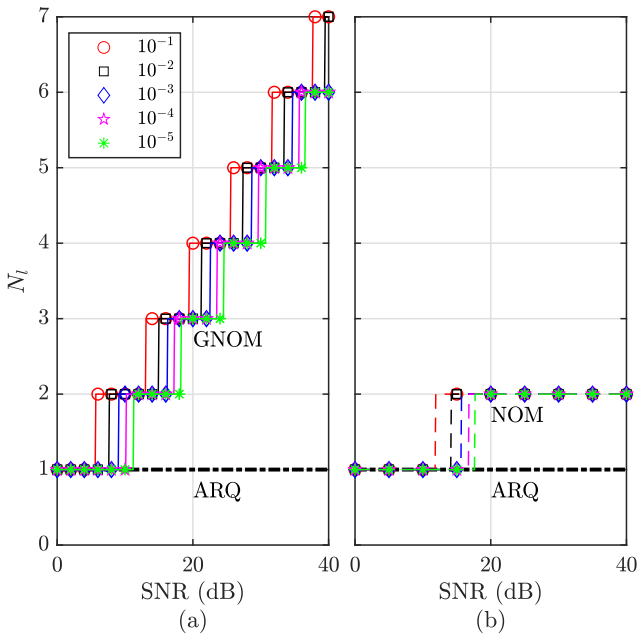


Figure 8.3: Spectral efficiency, N_l, of the proposed GNOM scheme for $\tau = 10^{-1}$–10^{-5} in AWGN channel where $L = 32$.

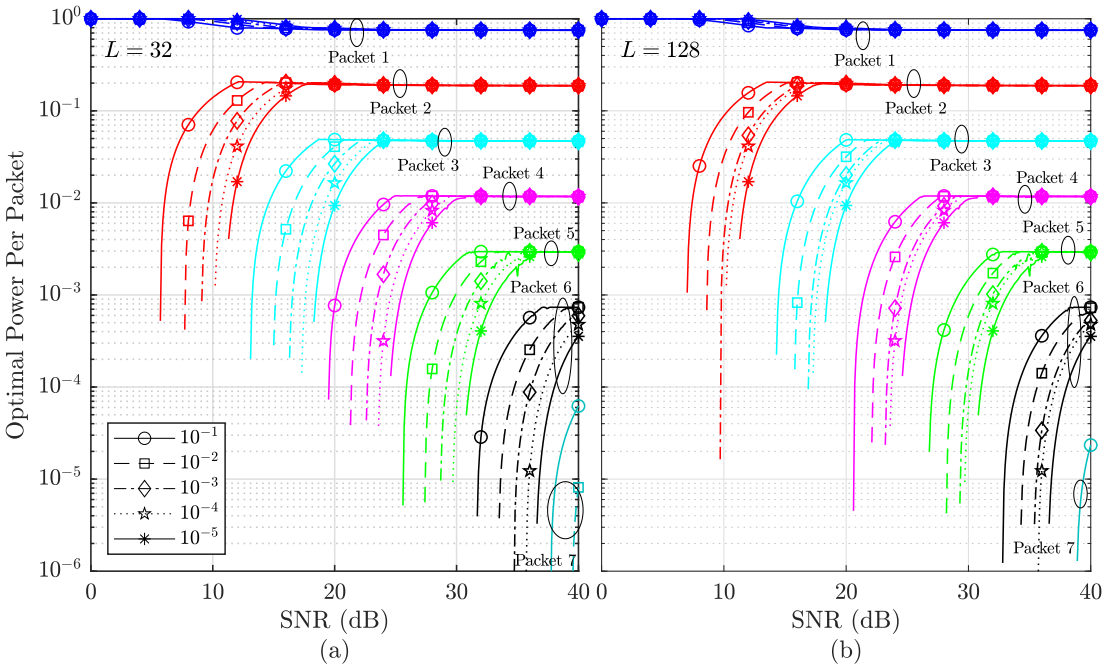


Figure 8.4: Power allocation per packet for GNOM packets for different PER thresholds in AWGN channel. (a) $L = 32$. (b) $L = 128$.

for the secondary packets get, whereas the primary packet power coefficient increases. It is worth mentioning that at relatively high SNRs, the power coefficients per packet converge regardless of PER threshold.

The average power allocations per packet for the proposed GNOM in a Rayleigh fading channel is demonstrated in Figure 8.5. Before each transmission, the value of

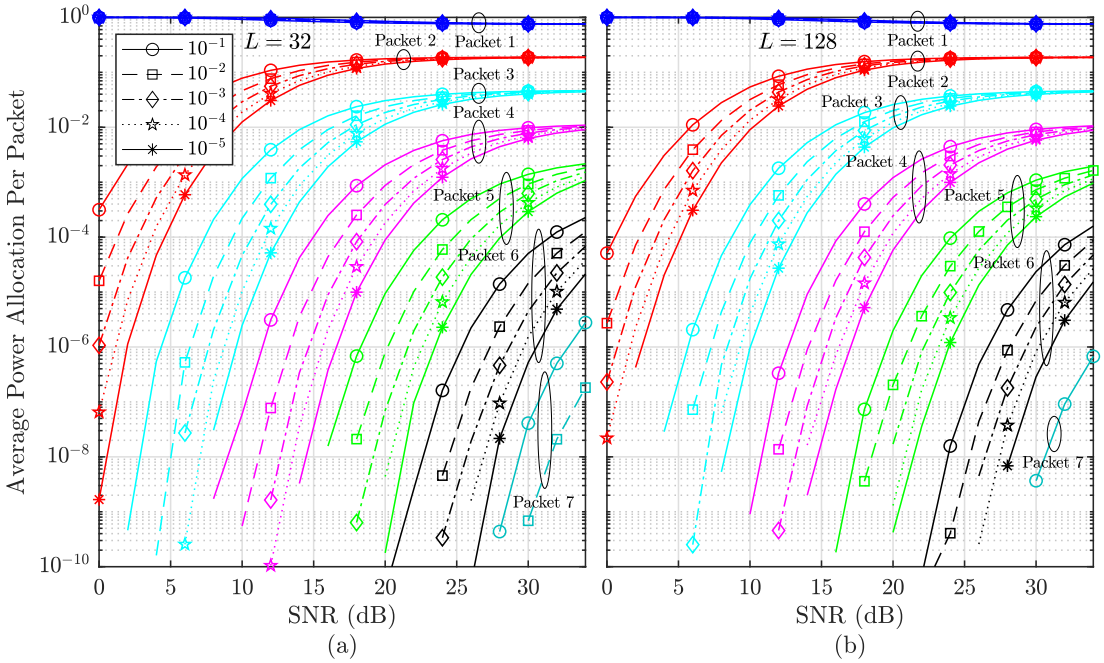


Figure 8.5: Average power allocations per packet for GNOM using various PER thresholds in Rayleigh fading channels. (a) $L = 32$. (b) $L = 128$.

N and the power coefficients for the N packets are determined using Algorithm 8.1. The average power allocation per packet is computed as the ratio of the sum of powers allocated per packet to the total channel uses. As depicted in the figure, when the PER threshold reduces from 10^{-1} to 10^{-5}, more power is allocated to the primary packet to satisfy its PER threshold. Therefore, decreasing the PER threshold for the primary packet increases its average power allocation. Such performance is obtained because the primary packet has the highest priority in terms of PER satisfaction. Moreover, similar to the AWGN case, the power coefficients per packet at high SNRs converge to the same value regardless of the PER threshold.

Throughput is considered one of the important metrics to examine the performance of a retransmission system. The throughput of the proposed GNOM system for various PER thresholds is compared to the conventional ARQ and GNOM systems in Figure 8.6, where the throughput is defined as the total number of successfully received packets divided by a total number of transmission sessions. The spectral efficiency is considered as an upper bound for the throughput because the spectral efficiency does not consider dropped and retransmitted packets. From the results, it is evident that a significant throughput gain is achieved using the proposed GNOM compared to the conventional ARQ and NOM systems. For example, considering the conventional ARQ system as a benchmark, the throughput improvement ratio of GNOM exceeds

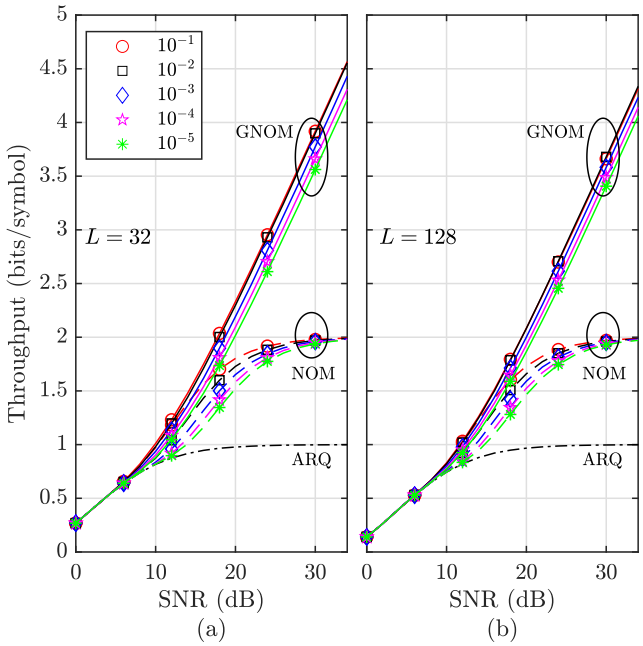


Figure 8.6: Throughput for various PER thresholds. (a) $L = 32$. (b) $L = 128$.

350% at SNR of 30 dB for $L = 32$, whereas NOM achieves only 200% throughput improvement ratio. The impact of the PER threshold is generally mild for the considered range of PER thresholds, where the difference is about 0.3 bits/symbol for SNRs more than 25 dB. As can be noted from the figure, the throughput degrades by decreasing the PER threshold because higher power is required to satisfy the PER constraint. Therefore, the number of multiplexed packets is reduced. Using the same justification, the throughput is improved when the PER is relaxed because more packets can be multiplexed at the transmitter. However, the throughput will not consistently improve because increasing the PER thresholds will increase the packet retransmission rates. The GNOM, NOM and the conventional ARQ are performing roughly the same at low SNRs because the outcome of the adaptation process most likely will result in a single packet transmission.

The packet drop rate (PDR) for GNOM, NOM and the conventional ARQ is demonstrated in Figure 8.7. The PDR is defined as the ratio of the total number of dropped packets to M. The figure shows that PDR improves with a more stringent PER threshold, which is expected because the power allocated to each packet is inversely proportional to the PER threshold. Nevertheless, it can be noted that the PDR of the three systems tend to converge to the same PDR for the PER threshold of 10^{-5}, unlike the 10^{-1} case. According to [216], the PDR required for various applications classified by ITU is 10^{-3}. Therefore, it can be noted from Figure 8.7 that the proposed system can

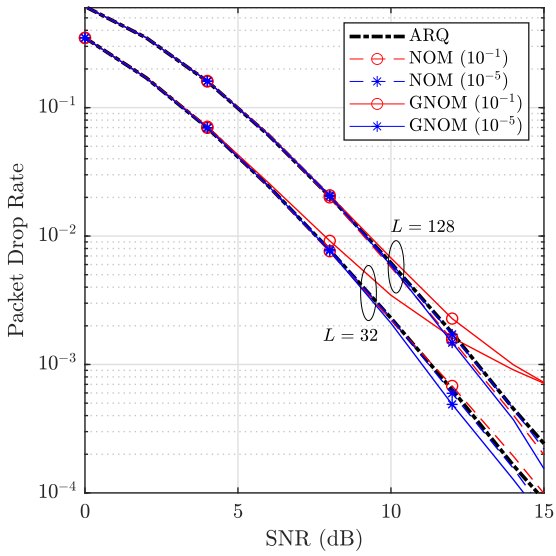


Figure 8.7: Packet drop rate. (a) $L = 32$. (b) $L = 128$.

successfully achieve a PDR of 10^{-3} for all demonstrated thresholds at relatively low SNRs.

In ARQ systems, a packet may undergo multiple transmissions before it is successfully received. Consequently, such packets will experience significant delay. Typically, more transmissions are performed in the low SNR region, whereas a single transmission is generally sufficient for a packet to be received successfully at high SNRs. The average number of transmission rounds for various PER thresholds is depicted in Figure 8.8, where the average number of transmissions per packet is defined as the ratio of the total channel uses to M. As can be noted from the figure, the average number of transmission approaches $\mathcal{M}$ at extremely low SNRs. However, the performance is significantly improved when SNR is increased. Another riveting result is that the GNOM and NOM outperform the conventional ARQ, where the delay may become less than one in the latter. This can never be achieved by conventional ARQ because the total time required to transmit M packets can be much less than $M \times T_{PCK}$, where T_{PCK} is the time interval of one packet. Such performance can never be achieved with conventional ARQ because in the best-case scenario, one packet is transmitted in each T_{PCK} time interval, whereas two packets are transmitted with NOM and N_{cap} packets are transmitted with GNOM.

The average consumed power per correctly received packet is shown in Figure 8.9. The average consumed power is defined as the total power used to transmit M packets divided by the total number of correctly received packets. The power used to transmit

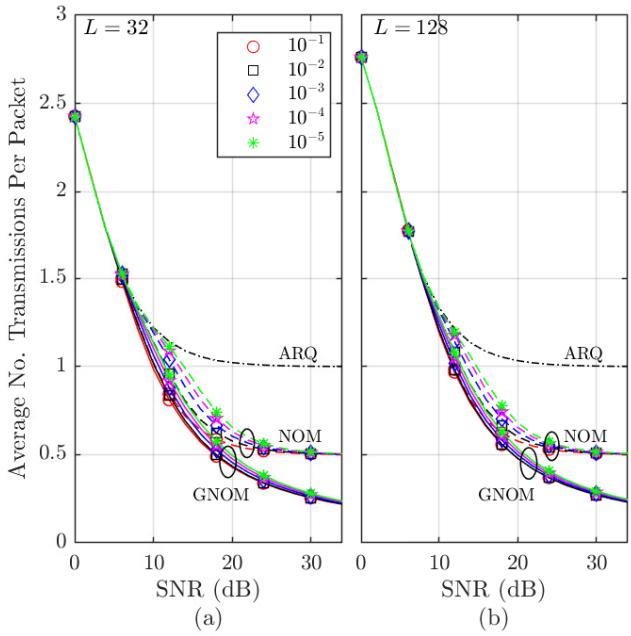


Figure 8.8: Average number of transmissions per packet. (a) $L = 32$. (b) $L = 128$.

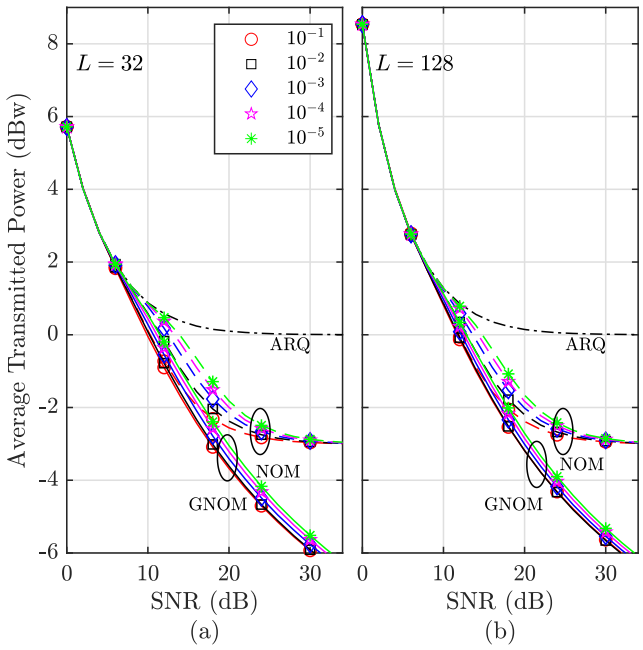


Figure 8.9: Average consumed power per packet. (a) $L = 32$. (b) $L = 128$.

each packet is normalized to unity. As can be noted from Figure 8.9a, the average power per packet for $L = 32$ is about 6 dBw at SNR of 0 dB. Such high power consumption is partially due to the retransmission process and since the dropped packets' power is added to the total consumed power. However, the power consumption decreases significantly by increasing SNR, where at SNR of 30 dB it becomes about -6

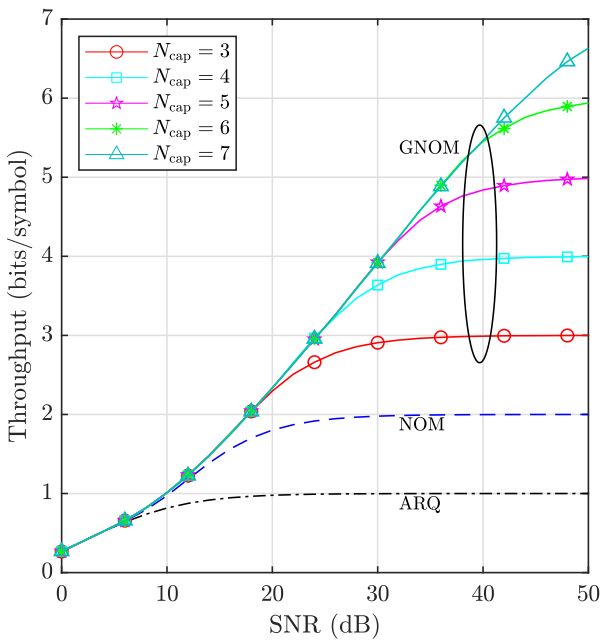


Figure 8.10: Throughput of GNOM for different N_{cap} compared to the ARQ and NOM [2], where $\tau = 10^{-1}$ and $L = 32$.

dBw and -3 dBw for the GNOM and NOM, respectively. It is worth noting that Figure 8.9 is generated with a fixed transmission power for the three systems. Although this might seem unfair because power adaptation can be also applied to conventional ARQ, the efficiency of the power amplifier will be dropped significantly due to the small transmission power. Therefore, the three systems consider a fixed transmission power equal to $p_{\max} = 1$.

Figure 8.10 compares the throughput of the proposed GNOM with the conventional ARQ and NOM systems. The PER threshold $\tau = 10^{-1}$, $L = 32$ and $N_{\text{cap}} = 3, 4, \ldots, 7$. The results in the figure confirm that GNOM outperforms the other two benchmarks at moderate and high SNRs, while the throughput of the considered systems is comparable for SNRs less than 8 dB. The throughput for N_{cap} is very close to the NOM because the power coefficients for the NOM and GNOM are roughly identical, and thus, the case of $N_{\text{cap}} = 2$ is not shown in the figure. Moreover, it can be noted that the GNOM throughput for all N_{cap} values is generally equivalent at low SNRs. Such performance is obtained because the number of packets superimposed at the transmitter is generally less than N_{cap}, particularly at low SNRs. The convergence of the throughput to N_{cap} is achieved at lower SNRs because superimposing large number if packets introduce significant interference, and thus may only be achieved at high SNRs.

8.7 Conclusions

In this chapter, a novel GNOM transmission scheme for high power and spectrum efficient communications is proposed and evaluated under various packet lengths and PER thresholds. In the proposed GNOM scheme, several packets are adaptively multiplexed in the power domain where each multiplexed packet contains new and re-transmitted data symbols to improve the system throughput and spectral efficiency. The proposed system can be configured to trade-off the complexity and performance to suit the specifications and requirements of various types of IoT nodes and applications by adjusting the maximum number of packets that can be combined in each transmission session. The obtained results demonstrated that at high SNRs, the proposed GNOM achieves significant spectral efficiency gain over the conventional ARQ and NOM systems. For example, the throughput improvement ratio of the GNOM over ARQ and NOM is about 450% and 225%, respectively for SNR of 35 dB, $L = 32$ and $N_{\text{cap}} = 7$. Similarly, it was shown that GNOM outperforms conventional ARQ and NOM in terms of power efficiency and delay. For example, because the average number of combined packets is 4 at SNR of 35 dB and $L = 32$, then the average power per packet and the average number of transmissions are only 23% and 50% of the ARQ and NOM, respectively.

Chapter 9

Conclusions and Future Directions

9.1 Conclusions

This thesis studied non-orthogonal multiple access (NOMA) for dense future wireless networks while focusing on performance analysis and optimisation. The necessary technical background is provided for the readers to understand the content of the work chapters. Also, a comprehensive literature review of the error rate performance of NOMA is supplemented to highlight the advancements in the field. Furthermore, the aim of the thesis was to study the error rate and throughput performance of NOMA systems. Although the bit error rate (BER) performance of conventional NOMA systems is derived for various modulation orders, number of users, and number of receiving antennas; the BER is derived for improved NOMA schemes such as the data-aware interference exploitation based power assignment and the joint-multiuser Gray-labelling. In addition, the power assignment is optimised for various settings based on the derived expressions to achieve a certain objective while satisfying the users' quality of service (QoS) requirements. Moreover, the throughput of NOMA is analysed and evaluated for cognitive NOMA, adaptive modulation NOMA systems and automatic repeat request (ARQ)-based non-orthogonal multiplexing (NOM) systems. In spite of the degraded BER performance of NOMA due to the inter-user interference (IUI), its throughput outperforms the orthogonal multiple access (OMA) counterpart at moderate and high signal to noise ratio (SNR) values, where the performance gain at high SNR values equals to the number of multiplexed users. Finally, the presented analysis in this thesis may serve as the first step for the system designer to gain insights and quantify the NOMA BER and throughput performance for various settings and scenarios.

9.2 Future Research Directions

This section aims to highlight future directions and open research problems of NOMA systems while focusing on the error rate performance. These future directions include the error rate performance analysis of NOMA integrated with intelligent reflecting surface (IRS), joint radar-communications, millimeter wave (mmWave) and terahertz (THz) communications. Moreover, the open research problems include the design and analysis of low complexity detectors for NOMA systems, as well as the error rate performance analysis of the non-ideal NOMA systems due to hardware imperfections. Furthermore, error rate analysis of new NOMA variants and machine learning-enabled NOMA are presented as possible future directions.

9.2.1 Hybrid IRS

IRS can enhance the system performance by controlling the propagation of signals in the environment via passive smart metasurface elements. However, there are several challenges, such as acquiring channel estimates between all metasurface elements and users. Since the IRS is passive, it cannot transmit pilot/training signals to facilitate channel estimation. Hence, novel channel estimation methods should be considered for IRS, such as estimating the cascaded user-IRS-base station (BS) channels, or by considering semi-passive IRS models which have sensing capabilities [233]. Also, the phase cannot be controlled continuously in practice. Therefore, it is crucial to consider quantised phase control.

The error rate performance of IRS-assisted NOMA has been investigated for the downlink (DL) direction only, where the assumptions of continuous and perfect phase are assumed [234, 235]. Hence, analyzing the error rate performance under imperfect and quantised phase is worth studying to get better practical insights. Moreover, the design and analysis for the uplink (UL) direction is a promising direction as well. Furthermore, studying the system performance while considering the direct link between the users and the BS is still an open research question. Hence, the design and analysis of the dual function IRS in DL and UL-NOMA systems is attractive.

9.2.2 Joint Radar-Communications Systems

Joint radar-communications systems are promising to solve the spectrum scarcity problem as both tasks are processed on the same frequency band. Hence, a single beamformer can be optimised to satisfy the requirements for both signals. For example, a certain error rate performance or sum-rate requirements for the communications part, while certain sensing power for the sensing part. While NOMA is used to serve multiple users, it can also be used for sensing simultaneously [236]. Investigating the error performance of the communications part and the effectiveness of the sensing part of the NOMA-based joint radar-communications system while considering proper signalling is still an open research problem.

9.2.3 mmWave and THz Communications

Higher frequency bands such as mmWave bands are currently used in the fifth generation (5G) standard to satisfy the growth in higher data requirements since wider bandwidths are available at such frequencies. However, current trends indicate that THz bands, which are wider than millimetre bands, are considered a key technology for future wireless networks [81]. Nonetheless, the propagation limitations become more challenging and the communications would be directional and take place mostly in line-of-sight (LOS) environments. Therefore, the near-far problem would exist even if the users are relatively in close proximity. Consequently, studying the best conditions for pairing the NOMA users under certain error rate performance requirements can provide a better understanding of the practical scenarios in which NOMA performance would be maximised. Furthermore, technologies such as cooperative communications and IRS can be used to improve network coverage.

9.2.4 Low Complexity Detectors

One of the crucial concerns for deploying NOMA with more than two-user per resource block is the associated detectors' complexity. While either successive interference cancellation (SIC) or joint-multiuser maximum likelihood detector (JMLD) computational complexities might limit the applications of NOMA, designing new detectors with lower complexity for the UL or DL is a promising research direction. Most importantly is the UL, since JMLD is introduced as a solution for the SIC error floor problem [68]. However, it should be noted that the adoption of NOMA in standards and industry has been limited due to its trade-off between performance gain and

complexity [237, 238]. These practical considerations highlight the need to address the complexity issue while maintaining acceptable performance. Therefore, further research and development are required to explore low complexity detectors for NOMA systems, providing motivation and future directions for error rate analysis.

9.2.5 Hardware Imperfections

Although several articles studied the impact of imperfect SIC and channel estimation errors, the literature works studying the impact of residual hardware impairment (RHI), I/Q imbalance and phase noise are limited [198, 239–241]. In addition, problems including frame synchronization, time synchronization and carrier synchronization over the performance of NOMA systems are missing from the literature. Furthermore, the works of orthogonal frequency division multiplexing (OFDM)-NOMA systems reported in this survey mostly assume ideal conditions in terms of time and frequency synchronization, as well as cyclic prefix (CP) length which is assumed to be longer than the channel maximum delay spread, and hence the inter-symbol interference is removed perfectly. However, such ideal conditions cannot be achieved in practical scenarios, where imperfect time and frequency synchronization are likely to occur. Additionally, OFDM-NOMA is prone to user mobility, which introduces Doppler shift and loss of subcarriers orthogonality. Applying orthogonal time frequency space (OTFS) and studying the error rate performance of NOMA systems under the previously mentioned imperfections can be important directions for future research.

9.2.6 Variants

While NOMA is mainly recognized as a beyond 5G multiple access technique due to its ability to augment the spectral efficiency (SE) of communication systems, its advantages are not limited to SE enhancement only, they extend to other several aspects of the communication systems such as reliability, energy efficiency and latency. In the following, we highlight several more general variants of NOMA that may offer significant improvements to the communications system.

9.2.6.1 Multirate NOMA

Multirate NOMA has been introduced to improve the reliability of the NOMA system [242]. Unlike conventional NOMA systems, mutlirate NOMA allows pairing users with different symbol rates, thus the energy dimension is exploited as another degree

of freedom when performing power assignment (PA). Error rate analysis of multirate NOMA systems for UL and DL, as well as the design of optimal low-complexity detectors, are still open research problems. Further problems that need investigation are user pairing, optimal PA and modulation orders selection.

9.2.6.2 NOM

NOM has been proposed for hybrid automatic repeat request (HARQ)-based wireless systems to allow multiplexing new packet transmission with retransmissions such that the energy efficiency and latency are improved [2]. While [124] considers generalizing NOM to more than two packets, the PA is assumed fixed during the retransmissions. Allowing adaptive power and considering higher modulation orders can improve performance further. Nonetheless, the design of a low-complexity detector and low-complexity algorithm to solve the PA and the number of multiplexed packets are important research directions.

9.2.6.3 RSMA

While noting that the sheer number of articles reported in literature deals with single-input-single-output (SISO) settings, NOMA is effective for SISO designs and its SE reduces for most of the multiple antennas scenarios [243]. Clerckx *et al.* [244] presented rate-splitting multiple access (RSMA) as another variant of NOMA for and OMA for multiple antenna settings, where it is based on splitting the user's message into different parts such as common and private parts, each of which can be handled independently, leading to increasing both efficiency and reliability. Consequently, RSMA promises to achieve higher multiplexing gains and rates, robustness to user deployment settings, and serve a larger number of users [245–247]. In addition, latency caused by the long chain of SIC in NOMA is reduced due to more efficient use of SIC. Therefore, researchers are invited to study the error rate performance of RSMA for single-input-multiple-output (SIMO), multiple-input single-output (MISO) and multiple-input multiple-output (MIMO) settings.

9.2.7 Machine Learning

While both statistical-based analysis and machine learning-based analysis of NOMA are reported in the literature, the latter has the potential to revolutionize the performance of NOMA by leveraging data-driven models and advanced algorithms. More

specifically, machine learning is used in prediction and optimisation problems such as channel estimation, adaptive modulation, constellation design, and classification [168] as well as receiver design [248]. Nonetheless, analyzing the performance of machine learning-based approaches using the traditional error rate metrics including BER, symbol error rate (SER) and block error rate (BLER) can be very complicated due to the black-box nature of machine learning-based approaches. Therefore, adopting model-dependent evaluation metrics such as accuracy, computational complexity, and reliability besides the hybrid analytical approach can provide a comprehensive understanding of NOMA error rate performance.

Furthermore, evaluating the error rate of machine learning-enabled NOMA has the following challenges. For instance, limited labeled data, NOMA system complexity, scalability issues, generalization across diverse environments, and real-time processing constraints. Since NOMA networks have a dynamic nature, acquiring labeled data is challenging, and modeling interference management and resource allocation complexities is complicated. Scaling techniques to handle numerous users while ensuring generalization across scenarios is complex. In addition, real-time analysis is required for effective resource allocation.

References

[1] H. Yahya, E. Alsusa, and A. Al-Dweik, "Design and analysis of NOMA with adaptive modulation and power under BLER constraints," *IEEE Trans. Veh. Technol.*, vol. 71, pp. 11228–11233, Oct. 2022.

[2] A. Al-Dweik and Y. Iraqi, "High throughput wireless links for time-sensitive WSNs with reliable data requirements," *IEEE Sensors J.*, vol. 21, pp. 24890–24898, Nov. 2021.

[3] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Net.*, vol. 34, pp. 134–142, May/Jun. 2020.

[4] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?," *Nature Electron.*, vol. 3, no. 1, pp. 20–29, 2020.

[5] M. Scudellari, "How the pandemic might play out in 2021 and beyond," *Nature*, pp. 22–25, 2020.

[6] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, pp. 28–41, Sep. 2019.

[7] "Ericsson mobility report, Q2 update, August 2021," tech. rep., Ericsson, 2021. An optional note.

[8] "Cisco annual Internet report (2018–2023)," tech. rep., Cisco, 2020. An optional note.

[9] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *IEEE Veh. Technol. Conf. (VTC)*, (Dresden, Germany), pp. 1–5, Jun. 2013.

[10] A. I. Perez-Neira, M. Caus, and M. A. Vazquez, "Non-orthogonal transmission techniques for multibeam satellite systems," *IEEE Commun. Mag.*, vol. 57, pp. 58–63, Dec. 2019.

[11] Y. Liu, Z. Qin, Y. Cai, Y. Gao, G. Y. Li, and A. Nallanathan, "UAV communications based on non-orthogonal multiple access," *IEEE Wireless Commun.*,

vol. 26, pp. 52–57, Feb. 2019.

[12] H. Marshoud, S. Muhaidat, P. C. Sofotasios, S. Hussain, M. A. Imran, and B. S. Sharif, "Optical non-orthogonal multiple access for visible light communication," *IEEE Wireless Commun.*, vol. 25, pp. 82–88, Apr. 2018.

[13] M. Jain, N. Sharma, A. Gupta, D. Rawal, and P. Garg, "Performance analysis of NOMA assisted underwater visible light communication system," *IEEE Wireless Commun. Lett.*, vol. 9, pp. 1291–1294, Aug. 2020.

[14] V. Raghavan and J. Li, "Evolution of physical-layer communications research in the post-5G era," *IEEE Access*, vol. 7, pp. 10392–10401, Jan. 2019.

[15] A. R. Mishra, *Fundamentals of Network Planning and Optimisation 2G/3G/4G: Evolution to 5G.* John Wiley and Sons, Ltd, 2018.

[16] M. Kishk, A. Bader, and M. S. Alouini, "Aerial base station deployment in 6G cellular networks using tethered drones: The mobility and endurance tradeoff," *IEEE Veh. Technol. Mag.*, pp. 0–0, Sep. 2020.

[17] E. Calvanese Strinati, S. Barbarossa, J. L. Gonzalez-Jimenez, D. Ktenas, N. Cassiau, L. Maret, and C. Dehos, "6G: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication," *IEEE Veh. Technol. Mag.*, vol. 14, pp. 42–50, Sep. 2019.

[18] M. B. Shahab, S. J. Johnson, M. Shirvanimoghaddam, M. Chafii, E. Basar, and M. Dohler, "Index modulation aided uplink NOMA for massive machine type communications," *IEEE Wireless Commun. Lett.*, vol. 9, pp. 2159–2162, Dec. 2020.

[19] M. Vaezi, R. Schober, Z. Ding, and H. V. Poor, "Non-orthogonal multiple access: Common myths and critical questions," *IEEE Wireless Commun.*, vol. 26, pp. 174–180, Oct. 2019.

[20] K. Q. Zhang, *Digital Modulation*, ch. 4, pp. 70–102. John Wiley and Sons, Ltd, 2015.

[21] A. Goldsmith, *Wireless communications*. Cambridge, UK: Cambridge Univ. Press, 2005.

[22] M. Simon and M. Alouini, *Digital Communication Over Fading Channels*. Hoboken, New Jersey, USA: Wiley-Interscience, 2 ed., 2005.

[23] T. Assaf, A. Al-Dweik, M. S. E. Moursi, H. Zeineldin, and M. Al-Jarrah, "NOMA receiver design for delay-sensitive systems," *IEEE Syst. J.*, vol. 15, pp. 5606–5617, Dec. 2021.

[24] H. Yahya, E. Alsusa, and A. Al-Dweik, "Exact BER analysis of NOMA with arbitrary number of users and modulation orders," *IEEE Trans. Commun.*, vol. 69,

pp. 6330–6344, Sep. 2021.

[25] Y. Qi, X. Zhang, and M. Vaezi, "Over-the-air implementation of NOMA: New experiments and future directions," *IEEE Access*, vol. 9, pp. 135828–135844, 2021.

[26] I. Lee and J. Kim, "Average symbol error rate analysis for non-orthogonal multiple access with M-ary QAM signals in Rayleigh fading channels," *IEEE Commun. Lett.*, vol. 23, pp. 1328–1331, Aug. 2019.

[27] Y. Iraqi and A. Al-Dweik, "Power allocation for reliable SIC detection of rectangular QAM-based NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 70, pp. 8355–8360, Aug. 2021.

[28] F. Kara and H. Kaya, "A true power allocation constraint for non-orthogonal multiple access with M-QAM signalling," in *IEEE Microw. Theory Techn. Wireless Commun. (MTTW)*, vol. 1, (Riga, Latvia), pp. 7–12, Oct. 2020.

[29] A. Chauhan and A. Jaiswal, "Power allocation constraint and error rate analysis for multi-user PD-NOMA system employing M–ary PSK," *IEEE Commun. Lett.*, vol. 27, pp. 1452–1456, May 2023.

[30] H. Yahya, E. Alsusa, A. Al-Dweik, and M. Debbah, "Cognitive NOMA with blind transmission-mode identification," *IEEE Trans. Commun.*, vol. 71, pp. 2042–2058, Apr. 2023.

[31] E. C. Cejudo, H. Zhu, and O. Alluhaibi, "On the power allocation and constellation selection in downlink NOMA," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, (Toronto, ON, Canada), pp. 1–5, Sep. 2017.

[32] A. Hilario-Tacuri, "Closed-form expressions for symbol error rate of two-user NOMA systems," in *IEEE Int. Conf. Electron., Elect. Eng. Comput. (INTERCON)*, (Lima, Peru), pp. 1–4, Sep. 2021.

[33] S. Bisen, P. Shaik, J. Jose, J. Nebhen, O. Krejcar, and V. Bhatia, "ASER analysis of generalized hexagonal QAM schemes for NOMA systems over Nakagami-m fading channels," *Int. J. Commun. Syst.*, vol. 36, p. e5440, Jan. 2023.

[34] T. Assaf, A. J. Al-Dweik, M. S. E. Moursi, H. Zeineldin, and M. Al-Jarrah, "Exact bit error-rate analysis of two-user NOMA using QAM with arbitrary modulation orders," *IEEE Commun. Lett.*, vol. 24, pp. 2705–2709, Dec. 2020.

[35] F. Kara and H. Kaya, "BER performances of downlink and uplink NOMA in the presence of SIC errors over fading channels," *IET Commun.*, vol. 12, pp. 1834–1844(10), Sep. 2018.

[36] M. Jain, S. Soni, N. Sharma, and D. Rawal, "Performance analysis at far and near user in NOMA based system in presence of SIC error," *Int. J. Electron.*

Commun. (AEU), vol. 114, p. 152993, Feb. 2020.

[37] A. Hilario-Tacuri, J. Maldonado, M. Revollo, and H. Chambi, "Bit error rate analysis of NOMA-OFDM in 5G systems with non-linear HPA with memory," *IEEE Access*, vol. 9, pp. 83709–83717, Jun. 2021.

[38] K. Chung, "Impact of channel estimation errors on SIC performance of NOMA in 5G systems," *J. Convergence Inf. Technol.*, vol. 10, pp. 22–27, Sep. 2020.

[39] K. Chung, "Optimal detection for NOMA systems with correlated information sources of interactive mobile users," *J. Korea Inst. Electron. Commun. Sci.*, vol. 15, pp. 651–658, Aug. 2020.

[40] K. Chung, "On negative correlation bit-to-symbol(: B2S) mapping for NOMA with correlated information sources in 5G systems," *J. Korea Inst. Electron. Commun. Sci.*, vol. 15, pp. 881–888, Oct. 2020.

[41] N. Jaiswal and N. Purohit, "Performance analysis of NOMA-enabled vehicular communication systems with transmit antenna selection over double Nakagami-*m* fading," *IEEE Trans. Veh. Technol.*, vol. 70, pp. 12725–12741, Dec. 2021.

[42] R. Makkar, V. Kotha, D. Rawal, V. K. Chakka, and N. Sharma, "Performance analysis of OSTBC in NOMA assisted downlink system with SIC errors," in *IEEE Veh. Technol. Conf. (VTC)*, (London, UK), pp. 1–5, Sep. 2022.

[43] L. Bariah, S. Muhaidat, and A. Al-Dweik, "Error probability analysis of non-orthogonal multiple access over Nakagami-*m* fading channels," *IEEE Trans. Commun.*, vol. 67, pp. 1586–1599, Feb. 2019.

[44] L. Bariah, B. Selim, L. Mohjazi, S. Muhaidat, P. C. Sofotasios, and W. Hamouda, "Pairwise error probability of non-orthogonal multiple access with I/Q imbalance," in *UK/ China Emerging Technol. (UCET)*, (Glasgow, UK), pp. 1–4, Aug. 2019.

[45] A. Abushattal, S. Althunibat, M. Qaraqe, and H. Arslan, "A secure downlink NOMA scheme against unknown internal eavesdroppers," *IEEE Wireless Commun. Lett.*, vol. 10, pp. 1281–1285, Jun. 2021.

[46] T. Alshamaseen, S. Althunibat, M. Qaraqe, and H. Alashaary, "Phase-assisted NOMA based key distribution for IoT networks," *Trans. Emerg. Telecommun. Technol.*, p. e4738, 2023.

[47] A. M. A. Alkhazzar and H. Aghaeinia, "Symbol error rate analysis of non-orthogonal multiple access systems," *Phys. Commun.*, vol. 46, p. 101295, Feb. 2021.

[48] T. Assaf, A. Al-Dweik, M. E. Moursi, and H. Zeineldin, "Exact BER performance analysis for downlink NOMA systems over Nakagami-m fading channels," *IEEE Access*, vol. 7, pp. 134539–134555, Sep. 2019.

[49] J.-R. Garnier, A. Fabre, H. Farès, and R. Bonnefoi, "On the performance of QPSK modulation over downlink NOMA: From error probability derivation to SDR-based validation," *IEEE Access*, vol. 8, pp. 66495–66507, Mar. 2020.

[50] M. Aldababsa, C. Göztepe, G. K. Kurt, and O. Kucur, "Bit error rate for NOMA network," *IEEE Commun. Lett.*, vol. 24, pp. 1188–1191, Jun. 2020.

[51] A. Almohamad, S. Althunibat, M. Hasna, and K. Qaraqe, "On the error performance of non-orthogonal multiple access systems," in *Int. Conf. Inf. Commun. Technol. Convergence (ICTC)*, (Jeju Island, Korea), pp. 116–121, Oct. 2020.

[52] N. Jaiswal and N. Purohit, "Performance of downlink NOMA-enabled vehicular communications over double Rayleigh fading channels," *IET Commun.*, vol. 14, pp. 3652–3660, Sep. 2020.

[53] S. Mukhtar and G. R. Begh, "On the performance analysis of downlink NOMA system over non-homogenous fading channel," in *IEEE Int. Conf. Comput. Power Commun. Technol. (GUCON)*, (Kuala Lumpur, Malaysia), pp. 1–6, Nov. 2021.

[54] A. S. Alqahtani and E. Alsusa, "Performance analysis of downlink NOMA system over α-η-μ generalized fading channel," in *IEEE Veh. Technol. Conf. (VTC)*, (Antwerp, Belgium), pp. 1–5, May 2020.

[55] B. M. ElHalawany, F. Jameel, D. B. da Costa, U. S. Dias, and K. Wu, "Performance analysis of downlink NOMA systems over κ-μ shadowed fading channels," *IEEE Trans Veh. Technol.*, vol. 69, pp. 1046–1050, Jan. 2020.

[56] X. Yan, H. Xiao, C. Wang, K. An, A. T. Chronopoulos, and G. Zheng, "Performance analysis of NOMA-based land mobile satellite networks," *IEEE Access*, vol. 6, pp. 31327–31339, Jun. 2018.

[57] A. Alqahtani, E. Alsusa, A. Al-Dweik, and M. Al-Jarrah, "Performance analysis for Downlink NOMA over α-μ generalized fading channels," *IEEE Trans. Veh. Technol.*, vol. 70, pp. 6814–6825, Jul. 2021.

[58] P. Kumar and K. Dhaka, "Average BER analysis of NOMA systems under TWDP fading," in *National Conf. Commun. (NCC)*, (Guwahati, India), pp. 1–6, Feb. 2023.

[59] L. Bariah, S. Muhaidat, P. C. Sofotasios, S. Gurugopinath, W. Hamouda, and H. Yanikomeroglu, "Non-orthogonal multiple access in the presence of additive generalized Gaussian noise," *IEEE Commun. Lett.*, vol. 24, pp. 2137–2141, Oct.

2020.

[60] K. Aslan and T. Gucluoglu, "Performance analysis of downlink-NOMA over generalized-k fading channels," *Phys. Commun.*, vol. 48, p. 101414, Jul. 2021.

[61] X. Wang, F. Labeau, and L. Mei, "Closed-form BER expressions of QPSK constellation for uplink non-orthogonal multiple access," *IEEE Commun. Lett.*, vol. 21, pp. 2242–2245, Oct. 2017.

[62] X. Chen, H. Yin, and G. Wei, "Impact of frequency offset on the performance of uplink OFDM-NOMA systems," in *IEEE Int. Conf. Comput. Commun. (ICCC)*, (Chengdu, China), pp. 168–173, Dec. 2018.

[63] F. Wei, T. Zhou, T. Xu, and H. Hu, "BER analysis for uplink NOMA in asymmetric channels," *IEEE Commun. Lett.*, vol. 24, pp. 2435–2439, Nov. 2020.

[64] R. P. Sirigina, A. S. Madhukumar, and M. Bowyer, "BER analysis of interference-limited synchronous wireless networks with line-of-sight links," *IEEE Commun. Lett.*, vol. 23, pp. 542–545, Mar. 2019.

[65] J. S. Yeom, H. S. Jang, K. S. Ko, and B. C. Jung, "BER performance of uplink NOMA with joint maximum-likelihood detector," *IEEE Trans. Veh. Technol.*, vol. 68, pp. 10295–10300, Oct. 2019.

[66] C. Liu and N. C. Beaulieu, "Exact SER and BER analysis on jointly optimal maximum likelihood detection of two QPSK signals with phase offset," *IEEE Trans. Veh. Technol.*, vol. 70, pp. 11695–11709, Nov. 2021.

[67] P. Shachi, T. Sanjana, K. R. Sudhindra, and M. N. Suma, "Outage and error-rate analysis of NOMA-inspired multicarrier underlay cognitive radio system," *IEEE Access*, vol. 10, pp. 113956–113968, Nov. 2022.

[68] H. Semira, F. Kara, H. Kaya, and H. Yanikomeroglu, "Multi-user joint maximum-likelihood detection in uplink NOMA-IoT networks: Removing the error floor," *IEEE Wireless Commun. Lett.*, vol. 10, pp. 2459–2463, Nov. 2021.

[69] H. Semira, F. Kara, H. Kaya, and H. Yanikomeroglu, "Error performance analysis of multiuser detection in uplink-NOMA with adaptive M-QAM," *IEEE Wireless Commun. Lett.*, vol. 11, pp. 1654–1658, Aug. 2022.

[70] E. C. Cejudo, H. Zhu, and J. Wang, "Resource allocation in multicarrier NOMA systems based on optimal channel gain ratios," *IEEE Trans. Wireless Commun.*, vol. 21, pp. 635–650, Jan. 2022.

[71] T. Assaf, A. Al-Dweik, M. S. E. Moursi, and H. Zeineldin, "Efficient bit loading algorithm for OFDM-NOMA systems with BER constraints," *IEEE Trans. Veh. Technol.*, vol. 71, pp. 423–436, Nov. 2022.

[72] H. Yahya, E. A. Alsusa, and A. Al-Dweik, "Min-Max design and analysis of

NOMA with adaptive modulation under BLER constraints," in *IEEE Veh. Technol. Conf. (VTC)*, (London, UK), pp. 1–5, Sep. 2022.

[73] W. Yu, H. Jia, and L. Musavian, "Joint adaptive M-QAM modulation and power adaptation for a downlink NOMA network," *IEEE Trans. Commun.*, vol. 70, pp. 783–796, Feb. 2022.

[74] K. Wang, T. Zhou, T. Xu, H. Hu, and X. Tao, "Asymmetric adaptive modulation for uplink NOMA systems," *IEEE Trans. Commun.*, vol. 69, pp. 7222–7235, Nov. 2021.

[75] Q. He, Y. Hu, and A. Schmeink, "Closed-form symbol error rate expressions for non-orthogonal multiple access systems," *IEEE Trans. Veh. Technol.*, vol. 68, pp. 6775–6789, Jul. 2019.

[76] Y. Wang, J. Wang, L. Ma, Y. Huang, and C. Zhao, "Minimum error performance of downlink non-orthogonal multiple access systems," in *IEEE Veh. Technol. Conf. (VTC)*, (Honolulu, HI, USA), pp. 1–5, Sep. 2019.

[77] Y. Wang, J. Wang, D. W. Kwan Ng, R. Schober, and X. Gao, "A minimum error probability NOMA design," *IEEE Trans. Wireless Commun.*, vol. 20, pp. 4221–4237, Jul. 2021.

[78] A. K. Dutta, "MBER criterion assisted power NOMA design and performance analysis with estimated channel," *IEEE Trans. Veh. Technol.*, vol. 68, pp. 11816–11826, Dec. 2019.

[79] Z. Dong, H. Chen, J. K. Zhang, L. Huang, and B. Vucetic, "Uplink non-orthogonal multiple access with finite-alphabet inputs," *IEEE Trans. Wireless Commun.*, vol. 17, pp. 5743–5758, Sep. 2018.

[80] C. Wang, Y. Wang, and P. Xiao, "Power allocation based on SINR balancing for NOMA systems with imperfect channel estimation," in *Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, (Gold Coast, Australia), pp. 1–6, Dec. 2019.

[81] 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, pp. 2192–2235, Fourth Quart. 2020.

[82] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, pp. 2294–2323, Third Quart. 2018.

[83] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, pp. 2347–2381, Dec. 2017.

[84] M. Shirvanimoghaddam *et al.*, "Massive non-orthogonal multiple access for cellular IoT: Potentials and limitations," *IEEE Commun. Mag.*, vol. 55, pp. 55–61, Sep. 2017.

[85] Z. Wei, L. Yang, D. W. K. Ng, J. Yuan, and L. Hanzo, "On the performance gain of NOMA over OMA in uplink communication systems," *IEEE Trans. Commun.*, vol. 68, pp. 536–568, Jan. 2020.

[86] Z. Xiang, W. Yang, Y. Cai, Z. Ding, Y. Song, and Y. Zou, "NOMA-assisted secure short-packet communications in IoT," *IEEE Wireless Commun.*, vol. 27, pp. 8–15, Aug. 2020.

[87] N. I. Miridakis and D. D. Vergados, "A survey on the successive interference cancellation performance for single-antenna and multiple-antenna OFDM systems," *IEEE Commun. Surveys Tuts.*, vol. 15, pp. 312–335, 1st Quart. 2013.

[88] M. Aldababsa, A. Khaleel, and E. Basar, "STAR-RIS-NOMA networks: An error performance perspective," *IEEE Commun. Lett.*, vol. 26, pp. 1784–1788, Aug. 2022.

[89] M. Qiu, Y. C. Huang, S. L. Shieh, and J. Yuan, "A lattice-partition framework of downlink non-orthogonal multiple access without SIC," *IEEE Trans. Commun.*, vol. 66, pp. 2532–2546, Jun. 2018.

[90] E. M. Almohimmah and M. T. Alresheedi, "Error analysis of NOMA-based VLC systems with higher order modulation schemes," *IEEE Access*, vol. 8, pp. 2792–2803, Dec. 2020.

[91] H. Xu and N. Pillay, "Threshold based signal detection and the average symbol error probability for downlink NOMA systems with M-ary QAM," *IEEE Access*, vol. 8, pp. 156677–156685, Aug. 2020.

[92] X. Liu, Y. Wang, F. Zhou, and R. Q. Hu, "BER analysis for NOMA-enabled visible light communication systems with M-PSK," in *Int. Conf. Wireless Commun. Signal Process. (WCSP)*, (Hangzhou, China), pp. 1–7, Oct. 2018.

[93] X. Liu, Z. Chen, Y. Wang, F. Zhou, Y. Luo, and R. Q. Hu, "BER analysis of NOMA-enabled visible light communication systems with different modulations," *IEEE Trans. Veh. Technol.*, vol. 68, pp. 10807–10821, Nov. 2019.

[94] F. Kara and H. Kaya, "Performance analysis of SSK-NOMA," *IEEE Trans. Veh. Technol.*, vol. 68, pp. 6231–6242, Jul. 2019.

[95] H. Yahya, A. Al-Dweik, and E. Alsusa, "Power-tolerant NOMA using data-aware adaptive power assignment for IoT systems," *IEEE Internet Things J.*, vol. 8, pp. 14896–14907, Oct. 2021.

[96] H. Marshoud, P. C. Sofotasios, S. Muhaidat, G. K. Karagiannidis, and B. S.

Sharif, "On the performance of visible light communication systems with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 16, pp. 6350–6364, Oct. 2017.

[97] J. Zhang, X. Wang, T. Hasegawa, and T. Kubo, "Downlink non-orthogonal multiple access (NOMA) constellation rotation," in *IEEE Veh. Technol. Conf. (VTC)*, (Montreal, QC, Canada), pp. 1–5, Sep. 2016.

[98] S. Shieh, C. Lin, Y. Huang, and C. Wang, "On Gray labeling for downlink non-orthogonal multiple access without SIC," *IEEE Commun. Lett.*, vol. 20, pp. 1721–1724, Sep. 2016.

[99] A. J. Al-Dweik, M. Mayhew, R. Muresan, S. M. Ali, and A. Shami, "Using technology to make roads safer: Adaptive speed limits for an intelligent transportation system," *IEEE Veh. Technol. Mag.*, vol. 12, pp. 39–47, Mar. 2017.

[100] A. Al-Dweik, R. Muresan, M. Mayhew, and M. Lieberman, "IoT-based multi-functional scalable real-time enhanced road side unit for intelligent transportation systems," in *IEEE Canadian Conf. Electrical Computer Engineering (IEEE CCECE)*, (Windsor, ON, Canada), pp. 1–6, Mar. 2017.

[101] L. Yang, R. Muresan, A. Al-Dweik, and L. J. Hadjileontiadis, "Image-based visibility estimation algorithm for intelligent transportation systems," *IEEE Access*, vol. 6, pp. 76728–76740, Nov. 2018.

[102] A. Hoglund, X. Lin, O. Liberg, A. Behravan, E. A. Yavuz, M. Van Der Zee, Y. Sui, T. Tirronen, A. Ratilainen, and D. Eriksson, "Overview of 3GPP release 14 enhanced NB-IoT," *IEEE Netw.*, vol. 31, pp. 16–22, Nov./Dec. 2017.

[103] M. A. Al-Jarrah, M. A. Yaseen, A. Al-Dweik, O. A. Dobre, and E. Alsusa, "Decision fusion for IoT-based wireless sensor networks," *IEEE Internet Things J.*, vol. 7, pp. 1313–1326, Feb. 2020.

[104] R. Zhang and L. Hanzo, "A unified treatment of superposition coding aided communications: Theory and practice," *IEEE Commun. Surveys Tuts.*, vol. 13, pp. 503–520, 3rd Quart. 2011.

[105] B. Selim, S. Muhaidat, P. C. Sofotasios, A. Al-Dweik, B. S. Sharif, and T. Stouraitis, "Radio-frequency front-end impairments: Performance degradation in nonorthogonal multiple access communication systems," *IEEE Veh. Technol. Mag.*, vol. 14, pp. 89–97, Mar. 2019.

[106] X. Wan, X. Zhu, Y. Jiang, Y. Liu, and J. Zhao, "An interference alignment and ICA-based semiblind dual-user downlink NOMA system for high-reliability low-latency IoT," *IEEE Internet Things J.*, vol. 7, pp. 10837–10851, Apr. 2020.

[107] J. Chu, X. Chen, C. Zhong, and Z. Zhang, "Robust design for NOMA-based

multibeam LEO satellite Internet of Things," *IEEE Internet Things J.*, vol. 8, pp. 1959–1970, Feb. 2021.

[108] Y. Xu, R. Q. Hu, and G. Li, "Robust energy-efficient maximization for cognitive NOMA networks under channel uncertainties," *IEEE Internet Things J.*, vol. 7, pp. 8318–8330, Sep. 2020.

[109] Q. Qi, X. Chen, and D. W. K. Ng, "Robust beamforming for NOMA-based cellular massive IoT with SWIPT," *IEEE Trans. Signal Process.*, vol. 68, pp. 211–224, Dec. 2019.

[110] M. Qiu, Y. C. Huang, and J. Yuan, "Downlink non-orthogonal multiple access without SIC for block fading channels: An algebraic rotation approach," *IEEE Trans. Wireless Commun.*, vol. 18, pp. 3903–3918, Aug. 2019.

[111] *GPP TS 36.211 V10.5.0, Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation, (Release 10)*, June 2012.

[112] E. Alsusa and C. Masouros, "Adaptive code allocation for interference management on the downlink of DS-CDMA systems," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2420–2424, Jul. 2008.

[113] C. Masouros and E. Alsusa, "Dynamic linear precoding for the exploitation of known interference in MIMO broadcast systems," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1396–1404, Mar. 2009.

[114] T. M. Cover and J. A. Thomas, *Elements of information theory*. Hoboken, New Jersey, USA: Wiley-Interscience, 2 ed., 2006.

[115] T. F. Wong, "Numerical calculation of symmetric capacity of rayleigh fading channel with BPSK/QPSK," *IEEE Commun. Lett.*, vol. 5, pp. 328–330, Aug. 2001.

[116] E. Baccarelli and A. Fasano, "Some simple bounds on the symmetric capacity and outage probability for QAM wireless channels with Rice and Nakagami fadings," *IEEE J. Sel. Areas Commun.*, vol. 18, pp. 361–368, Mar. 2000.

[117] Q. Yu, H. Chen, and W. Meng, "A unified multiuser coding framework for multiple-access technologies," *IEEE Syst. J.*, vol. 13, pp. 3781–3792, Dec. 2019.

[118] L. Zhang *et al.*, "Layered-division-multiplexing: Theory and practice," *IEEE Trans. Broadcast.*, vol. 62, pp. 216–232, Mar. 2016.

[119] Z. Ding *et al.*, "A state-of-the-art survey on reconfigurable intelligent surface-assisted non-orthogonal multiple access networks," *Proc. IEEE*, pp. 1–22, 2022.

[120] H. Yahya, E. Alsusa, A. Al-Dweik, *et al.*, "Error rate analysis of NOMA: Principles, survey and future directions," 2022.

[121] S. Ahn and D. Yoon, "Hierarchical circular θ-QAM," *IEEE Trans. Broadcast.*, vol. 67, pp. 343–350, Mar. 2021.

[122] L. Zhang *et al.*, "Using layered-division-multiplexing to deliver multi-layer mobile services in ATSC 3.0," *IEEE Trans. Broadcast.*, vol. 65, pp. 40–52, Mar. 2019.

[123] Y. Yuan, Z. Yuan, and L. Tian, "5G non-orthogonal multiple access study in 3GPP," *IEEE Commun. Mag.*, vol. 58, pp. 90–96, Jul. 2020.

[124] H. Yahya, A. Al-Dweik, Y. Iraqi, E. Alsusa, and A. Ahmed, "A power and spectrum efficient uplink transmission scheme for QoS-constrained IoT networks," *IEEE Internet Things J.*, vol. 9, pp. 17425–17439, Sep. 2022.

[125] J. Yao, Q. Zhang, and J. Qin, "Injection optimization for uniform QAM-based non-orthogonal multiplexing systems," *IEEE Trans. Broadcast.*, vol. 68, pp. 927–934, Dec. 2022.

[126] A. Al-Dweik, Y. Iraqi, K.-H. Park, M. Al-Jarrah, E. Alsusa, and M.-S. Alouini, "Efficient NOMA design without channel phase information using amplitude-coherent detection," *IEEE Trans. Commun.*, vol. 70, pp. 245–263, Jan. 2022.

[127] C. Liu, N. C. Beaulieu, J. Cheng, S. Wu, C. Jiang, and H. Yang, "An analysis of the error rate performance for uplink asynchronous signal detection in non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 70, pp. 4591–4606, Jul. 2022.

[128] X. Yue and Y. Liu, "Performance analysis of intelligent reflecting surface assisted NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 21, pp. 2623–2636, Apr. 2022.

[129] K.-H. Lee, J. S. Yeom, J. Joung, and B. C. Jung, "Performance analysis of uplink NOMA with constellation–rotated STLC for IoT networks," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 705–717, Apr. 2022.

[130] N. Ye, A. Wang, X. Li, W. Liu, X. Hou, and H. Yu, "On constellation rotation of NOMA with SIC receiver," *IEEE Commun. Lett.*, vol. 22, pp. 514–517, Mar. 2018.

[131] C. Yan, A. Harada, A. Benjebbour, Y. Lan, A. Li, and H. Jiang, "Receiver design for downlink non-orthogonal multiple access (NOMA)," in *IEEE Veh. Technol. Conf. (VTC)*, Jul. 2015.

[132] J. Yeom, E. Chu, B. Jung, and H. Jin, "Performance analysis of diversity-controlled multi-user superposition transmission for 5G wireless networks," *Sensors*, vol. 18, p. 536, Feb. 2018.

[133] W. Han, Y. Wang, D. Tang, R. Hou, and X. Ma, "Study of the symbol error rate/bit error rate in non-orthogonal multiple access (NOMA) systems," in *IEEE/CIC Int. Conf. Commun. in China (ICCC Workshops)*, (Beijing, China), pp. 101–106, Aug. 2018.

[134] W. Han, X. Ma, D. Tang, and N. Zhao, "Study of SER and BER in NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 70, pp. 3325–3340, May 2021.

[135] Q. Gao, M. Jia, Q. Guo, and X. Gu, "On the analysis and optimization of BER performance of symmetrical coding based NOMA," in *IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, pp. 1–6, May 2021.

[136] M. Mamatov and J. Nuritdinov, "Some properties of the sum and geometric differences of Minkowski," *J. Appl. Mathematics Physics*, vol. 8, no. 10, pp. 2241–2255, 2020.

[137] S.-L. Shieh and Y.-C. Huang, "A simple scheme for realizing the promised gains of downlink nonorthogonal multiple access," *IEEE Trans. Commun.*, vol. 64, pp. 1624–1635, Apr. 2016.

[138] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, pp. 1074–1080, Jul. 2002.

[139] H. Yahya, E. Alsusa, and A. Al-Dweik, "Joint-multiuser Gray-labeling for NOMA." `https://codeocean.com/capsule/8158634/tree/v1`, Nov. 2022.

[140] X. Pang, J. Tang, N. Zhao, X. Zhang, and Y. Qian, "Energy-efficient design for mmWave-enabled NOMA-UAV networks," *Science China Information Sciences*, vol. 64, no. 4, pp. 1–14, 2021. Art. no. 140303.

[141] N. Zhao, Y. Li, S. Zhang, Y. Chen, W. Lu, J. Wang, and X. Wang, "Security enhancement for NOMA-UAV networks," *IEEE Trans. Veh. Technol.*, vol. 69, pp. 3994–4005, Apr. 2020.

[142] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys Tuts.*, vol. 11, pp. 116–130, 1st Quart. 2009.

[143] M. Vaezi, G. A. Aruma Baduge, Y. Liu, A. Arafa, F. Fang, and Z. Ding, "Interplay between NOMA and other emerging technologies: A survey," *IEEE Trans. Cognitive Commun. Netw.*, vol. 5, pp. 900–919, Dec. 2019.

[144] F. Zhou *et al.*, "State of the art, taxonomy, and open issues on cognitive radio networks with NOMA," *IEEE Wireless Commun.*, vol. 25, pp. 100–108, Apr. 2018.

[145] V. Singh, S. Solanki, P. Upadhyay, D. Costa, and J. Moualeu, "Performance

analysis of hardware-impaired overlay cognitive satellite–terrestrial networks with adaptive relaying protocol," *IEEE Syst. J.*, vol. 15, pp. 192–203, Mar. 2021.

[146] S. Solanki, P. Upadhyay, D. Costa, P. Bithas, and A. Kanatas, "Performance analysis of cognitive relay networks with RF hardware impairments and CEEs in the presence of primary users' interference," *IEEE Trans. Cognitive Commun. Netw.*, vol. 4, no. 2, pp. 406–421, 2018.

[147] L. Lv *et al.*, "Cognitive non-orthogonal multiple access with cooperative relaying: A new wireless frontier for 5G spectrum sharing," *IEEE Commun. Mag.*, vol. 56, pp. 188–195, Apr. 2018.

[148] C. Singh and P. Upadhyay, "Overlay cognitive IoT-based full-duplex relaying NOMA systems with hardware imperfections," *IEEE Internet Things J.*, vol. 9, pp. 6578–6596, Sep. 2021.

[149] V. Singh and P. Upadhyay, "Exploiting FD/HD cooperative-NOMA in underlay cognitive hybrid satellite-terrestrial networks," *IEEE Trans. Cognitive Commun. Networking*, vol. 8, pp. 246–262, Jan. 2021.

[150] A. Shukla, V. Singh, P. Upadhyay, A. Kumar, and J. Moualeu, "Performance analysis of energy harvesting-assisted overlay cognitive NOMA systems with incremental relaying," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1558–1576, 2021.

[151] Y. Liu, Z. Ding, M. Elkashlan, and J. Yuan, "Nonorthogonal multiple access in large-scale underlay cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 65, pp. 10152–10157, Dec. 2016.

[152] S. Arzykulov *et al.*, "On the outage of underlay CR-NOMA networks with detect-and-forward relaying," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, pp. 795–804, Sep. 2019.

[153] L. Luo, Q. Li, and J. Cheng, "Performance analysis of overlay cognitive NOMA systems with imperfect successive interference cancellation," *IEEE Trans. Commun.*, vol. 68, pp. 4709–4722, May 2020.

[154] A. Bhowmick, S. Dhar Roy, and S. Kundu, "Throughput maximization of a NOMA-based energy-harvesting UAV assisted CR network," *IEEE Trans. Veh. Technol.*, vol. 71, pp. 362–374, Jan. 2022.

[155] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, pp. 6010–6023, Aug. 2016.

[156] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems," *IEEE*

Trans. Wireless Commun., vol. 15, pp. 7244–7257, Nov. 2016.

[157] Z. Ding, L. Dai, and H. V. Poor, "MIMO-NOMA design for small packet transmission in the Internet of Things," *IEEE Access*, vol. 4, pp. 1393–1405, Apr. 2016.

[158] Z. Yang *et al.*, "Power allocation study for non-orthogonal multiple access networks with multicast-unicast transmission," *IEEE Trans. Wireless Commun.*, vol. 17, pp. 3588–3599, Jun. 2018.

[159] L. Lv, J. Chen, Q. Ni, and Z. Ding, "Design of cooperative non-orthogonal multicast cognitive multiple access for 5G systems: User scheduling and performance analysis," *IEEE Trans. Commun.*, vol. 65, pp. 2641–2656, Jun. 2017.

[160] Z. Yang, Z. Ding, P. Fan, and Z. Ma, "Outage performance for dynamic power allocation in hybrid non-orthogonal multiple access systems," *IEEE Commun. Lett.*, vol. 20, pp. 1695–1698, Aug. 2016.

[161] Z. Chen, Z. Ding, X. Dai, and G. K. Karagiannidis, "On the application of quasi-degradation to MISO-NOMA downlink," *IEEE Trans. Signal Process.*, vol. 64, pp. 6174–6189, Dec. 2016.

[162] J. Shi *et al.*, "Energy efficient resource allocation in hybrid non-orthogonal multiple access systems," *IEEE Trans. Commun.*, vol. 67, pp. 3496–3511, May 2019.

[163] H. Abumarshoud, H. Alshaer, and H. Haas, "Dynamic multiple access configuration in intelligent LiFi attocellular access points," *IEEE Access*, vol. 7, pp. 62126–62141, May 2019.

[164] N. Nomikos *et al.*, "Hybrid NOMA/OMA with buffer-aided relay selection in cooperative networks," *IEEE J. Sel. Topics Signal Process.*, vol. 13, pp. 524–537, Jun. 2019.

[165] S. Luo and K. C. Teh, "Adaptive transmission for cooperative NOMA system with buffer-aided relaying," *IEEE Commun. Lett.*, vol. 21, pp. 937–940, Apr. 2017.

[166] K. Janghel and S. Prakriya, "Performance of adaptive OMA/cooperative-NOMA scheme with user selection," *IEEE Commun. Lett.*, vol. 22, pp. 2092–2095, Oct. 2018.

[167] P. Xu *et al.*, "Performance analysis of buffer-aided hybrid NOMA/OMA in cooperative uplink system," *IEEE Access*, vol. 7, pp. 168759–168773, Nov. 2019.

[168] M. Choi2019-TVT, D. Yoon, and J. Kim, "Blind signal classification for non-orthogonal multiple access in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 68, pp. 9722–9734, Oct. 2019.

[169] M. Choi and J. Kim, "Blind signal classification analysis and impact on user pairing and power allocation in nonorthogonal multiple access," *IEEE Access*, vol. 8, pp. 100916–100929, Jun. 2020.

[170] M. Naderpour and H. K. Bizaki, "Low overhead NOMA receiver with automatic modulation classification techniques," *IET Commun.*, vol. 14, pp. 768–774, Apr. 2020.

[171] T. Li, Y. Li, and O. A. Dobre, "Modulation classification based on fourth-order cumulants of superposed signal in NOMA systems," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 2885–2897, 2021.

[172] N. Zhang, K. Cheng, and G. Kang, "A machine-learning-based blind detection on interference modulation order in NOMA systems," *IEEE Commun. Lett.*, vol. 22, pp. 2463–2466, Dec. 2018.

[173] H. Yahya, E. Alsusa, and A. Al-Dweik, "Blind receiver design for downlink Cognitive-Radio NOMA networks," in *IEEE Int. Commun. Conf. (ICC) Workshops*, (Montreal, Canada), pp. 1–6, Jul. 2021.

[174] H. Yahya, A. Al-Dweik, E. Alsusa, and M. Debbah, "Priority-based dynamic IoT-downlink communications with blind transmission-mode recognition," *IEEE Internet Things Mag.*, vol. 5, pp. 106–112, Sep. 2022.

[175] W. Liang, Z. Ding, Y. Li, and L. Song, "User pairing for downlink non-orthogonal multiple access networks using matching algorithm," *IEEE Trans. Commun.*, vol. 65, pp. 5319–5332, Dec. 2017.

[176] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "On the sum rate of mimo-noma and mimo-oma systems," *IEEE Wireless Commun. Lett.*, vol. 6, pp. 534–537, Apr. 2017.

[177] A. Saci, A. Al-Dweik, A. Shami, and Y. Iraqi, "One-shot blind channel estimation for OFDM systems over frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 65, pp. 5445–5458, Dec. 2017.

[178] V. Dao, L. Hoang, S. Girs, and E. Uhlemann, "Outage performance of pairwise NOMA allowing a dynamic decoding order and optimal pairs of power levels," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1886–1906, 2020.

[179] J. Craig, "A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations," in *MILCOM 91 - Conference record*, (McLean, VA, USA), pp. 571–575 vol.2, Nov. 1991.

[180] H. Yahya, E. Alsusa, and A. Al-Dweik, "Hybrid cognitive-radio NOMA with blind transmission mode identification and BER constraints," *TechRxiv*, Aug. 2021.

[181] D. Sadhwani, A. Powari, and N. Mehta, "New, simple and accurate approximation for the Gaussian Q function with applications," *IEEE Commun. Lett.*, vol. 26, pp. 518–522, Mar. 2022.

[182] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*. Academic press, 2014.

[183] M. Bóna, *A walk through combinatorics: an introduction to enumeration and graph theory*. MA, USA: World Scientific Pub. Co. Inc., 2nd ed., 2006.

[184] D.-T. Do, A.-T. Le, C.-B. Le, and B. M. Lee, "On exact outage and throughput performance of cognitive radio based non-orthogonal multiple access networks with and without D2D link," *Sensors*, vol. 19, Jul. 2019.

[185] T. M. Hoang, L. T. Dung, B. C. Nguyen, and T. Kim, "Outage and throughput analysis of UAV-assisted NOMA relay systems with indoor and outdoor users," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, pp. 2633–2647, Jun 2023.

[186] M. Tariq *et al.*, "Computational power evaluation for energy-constrained wireless communications systems," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 308–319, 2020.

[187] H. Y. Hsieh, M. J. Yang, and C. H. Wang, "Fair resource allocation using the MCS map for multi-user superposition transmission (MUST)," in *IEEE PIMRC*, (Spain), pp. 1–7, Dec. 2016.

[188] S. Kim, H. Kim, and D. Hong, "Joint power allocation and MCS selection in downlink NOMA system," in *IEEE PIMRC*, (Italy), pp. 1–4, Dec. 2018.

[189] H. Jia and L. Musavian, "Performance analysis for NOMA with M-QAM modulation," in *IEEE VTC*, (Belgium), pp. 1–5, Jun. 2020.

[190] Y. Wang *et al.*, "Effective throughput maximization of NOMA with practical modulations," *IEEE J. Sel. Areas Commun.*, vol. 40, pp. 1084–1100, Apr. 2022.

[191] J. Choi, "On the power allocation for a practical multiuser superposition scheme in NOMA systems," *IEEE Commun. Lett.*, vol. 20, pp. 438–441, Mar. 2016.

[192] M. AbdelMoniem *et al.*, "Enhanced NOMA system using adaptive coding and modulation based on LSTM neural network channel estimation," *Applied Sciences*, vol. 9, Jul. 2019.

[193] Z. Zheng, L. Yuan, and F. Fang, "Performance analysis of fountain coded non-orthogonal multiple access with finite blocklength," *IEEE Wireless Commun. Lett.*, vol. 10, pp. 1752–1756, Aug. 2021.

[194] M. Le-Tran, T.-H. Vu, and S. Kim, "Performance analysis of optical backhauled cooperative NOMA visible light communication," *IEEE Trans. Veh. Technol.*,

vol. 70, pp. 12932–12945, Dec. 2021.

[195] T.-H. Vu, T.-V. Nguyen, T.-T. Nguyen, V. N. Q. Bao, and S. Kim, "Short-packet communications in NOMA-CDRT IoT networks with cochannel interference and imperfect SIC," *IEEE Trans. Veh. Technol.*, vol. 71, pp. 5552–5557, May 2022.

[196] T.-H. Vu, T.-V. Nguyen, T.-T. Nguyen, and S. Kim, "Performance analysis and deep learning design of wireless powered cognitive NOMA IoT short-packet communications with imperfect CSI and SIC," *IEEE Internet Things J.*, vol. 9, pp. 10464–10479, Jul. 2022.

[197] T.-H. Vu, T.-V. Nguyen, D. B. d. Costa, and S. Kim, "Intelligent reflecting surface-aided short-packet non-orthogonal multiple access systems," *IEEE Trans. Veh. Technol.*, vol. 71, pp. 4500–4505, Apr. 2022.

[198] N. P. Le and K. N. Le, "Uplink NOMA short-packet communications with residual hardware impairments and channel estimation errors," *IEEE Trans. Veh. Technol.*, vol. 71, pp. 4057–4072, Apr. 2022.

[199] B. Sklar, *Digital Communications: Fundamentals and Applications*. Novell-Bibliothek, Prentice-Hall PTR, 2001.

[200] H. Yahya, E. Alsusa, and A. Al-Dweik, "Exact BER analysis of NOMA with arbitrary number of users and modulation orders," 2021.

[201] G. K. Karagiannidis and A. S. Lioumpas, "An improved approximation for the Gaussian Q-function," *IEEE Commun. Lett.*, vol. 11, pp. 644–646, Aug. 2007.

[202] B. Zhang, K. Wang, K. Yang, and G. Zhang, "IRS-assisted short packet wireless energy transfer and communications," *IEEE Wireless Commun. Lett.*, vol. 11, pp. 303–307, Feb. 2022.

[203] A. Ahmed, M. Naeem, and A. Al-Dweik, "Joint optimization of sensors association and UAVs placement in IoT applications with practical network constraints," *IEEE Access*, vol. 9, pp. 7674–7689, Jan. 2021.

[204] N. Hossein Motlagh, T. Taleb, and O. Arouk, "Low-altitude unmanned aerial vehicles-based internet of things services: Comprehensive survey and future perspectives," *IEEE Internet Things J.*, vol. 3, pp. 899–922, Dec. 2016.

[205] S. Zhang, J. Liu, and W. Sun, "Stochastic geometric analysis of multiple unmanned aerial vehicle-assisted communications over internet of things," *IEEE Internet Things J.*, vol. 6, pp. 5446–5460, Jun. 2019.

[206] Z. Lin *et al.*, "Supporting IoT with rate-splitting multiple access in satellite and aerial-integrated networks," *IEEE Internet Things J.*, vol. 8, pp. 11123–11134, Jul. 2021.

[207] J. Contreras-Castillo, S. Zeadally, and J. A. Guerrero-Ibañez, "Internet of vehicles: Architecture, protocols, and security," *IEEE Internet Things J.*, vol. 5, pp. 3701–3709, Oct. 2018.

[208] L.-L. Wang, J.-S. Gui, X.-H. Deng, F. Zeng, and Z.-F. Kuang, "Routing algorithm based on vehicle position analysis for internet of vehicles," *IEEE Internet Things J.*, vol. 7, pp. 11701–11712, Dec. 2020.

[209] Y. Ni, J. He, L. Cai, J. Pan, and Y. Bo, "Joint roadside unit deployment and service task assignment for internet of vehicles (IoV)," *IEEE Internet of Things J.*, vol. 6, pp. 3271–3283, Apr. 2019.

[210] K. Z. Ghafoor, L. Kong, D. B. Rawat, E. Hosseini, and A. S. Sadiq, "Quality of service aware routing protocol in software-defined internet of vehicles," *IEEE Internet Things J.*, vol. 6, pp. 2817–2828, Apr. 2019.

[211] L. Cui *et al.*, "A blockchain-based containerized edge computing platform for the internet of vehicles," *IEEE Internet Things J.*, vol. 8, pp. 2395–2408, Feb. 2021.

[212] Y. Xu, , G. Gui, H. Gacanin, and F. Adachi, "A survey on resource allocation for 5G heterogeneous networks: Current research, future trends, and challenges," *IEEE Commun. Surveys Tut.*, vol. 23, pp. 668–695, 2nd Quart. 2021.

[213] Z. Lin *et al.*, "Secure and energy efficient transmission for RSMA-based cognitive satellite-terrestrial networks," *IEEE Wireless Commun. Lett.*, vol. 10, pp. 251–255, Feb. 2021.

[214] J. Ding, M. Mahyar, C. Ranaweera, and J. Choi, "IoT connectivity technologies and applications: A survey," *IEEE Access*, vol. 8, pp. 67646–67673, Apr. 2020.

[215] M. A. Al-Jarrah, A. Al-Dweik, M. Kalil, and S. S. Ikki, "Decision fusion in distributed cooperative wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 68, pp. 797–811, Jan. 2019.

[216] A. Ahmed, A. Al-Dweik, Y. Iraqi, H. Mukhtar, M. Naeem, and E. Hossain, "Hybrid automatic repeat request (HARQ) in wireless communications systems and standards: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 23, pp. 2711–2752, 4th Quart. 2021.

[217] "3GPP 38 series of specifications: 5G New Radio." `https://www.3gpp.org/DynaReport/38-series.htm`. Accessed: 2020-08-18.

[218] D. Cai, Z. Ding, P. Fan, and Z. Yang, "On the performance of NOMA with hybrid ARQ," *IEEE Trans. Veh. Technol.*, vol. 67, pp. 10033–10038, Oct. 2018.

[219] Z. Yu, C. Zhai, W. Ni, and D. Wang, "Non-orthogonal multiple access with

cooperative truncated arq and relay selection," *IEEE Access*, vol. 7, pp. 56228–56243, Apr. 2019.

[220] F. Cao, Y. Song, and Y. Yang, "ARQ assisted short-packet communications for NOMA networks over Nakagami-*m* fading channels," *IEEE Access*, vol. 8, pp. 158263–158272, Aug. 2020.

[221] S. Park and S. Choi, "Performance of symbol-level combining and bit-level combining in MIMO multiple ARQ systems," *IEEE Trans. Commun.*, vol. 64, pp. 1517–1528, Apr. 2016.

[222] R. Mai, T. V. K. Chaitanya, and T. Le-Ngoc, "Progressive hybrid precoding and combining for massive MIMO ARQ systems," *IEEE Access*, vol. 6, pp. 34503–34515, Jun. 2018.

[223] A. U. Rehman, C. Dong, L. L. Yang, and L. Hanzo, "Performance of cognitive stop-and-wait hybrid automatic repeat request in the face of imperfect sensing," *IEEE Access*, vol. 4, pp. 5489–5508, Jul. 2016.

[224] H. Mukhtar, A. Al-Dweik, M. Al-Mualla, and A. Shami, "Low complexity power optimization algorithm for multimedia transmission over wireless networks," *IEEE J. Sel. Topics Signal Process.*, vol. 9, pp. 113–124, Feb. 2015.

[225] H. Mukhtar, A. Al-Dweik, M. Al-Mualla, and A. Shami, "Adaptive hybrid ARQ system using turbo product codes with hard/soft decoding," *IEEE Commun. Lett.*, vol. 17, pp. 2132–2135, Nov. 2013.

[226] K. Vakilinia *et al.*, "Optimizing transmission lengths for limited feedback with nonbinary LDPC examples," *IEEE Trans. Commun.*, vol. 64, pp. 2245–2257, Jun. 2016.

[227] K. Miura, Y. Kishiyama, and K. Higuchi, "Low latency HARQ method using early retransmission before channel decoding based on superposition coding," in *Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, (Taipei, Taiwan), pp. 1–2, Feb. 2019.

[228] A. Khreis, F. Bassi, P. Ciblat, and P. Duhamel, "Multi-layer HARQ with delayed feedback," *IEEE Trans. Wireless Commun.*, vol. 19, pp. 6224–6237, Sep. 2020.

[229] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-s. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, pp. 721–742, Second Quart. 2017.

[230] Y. Yin, M. Liu, G. Gui, H. Gacanin, H. Sari, and F. Adachi, "QoS-oriented dynamic power allocation in NOMA-based wireless caching networks," *IEEE Wireless Commun. Lett.*, vol. 10, pp. 82–86, Jan. 2021.

[231] A. E. Mostafa, Y. Zhou, and V. W. S. Wong, "Connectivity maximization for

narrowband IoT systems with NOMA," in *IEEE Int. Conf. Commun. (ICC)*, (Paris, France), pp. 1–6, May 2017.

[232] W. U. Khan, J. Liu, F. Jameel, V. Sharma, R. Jäntti, and Z. Han, "Spectral efficiency optimization for next generation NOMA-enabled IoT networks," *IEEE Trans. Veh. Technol.*, vol. 69, pp. 15284–15297, Dec. 2020.

[233] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, pp. 3313–3351, May 2021.

[234] V. C. Thirumavalavan and T. S. Jayaraman, "BER analysis of reconfigurable intelligent surface assisted downlink power domain NOMA system," in *Int. Conf. Commun. Syst. Netw. (COMSNETS)*, (Bengaluru, India), pp. 519–522, Mar. 2020.

[235] L. Bariah, S. Muhaidat, P. C. Sofotasios, F. E. Bouanani, O. A. Dobre, and W. Hamouda, "Large intelligent surface-assisted nonorthogonal multiple access for 6G networks: Performance analysis," *IEEE Internet Things J.*, vol. 8, pp. 5129–5140, Apr. 2021.

[236] Z. Wang, Y. Liu, X. Mu, Z. Ding, and O. A. Dobre, "NOMA empowered integrated sensing and communication," *IEEE Commun. Lett.*, vol. 26, pp. 677–681, Mar. 2022.

[237] 3GPP, *Study on downlink multiuser superposition transmission (MUST) for LTE (Release 13)*. PhD thesis, 2015.

[238] H. Lee, S. Kim, and J.-H. Lim, "Multiuser superposition transmission (MUST) for LTE-A systems," in *IEEE Int. Conf. Commun. (ICC)*, (Kuala Lumpur, Malaysia), pp. 1–6, May 2016.

[239] A. Kureev, I. Levitsky, and E. Khorov, "Experimental study of constellation rotation in NOMA Wi-Fi networks," *J. Commun. Technol. Electron.*, vol. 65, pp. 1525–1530, Jan. 2020.

[240] E. Khorov, A. Kureev, I. Levitsky, and I. F. Akyildiz, "A phase noise resistant constellation rotation method and its experimental validation for NOMA Wi-Fi," *IEEE J. Sel. Areas Commun.*, vol. 40, pp. 1346–1354, Apr. 2022.

[241] L. Mohjazi, L. Bariah, S. Muhaidat, P. C. Sofotasios, O. Onireti, and M. A. Imran, "Error probability analysis of non-orthogonal multiple access for relaying networks with residual hardware impairments," in *IEEE Int. Symp. Personal Indoor Mobile Radio Commun. (PIMRC)*, (Istanbul, Turkey), pp. 1–6, Sep. 2019.

[242] A. Al-Dweik, E. Alsusa, O. A. Dobre, and R. Hamila, "Multirate NOMA for improving connectivity in 6G communications networks," 2023.

[243] 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, 2021.

[244] B. Clerckx, Y. Mao, E. A. Jorswieck, J. Yuan, D. J. Love, E. Erkip, and D. Niyato, "A primer on rate-splitting multiple access: Tutorial, myths, and frequently asked questions," *IEEE Jo. Sel. Areas Commun.*, vol. 41, pp. 1265–1308, May 2023.

[245] F. Xiao, X. Li, L. Yang, H. Liu, and T. A. Tsiftsis, "Outage performance analysis of RSMA-aided semi-grant-free transmission systems," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 253–268, 2023.

[246] A. Bansal, K. Singh, B. Clerckx, C.-P. Li, and M.-S. Alouini, "Rate-splitting multiple access for intelligent reflecting surface aided multi-user communications," *IEEE Trans. Veh. Technol.*, vol. 70, pp. 9217–9229, Sep. 2021.

[247] 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, 2021.

[248] C.-P. Wei, H. Yang, C.-P. Li, and Y.-M. Chen, "SCMA decoding via deep learning," *IEEE Wireless Commun. Lett.*, vol. 10, pp. 878–881, Apr. 2021.

[249] W. Lindsey, "Error probabilities for Rician fading multichannel reception of binary andn-ary signals," *IEEE Trans. Inf. Theory*, vol. 10, pp. 339–350, Oct. 1964.

Appendix A

Definitions of $\mathcal{A}_{l,n}$ and Optimal Power Assignment

The definitions of $\mathcal{A}_{l,n}$ are listed in Table A.1 and Table A.2. Note that $\mathcal{A}_{l,n}$ has the general form of $\mathcal{A}_{i,n} = \frac{2|h_n|^2 g_{i,n}}{N_0} = g_{i,n}\gamma_n$. Moreover, the optimal power assignment for $N = 2$ and $N = 3$ using various modulation orders is shown in Table A.3 and Table A.4.

Table A.1: Definitions of $\mathcal{A}_{l,n}$, $N = 2$.

$\mathcal{A}_{1,1}$	$\frac{\left(A_{11}-A_{\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{5,1}$	$\frac{\left(A_{11}+3A_{\grave{1}1}\right)^2}{4\gamma_1^{-1}}$
$\mathcal{A}_{2,1}$	$\frac{\left(3A_{11}+A_{\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{1,2}$	$\frac{\left(A_{11}\right)^2}{\gamma_2^{-1}}$
$\mathcal{A}_{3,1}$	$\frac{\left(A_{\grave{1}1}\right)^2}{\gamma_1^{-1}}$	$\mathcal{A}_{2,2}$	$\frac{\left(A_{\grave{1}1}\right)^2}{\gamma_2^{-1}}$
$\mathcal{A}_{4,1}$	$\frac{\left(A_{11}\right)^2}{\gamma_1^{-1}}$		

Table A.2: Definitions of $\mathcal{A}_{l,n}$, $N = 3$.

$\mathcal{A}_{1,1}$	$\frac{\left(A_{111}-A_{\grave{1}11}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{21,1}$	$\frac{\left(A_{1\grave{1}1}\right)^2}{\gamma_1^{-1}}$
$\mathcal{A}_{2,1}$	$\frac{\left(2A_{111}-A_{1\grave{1}1}-A_{\grave{1}\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{22,1}$	$\frac{\left(3A_{1\grave{1}1}+A_{\grave{1}11}\right)^2}{4\gamma_1^{-1}}$
$\mathcal{A}_{3,1}$	$\frac{\left(2A_{111}+A_{1\grave{1}1}+A_{\grave{1}\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{23,1}$	$\frac{\left(A_{\grave{1}11}+A_{1\grave{1}1}-2A_{\grave{1}\grave{1}1}\right)^2}{4\gamma_1^{-1}}$
$\mathcal{A}_{4,1}$	$\frac{\left(3A_{111}+A_{\grave{1}11}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{24,1}$	$\frac{\left(3A_{\grave{1}\grave{1}1}+A_{1\grave{1}1}\right)^2}{4\gamma_1^{-1}}$
$\mathcal{A}_{5,1}$	$\frac{\left(A_{\grave{1}11}-A_{1\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{25,1}$	$\frac{\left(2A_{\grave{1}\grave{1}1}+A_{111}+A_{\grave{1}11}\right)^2}{4\gamma_1^{-1}}$
$\mathcal{A}_{6,1}$	$\frac{\left(A_{\grave{1}11}\right)^2}{\gamma_1^{-1}}$	$\mathcal{A}_{1,2}$	$\frac{\left(2A_{111}-A_{\grave{1}11}-A_{1\grave{1}1}\right)^2}{4\gamma_2^{-1}}$
$\mathcal{A}_{7,1}$	$\frac{\left(3A_{\grave{1}11}+A_{1\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{2,2}$	$\frac{\left(2A_{111}+A_{\grave{1}11}+A_{1\grave{1}1}\right)^2}{4\gamma_2^{-1}}$
$\mathcal{A}_{8,1}$	$\frac{\left(A_{1\grave{1}1}-A_{\grave{1}\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{3,2}$	$\frac{\left(A_{\grave{1}11}-A_{1\grave{1}1}\right)^2}{4\gamma_2^{-1}}$
$\mathcal{A}_{9,1}$	$\frac{\left(3A_{1\grave{1}1}+A_{\grave{1}\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{4,2}$	$\frac{\left(3A_{\grave{1}11}+A_{1\grave{1}1}\right)^2}{4\gamma_2^{-1}}$
$\mathcal{A}_{1\grave{1},1}$	$\frac{\left(2A_{1\grave{1}1}+A_{111}+A_{\grave{1}11}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{5,2}$	$\frac{\left(A_{1\grave{1}1}\right)^2}{\gamma_2^{-1}}$
$\mathcal{A}_{11,1}$	$\frac{\left(A_{111}+A_{\grave{1}11}-2A_{\grave{1}\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{6,2}$	$\frac{\left(A_{\grave{1}11}+A_{1\grave{1}1}-2A_{\grave{1}\grave{1}1}\right)^2}{4\gamma_2^{-1}}$
$\mathcal{A}_{12,1}$	$\frac{\left(A_{\grave{1}\grave{1}1}\right)^2}{\gamma_1^{-1}}$	$\mathcal{A}_{7,2}$	$\frac{\left(A_{\grave{1}\grave{1}1}\right)^2}{\gamma_2^{-1}}$
$\mathcal{A}_{13,1}$	$\frac{\left(2A_{\grave{1}\grave{1}1}+A_{\grave{1}11}+A_{1\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{8,2}$	$\frac{(A_{111})^2}{\gamma_2^{-1}}$
$\mathcal{A}_{14,1}$	$\frac{\left(2A_{111}-A_{\grave{1}11}-A_{1\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{9,2}$	$\frac{\left(A_{\grave{1}11}\right)^2}{\gamma_2^{-1}}$
$\mathcal{A}_{15,1}$	$\frac{(A_{111})^2}{\gamma_1^{-1}}$	$\mathcal{A}_{1\grave{1},2}$	$\frac{\left(3A_{1\grave{1}1}+A_{\grave{1}11}\right)^2}{4\gamma_2^{-1}}$
$\mathcal{A}_{16,1}$	$\frac{\left(2A_{111}+A_{\grave{1}11}+A_{1\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{11,2}$	$\frac{\left(2A_{\grave{1}\grave{1}1}+A_{\grave{1}11}+A_{1\grave{1}1}\right)^2}{4\gamma_2^{-1}}$
$\mathcal{A}_{17,1}$	$\frac{\left(2A_{\grave{1}11}-A_{1\grave{1}1}-A_{\grave{1}\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{1,3}$	$\frac{(A_{111})^2}{\gamma_3^{-1}}$
$\mathcal{A}_{18,1}$	$\frac{\left(2A_{\grave{1}11}+A_{1\grave{1}1}+A_{\grave{1}\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{2,3}$	$\frac{\left(A_{\grave{1}11}\right)^2}{\gamma_3^{-1}}$
$\mathcal{A}_{19,1}$	$\frac{\left(3A_{\grave{1}11}+A_{111}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{3,3}$	$\frac{\left(A_{1\grave{1}1}\right)^2}{\gamma_3^{-1}}$
$\mathcal{A}_{2\grave{1},1}$	$\frac{\left(A_{111}+A_{\grave{1}11}-2A_{1\grave{1}1}\right)^2}{4\gamma_1^{-1}}$	$\mathcal{A}_{4,3}$	$\frac{\left(A_{\grave{1}\grave{1}1}\right)^2}{\gamma_3^{-1}}$

Table A.3: Optimal power assignment including the system's average BER for different **m** and SNR, $N = 2$

	15 dB		30 dB		45 dB		60 dB	
m	α_1	BER	α_1	BER	α_1	BER	α_1	BER
$[2,2]$	0.145	5.3×10^{-2}	0.138	2.0×10^{-3}	0.138	6.3×10^{-5}	0.138	2.0×10^{-6}
$[4,4]$	0.151	9.1×10^{-2}	0.139	4.0×10^{-3}	0.138	1.3×10^{-4}	0.138	4.0×10^{-6}
$[8,8]$	6.2×10^{-2}	2.1×10^{-1}	5.2×10^{-2}	1.9×10^{-2}	5.0×10^{-2}	6.5×10^{-4}	5.0×10^{-2}	2.1×10^{-5}
$[16,16]$	5.2×10^{-2}	2.5×10^{-1}	4.7×10^{-2}	3.0×10^{-2}	4.6×10^{-2}	1.1×10^{-3}	4.6×10^{-2}	3.5×10^{-5}
$[64,64]$	1.1×10^{-2}	3.5×10^{-1}	1.3×10^{-2}	1.2×10^{-1}	1.3×10^{-2}	9.7×10^{-3}	1.3×10^{-2}	3.3×10^{-4}
$[4,2]$	0.244	5.6×10^{-2}	0.232	2.2×10^{-3}	0.232	6.9×10^{-5}	0.232	2.2×10^{-6}
$[2,4]$	0.103	7.9×10^{-2}	9.2×10^{-2}	3.3×10^{-3}	9.2×10^{-2}	1.9×10^{-4}	9.2×10^{-2}	3.3×10^{-6}
$[8,4]$	0.154	1.3×10^{-1}	0.136	7.3×10^{-3}	0.135	2.4×10^{-4}	0.135	7.6×10^{-6}
$[4,8]$	7.0×10^{-2}	1.7×10^{-1}	5.7×10^{-2}	9.9×10^{-3}	5.7×10^{-2}	3.2×10^{-4}	5.7×10^{-2}	1.0×10^{-5}
$[16,8]$	8.2×10^{-2}	2.1×10^{-1}	7.3×10^{-2}	2.0×10^{-2}	7.2×10^{-2}	7.0×10^{-4}	7.2×10^{-2}	2.2×10^{-5}
$[8,16]$	4.2×10^{-2}	2.4×10^{-1}	3.4×10^{-2}	2.8×10^{-2}	3.2×10^{-2}	1.0×10^{-3}	3.2×10^{-2}	3.2×10^{-5}
$[64,8]$	7.9×10^{-2}	2.4×10^{-1}	8.2×10^{-2}	4.1×10^{-2}	8.2×10^{-2}	1.7×10^{-3}	8.2×10^{-2}	5.4×10^{-5}
$[8,64]$	1.1×10^{-2}	3.4×10^{-1}	9.0×10^{-3}	8.0×10^{-2}	8.0×10^{-3}	3.8×10^{-3}	8.0×10^{-3}	1.2×10^{-4}

Table A.4: Optimal power assignment including the system's average BER for different **m** and SNR, $N = 3$

	15 dB			30 dB			45 dB			60 dB		
m	α_1	α_2	BER	α_1	α_2	BER	α_1	α_2	BER	α_1	α_2	BER
$[2,2,2]$	3.2×10^{-2}	0.180	1.8×10^{-1}	2.4×10^{-2}	0.158	1.2×10^{-2}	2.3×10^{-2}	0.156	3.8×10^{-4}	2.4×10^{-2}	0.160	1.2×10^{-5}
$[4,4,4]$	3.5×10^{-2}	0.187	2.5×10^{-1}	2.4×10^{-2}	0.160	2.2×10^{-2}	2.3×10^{-2}	0.156	7.6×10^{-4}	2.3×10^{-2}	0.156	2.4×10^{-5}
$[8,8,8]$	4.0×10^{-3}	6.6×10^{-2}	3.8×10^{-1}	3.0×10^{-3}	5.9×10^{-2}	1.2×10^{-1}	3.0×10^{-3}	5.5×10^{-2}	8.5×10^{-3}	3.0×10^{-3}	5.4×10^{-2}	2.8×10^{-4}
$[16,16,16]$	3.0×10^{-3}	3.8×10^{-2}	4.1×10^{-1}	3.0×10^{-3}	5.6×10^{-2}	1.7×10^{-1}	3.0×10^{-3}	5.4×10^{-2}	1.5×10^{-2}	2.0×10^{-3}	5.4×10^{-2}	5.0×10^{-4}
$[8,4,2]$	4.9×10^{-2}	0.289	2.1×10^{-1}	4.0×10^{-2}	0.263	2.1×10^{-2}	3.8×10^{-2}	0.261	7.2×10^{-4}	3.8×10^{-2}	0.262	2.3×10^{-5}
$[2,4,8]$	9.0×10^{-3}	9.6×10^{-2}	3.3×10^{-1}	6.0×10^{-3}	6.7×10^{-2}	4.6×10^{-2}	5.0×10^{-3}	0.127	3.8×10^{-3}	6.0×10^{-3}	6.4×10^{-2}	5.5×10^{-5}
$[16,8,4]$	1.7×10^{-2}	0.175	3.1×10^{-1}	1.3×10^{-2}	0.158	6.8×10^{-2}	1.1×10^{-2}	0.149	3.4×10^{-3}	1.1×10^{-2}	0.148	1.1×10^{-4}
$[4,8,16]$	7.0×10^{-3}	2.2×10^{-2}	3.9×10^{-1}	3.0×10^{-3}	3.9×10^{-2}	1.3×10^{-1}	2.0×10^{-3}	3.5×10^{-2}	7.3×10^{-3}	2.0×10^{-3}	3.5×10^{-2}	2.4×10^{-4}

Appendix B

Optimal Power Assignments for C-NOMA and PANOMA Systems

The optimal power assignment to minimize the average BER for $N = 2$ and $N = 3$ using various modulation orders is shown in Table B.1 and Table B.2. On the other hand, the optimal power assignment to maximize the capacity for $N = 2$ and $N = 3$ using various modulation orders is shown in Table B.3 and Table B.4

Table B.1: Optimal power assignment to minimize the system's average BER for different modulation orders under various E_b/N_0 conditions, $N = 2$.

		15 dB			30 dB			45 dB		
m	System	α_1	α_2	BER	α_1	α_2	BER	α_1	α_2	BER
[2,2]	C − NOMA	0.1419	0.8581	2.8700E−2	0.1382	0.8618	9.9863E−4	0.1381	0.8619	3.1681E−5
	PANOMA	0.1176	0.8824	2.8700E−2	0.1131	0.8869	9.9863E−4	0.1130	0.8870	3.1681E−5
[4,4]	C − NOMA	0.1419	0.8581	2.8700E−2	0.1382	0.8618	9.9863E−4	0.1381	0.8619	3.1681E−5
	PANOMA	0.1176	0.8824	2.8700E−2	0.1131	0.8869	9.9863E−4	0.1130	0.8870	3.1681E−5
[2,4]	C − NOMA	0.0960	0.9040	3.1200E−2	0.0917	0.9083	1.1000E−3	0.0916	0.9084	3.4770E−5
	PANOMA	0.0490	0.9510	2.9000E−2	0.0451	0.9549	1.0000E−3	0.0451	0.9549	3.1905E−5
	E − PANOMA	0.0492	$\alpha_2^{\Re} = 0.9508$ $\alpha_2^{\Im} = 0.6900$	2.8500E−2	0.0458	$\alpha_2^{\Re} = 0.9542$ $\alpha_2^{\Im} = 0.7020$	9.8690E−4	0.0500	$\alpha_2^{\Re} = 0.9500$ $\alpha_2^{\Im} = 0.6749$	3.1341E−5
[4,2]	C − NOMA	0.2359	0.7641	2.1400E−2	0.2317	0.7683	7.2871E−4	0.2307	0.7693	2.3100E−5
	PANOMA	0.2369	0.7631	2.1500E−2	0.2326	0.7674	7.3094E−4	0.2314	0.7686	2.3172E−5
	E − PANOMA	$\alpha_1^{\Re} = 0.2010$ $\alpha_1^{\Im} = 0.3957$	0.7990	2.0600E−2	$\alpha_1^{\Re} = 0.1992$ $\alpha_1^{\Im} = 0.3827$	0.8008	7.0035E−4	$\alpha_1^{\Re} = 0.2009$ $\alpha_1^{\Im} = 0.3932$	0.7991	2.2204E−5

Table B.2: Optimal power assignment to minimize the system's average BER for identical BPSK and identical QPSK under various E_b/N_0 conditions, $N = 3$.

		15 dB			30 dB			45 dB		
m	System	α_1	α_2	BER	α_1	α_2	BER	α_1	α_2	BER
$[2,2,2]$	C – NOMA	0.0277	0.1699	8.8400E−2	0.0234	0.1566	3.9000E−3	0.0233	0.1565	1.2631E−4
	PANOMA	0.0641	0.1946	9.0600E−2	0.0628	0.1988	4.2000E−3	0.0628	0.1995	1.3446E−4
$[4,4,4]$	C – NOMA	0.0277	0.1699	8.8400E−2	0.0234	0.1565	3.9000E−3	0.0233	0.1562	1.2630E−4
	PANOMA	0.1013	0.2201	9.0600E−2	0.1064	0.2283	4.2000E−3	0.1026	0.2265	1.3446E−4

Table B.3: Optimal power assignment to maximize the system's constrained capacity (bit/symbol) for different modulation orders under various E_b/N_0 conditions, $N = 2$.

		15 dB			30 dB			45 dB		
m	System	α_1	α_2	Sum Rate	α_1	α_2	Sum Rate	α_1	α_2	Sum Rate
$[2,2]$	C – NOMA	0.1405	0.8595	1.6243	0.1376	0.8624	1.9772	0.1377	0.8623	1.9990
	PANOMA	0.1159	0.8841	1.6243	0.1124	0.8876	1.9771	0.1125	0.8875	1.9990
$[4,4]$	C – NOMA	0.1405	0.8595	3.2486	0.1376	0.8624	3.9544	0.1377	0.8623	3.9979
	PANOMA	0.1159	0.8841	3.2486	0.1124	0.8876	3.9544	0.1125	0.8875	3.9979
$[2,4]$	C – NOMA	0.0792	0.9208	2.4518	0.0768	0.9232	2.9640	0.0768	0.9232	2.9984
	PANOMA	0.0309	0.9691	2.4355	0.0289	0.9711	2.9658	0.0290	0.9710	2.9984
	E – PANOMA	0.0314	$\alpha_2^{\Re} = 0.9686$ $\alpha_2^{\Im} = 0.8966$	2.4360	0.0294	$\alpha_2^{\Re} = 0.9706$ $\alpha_2^{\Im} = 0.8881$	2.9659	0.0296	$\alpha_2^{\Re} = 0.9704$ $\alpha_2^{\Im} = 0.8827$	2.9984
$[4,2]$	C – NOMA	0.2826	0.7174	2.5547	0.2734	0.7266	2.9742	0.2726	0.7274	2.9988
	PANOMA	0.2829	0.7171	2.5541	0.2740	0.7260	2.9742	0.2733	0.7267	2.9988
	E – PANOMA	$\alpha_1^{\Re} = 0.2494$ $\alpha_1^{\Im} = 0.4674$	0.7506	2.5690	$\alpha_1^{\Re} = 0.2424$ $\alpha_1^{\Im} = 0.4465$	0.7576	2.9752	$\alpha_1^{\Re} = 0.2416$ $\alpha_1^{\Im} = 0.4452$	0.7584	2.9989

Table B.4: Optimal power assignment to maximize the system's constrained capacity (bit/symbol) for identical BPSK and identical QPSK under various E_b/N_0 conditions, $N = 3$.

		15 dB			30 dB			45 dB		
m	System	α_1	α_2	Sum Rate	α_1	α_2	Sum Rate	α_1	α_2	Sum Rate
$[2,2,2]$	C – NOMA	0.0273	0.1673	1.7081	0.0233	0.1560	2.8887	0.0232	0.1556	2.9945
	PANOMA	0.1009	0.2170	1.6871	0.0985	0.2228	2.8829	0.1042	0.2267	2.9942
$[4,4,4]$	C – NOMA	0.0273	0.1673	3.4162	0.0233	0.1560	5.7774	0.0232	0.1556	5.9891
	PANOMA	0.1049	0.2193	3.3741	0.1023	0.2250	5.7659	0.1041	0.2267	5.9885

Appendix C

Average BER and Performance Gain Analyses for Gray-Labeling NOMA

C.1 Average BER Analysis

Furthermore, the average BER over Rician fading channels can be found by integrating the instantaneous BER over the probability density function (PDF) of the Rician fading channel, i.e $\overline{P}_{B_n} = \int_0^\infty P_n f\left(\gamma_{n,l}\right) d\gamma_{n,l}$. By following [22, Eq. (8.177)] and [249], the average BER can be expressed as

$$\overline{P}_n = \sum_j \frac{c_j}{\beta} \left[Q_1\left(u_j, v_j\right) - \frac{1+\Theta_j}{2\exp\left(\frac{u_j^2+v_j^2}{2}\right)} I_0\left(u_j, v_j\right) \right] \tag{C.1}$$

where $Q_1(\cdot)$ is the first order Marcum Q-function, $\Theta_j = \sqrt{\frac{p_j}{2+p_j}}$, $p_j = \frac{\left|\Delta_j\right|^2 \overline{\gamma}_{n,l}}{1+K_{n,l}}$ and

$$\{u_j, v_j\} = \sqrt{K_{n,l}\left(\frac{1+p_j}{2+p_j} \mp \Theta_j\right)}. \tag{C.2}$$

C.2 Performance Gain Analysis

C.2.1 U_1 Analysis

To illustrate the performance gain of the joint-multilayer Gray-mapping over the disjoint-multilayer Gray-mapping, $[M_1, M_2] = [4, 16]$ is considered. Let the difference between

the two resulting BER expressions be denoted as ΔP_1. By noting the BER expressions derived in [24, 34] for the disjoint-multilayer Gray-mapping and the expression in (5.28) for the joint-multilayer Gray-mapping, ΔP_1 for this case can be expressed as follows

$$\begin{aligned}\Delta P_1 = \frac{3}{2}Q\left(A_{12}\sqrt{\gamma_{n,l}}\right) - \frac{3}{2}Q\left(A_{\acute{1}2}\sqrt{\gamma_{n,l}}\right) + \frac{3}{4}Q\left(A_{\acute{1}1}\sqrt{\gamma_{n,l}}\right) \\ - \frac{3}{4}Q\left(A_{11}\sqrt{\gamma_{n,l}}\right) + \frac{1}{2}Q\left(A_{\acute{1}3}\sqrt{\gamma_{n,l}}\right) - \frac{1}{2}Q\left(A_{13}\sqrt{\gamma_{n,l}}\right) \\ + \frac{1}{2}Q\left(A_{16}\sqrt{\gamma_{n,l}}\right) - \frac{1}{2}Q\left(A_{\acute{1}6}\sqrt{\gamma_{n,l}}\right) + \frac{1}{4}Q\left(A_{\acute{1}5}\sqrt{\gamma_{n,l}}\right) \\ - \frac{1}{4}Q\left(A_{15}\sqrt{\gamma_{n,l}}\right)\end{aligned} \tag{C.3}$$

which can be approximated considering the most dominant terms as

$$\Delta P_1 \approx \frac{3}{4}\left[Q\left(A_{\acute{1}1}\sqrt{\gamma_{n,l}}\right) - Q\left(A_{11}\sqrt{\gamma_{n,l}}\right)\right]. \tag{C.4}$$

Therefore, while setting $K_{n,l} = 0$ at asymptotically high SNR, $\Delta\bar{P}_1^{\infty}$ can be expressed as

$$\Delta\bar{P}_1^{\infty} = \frac{3}{8\bar{\gamma}_{n,l}}\left[\frac{1}{A_{\acute{1}1}^2} - \frac{1}{A_{11}^2}\right] = \frac{1}{2\bar{\gamma}_{n,l}}\frac{3\left(5^{3/4}\sqrt{2\alpha_1\alpha_2}\right)}{5\alpha_1^2 - 2\sqrt{5}\alpha_1\alpha_2 + \alpha_2^2}. \tag{C.5}$$

It is noted that the term $\frac{1}{2\bar{\gamma}_{n,l}}$ is the single-user asymptotic BER, whereas the fraction including the power coefficient is related to NOMA. Considering the ranges of the power coefficients, it is seen that $\Delta\bar{P}_1^{\infty}$ is always positive indicating that the performance gain of the joint-multilayer Gray-mapping is ensured.

C.2.2 U_2 Analysis

Since the BER expressions for the joint-multilayer Gray-mapping and disjoint-multilayer Gray-mapping are identical for U_2. Thus, $\Delta P_2 = 0$.

Appendix D

GNOM BER Expressions

The generalized non-orthogonal multiplexing (GNOM) BER, $P_{k,l}^{(u)}$, can be computed using (8.9) such that $\mathbf{c}$ can be found in the following tables for different N_l and various packet indices. $\left|\Delta_j\right|$ can be computed using

$$\left|\Delta_j\right| = \left|\sum_{i=1}^{N_l} g_{j,i}\sqrt{\alpha_{i,l}^{(u)}}\right|. \tag{D.1}$$

Table D.1: BER parameters for $N_l = 2$.

k	$\mathbf{g}_1$	$\mathbf{g}_2$	$\mathbf{c}$
1	1	1	2
	−1	1	2
2	0	1	1
	−1	1	2
	2	1	2
	−2	1	−2
	1	1	−2

Table D.2: BER parameters for $N_l = 3$.

k	$\mathbf{g}_1$	$\mathbf{g}_2$	$\mathbf{g}_3$	$\mathbf{c}$
1	1	1	1	4
	1	1	−1	4
	1	−1	1	4
	−1	1	1	4
2	0	1	1	2
	0	−1	1	2
	2	1	−1	4
	1	−1	1	4
	−1	1	1	4
	2	1	1	4
	2	−1	1	−4
	1	1	−1	−4
	1	1	1	−4
	−2	1	1	−4
3	0	0	1	1
	0	−1	1	2
	0	2	1	2
	2	0	1	2
	2	1	−1	4
	2	−2	1	4
	2	2	1	4
	−1	1	1	4
	−2	1	1	4
	1	1	−1	4
	0	−2	1	−2
	0	1	1	−2
	−2	0	1	−2
	2	−1	1	−4
	−2	2	1	−4
	1	1	1	−4
	2	2	−1	−4
	1	−1	1	−4
	2	1	1	−4

Table D.3: BER parameters for $N_l = 4$.

k	$\mathbf{g}_1$	$\mathbf{g}_2$	$\mathbf{g}_3$	$\mathbf{g}_4$	$\mathbf{c}$
1	1	1	1	1	8
	1	1	1	−1	8
	1	1	−1	1	8
	1	−1	1	1	8
	−1	1	1	1	8
	1	−1	−1	1	8
	−1	1	−1	1	8
	−1	−1	1	1	8
2	0	1	1	1	4
	0	1	1	−1	4
	0	1	−1	1	4
	0	−1	1	1	4
	−1	1	1	1	8
	2	1	1	1	8
	1	−1	1	1	8
	1	−1	−1	1	8
	−1	1	−1	1	8
	2	1	1	−1	8
	2	1	−1	1	8
	−2	−1	1	1	8
	1	1	−1	1	−8
	−2	1	1	1	−8
	2	−1	−1	1	−8
	1	1	1	1	−8
	−1	−1	1	1	−8
	−2	1	−1	1	−8
	1	1	1	−1	−8
	2	−1	1	1	−8
3	0	0	−1	1	2
	0	0	1	1	2
	0	2	1	−1	4
	2	0	1	−1	4
	0	1	−1	1	4
	0	−1	1	1	4
	0	2	1	1	4
	2	0	1	1	4
	2	−1	−1	1	8
	2	2	1	−1	8
	−2	2	−1	1	8
	1	1	−1	1	8
	−1	1	1	1	8
	−2	1	1	1	8
	1	−1	−1	1	8
	−1	−1	1	1	8
	2	1	−1	1	8
	−2	−1	1	1	8
	2	−2	1	1	8
	2	2	1	1	8
	0	2	−1	1	−4
	0	1	1	1	−4
	2	0	−1	1	−4
	0	1	1	−1	−4
	0	−2	1	1	−4
	−2	0	1	1	−4
	1	1	1	1	−8
	2	−2	−1	1	−8
	2	2	−1	1	−8
	1	1	1	−1	−8
	1	−1	1	1	−8
	2	1	1	1	−8
	−1	1	−1	1	−8
	2	1	1	−1	−8
	−2	1	−1	1	−8
	2	−1	1	1	−8
	−2	−2	1	1	−8
	−2	2	1	1	−8

Table D.4: BER parameters for $N_l = 4$.

k	$\mathbf{g}_1$	$\mathbf{g}_2$	$\mathbf{g}_3$	$\mathbf{g}_4$	$\mathbf{c}$
4	0	0	0	1	1
	0	0	−1	1	2
	0	0	2	1	2
	0	2	0	1	2
	2	0	0	1	2
	0	2	1	−1	4
	2	0	1	−1	4
	0	2	−2	1	4
	0	2	2	1	4
	2	0	1	−1	4
	2	0	2	1	4
	2	−2	0	1	4
	2	2	0	1	4
	0	1	1	−1	4
	0	−1	1	1	4
	0	−2	1	1	4
	−2	0	1	1	4
	2	2	1	−1	8
	−2	2	−1	1	8
	2	−2	−2	1	8
	2	2	−2	1	8
	2	−2	2	1	8
	−2	1	1	1	8
	2	2	2	1	8
	1	1	1	−1	8
	−1	1	1	1	8
	−2	1	−1	1	8
	−2	−1	1	1	8
	−2	−2	1	1	8
	−1	1	−1	1	8
	−1	−1	1	1	8
	2	1	1	−1	8
	−2	2	1	1	8
	0	0	−2	1	−2
	0	0	1	1	−2
	0	−2	0	1	−2
	−2	0	0	1	−2
	0	2	−1	1	−4
	2	0	−1	1	−4
	0	−2	2	1	−4
	−2	0	2	−1	−4
	−2	2	0	1	−4
	0	1	1	1	−4
	0	2	2	−1	−4
	2	0	2	−1	−4
	2	2	0	−1	−4
	0	1	−1	1	−4
	0	2	1	1	−4
	2	0	1	1	−4
	1	1	1	1	−8
	2	−1	−1	1	−8
	1	1	−1	1	−8
	2	−2	−1	1	−8
	2	2	−1	1	−8
	1	−1	1	1	−8
	−2	2	−2	1	−8
	2	−2	1	1	−8
	2	1	1	1	−8
	−2	−2	2	1	−8
	−2	2	2	1	−8
	2	2	2	−1	−8
	2	2	1	1	−8
	1	−1	−1	1	−8
	2	1	−1	1	−8
	2	−1	1	1	−8