Resource Allocation in Multicarrier Non-Orthogonal Multiple Access (NOMA) Systems

Estela Carmona Cejudo

Supervisor: Dr. Huiling Zhu

University of Kent

Thesis submitted for the degree of

Doctor of Philosophy

in Electronic Engineering

October, 2020

Canterbury, United Kingdom

To Jack and Alexander, for making every single day bright and special.

Acknowledgment

First and foremost, I would like to express my gratitude to my supervisor Dr. Huiling Zhu for her invaluable guidance and for sharing her vast knowledge with me. She inspired and motivated me enormously to pursue a career in research. I would like to extend my sincere thanks to Prof. Jiangzhou Wang for his very insightful comments and contributions, and for always encouraging me to grow as a research scientist. I would like to offer my special thanks to Prof. Ali Ghrayeb and Dr. Osama Alluhaibi for their insightful comments and suggestions during the preparation of this work. I would like to express my appreciation to my fellow students and colleagues, and very specially to Dr. Hind Albasry, for a cherished time spent together. My appreciation also goes out to my mother, brother and sister, and to my dear friends Carmen C. A. and Virginia C. N. for their continuous encouragement and support during my Ph.D. Finally, I would like to thank my beloved sons Jack A. C. and Alexander A. C. for their tremendous patience all through my studies.

List of Publications

1. E. Carmona Cejudo, H. Zhu and O. Alluhaibi, "On the Power Allocation and Constellation Selection in Downlink NOMA," *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, 2017, pp. 1-5.

2. E. Carmona Cejudo, H. Zhu, J. Wang and O. Alluhaibi, "A Fast Algorithm for Resource Allocation in Downlink Multicarrier NOMA," *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, 2019, pp. 1-5.

3. E. Carmona Cejudo, H. Zhu, and J. Wang, "Resource Allocation in Downlink Multicarrier NOMA under a Fairness Constraint," *2020 IEEE 92nd Vehicular Technology Conference (VTC-Fall)*, 2020.

4. E. Carmona Cejudo, H. Zhu, and J. Wang, "Resource Allocation in Multicarrier NOMA Systems Based on Optimal Channel Gain Ratios," submitted to *IEEE Transactions on Wireless Communications* (under review), 2020.

5. E. Carmona Cejudo, H. Zhu, and J. Wang, "Resource Allocation in BER-Constrained Multicarrier NOMA Systems," submitted to *2021 IEEE International Conference on Communications (ICC)*, 2020.

Abstract

Non-orthogonal multiple access (NOMA) can improve the spectrum efficiency and enable massive connectivity in future wireless communications systems by multiplexing multiple users in a non-orthogonal manner. Many previous works in power-domain NOMA addressed research problems from the perspective of the channel capacity, assumed perfect successive interference cancellation (SIC) and considered the pairing of users with very distinct channel conditions. This can yield inefficient power allocations in terms of sum-rate. Further, the assumption of perfect SIC is not realistic in practical systems, where SIC error propagation greatly impacts the achievable bit error rate (BER) at the receivers.

By applying NOMA to multicarrier-based schemes, the capabilities of both can be enhanced through resource allocation, i.e. the assignment of radio resources to users under an optimization objective. However, resource allocation in multicarrier NOMA systems may lead to a nondeterministic polynomial time (NP)-hard problem requiring exhaustive search, which has prohibitive computational complexity. Instead, efficient algorithms that provide a good trade-off between system performance and implementation practicality are needed.

The contributions presented in this thesis are two-fold. First, new performance bounds on the BER of NOMA systems are provided. And second, a novel resource allocation scheme is presented, which can achieve a performance close to optimal with low computational complexity. The contributions are summarized as follows.

First, theoretical BER expressions are presented for multi-layer, multi-level quadrature amplitude modulation (QAM) in NOMA. To the best of the author's knowledge, this work represents the first attempt in developing such expressions. The optimal value of the power allocation factor in terms of BER is analytically derived. Further, the theoretical BER expressions are used for calculating the ratios of users' channel gains that maximize the sum-rate. Unlike previous research in NOMA, it is demonstrated that, in NOMA systems with QAM, the channel gains of two NOMA users must be of approximately the same order of magnitude in order to guarantee that inter-user interference can be overcome at the receivers. Additionally, accurate BER approximations are presented in the form of exponential functions. These are used for finding numerical boundaries for the values of the channel gain ratios of NOMA users that fulfill the BER constraints.

Second, the contributions on BER boundaries are applied to develop of a novel resource allocation scheme for multicarrier NOMA. A user pairing algorithm of quasi-linear complexity with respect to the number of users is proposed, based on the findings about NOMA optimal channel gain ratios and channel gain gaps. In contrast, the complexity of exhaustive search is of the order of the squared number of users. The problem of power and data rate allocation is solved by applying a Lagrangian optimization method based on the previously derived BER exponential approximations. The optimization result is applied to propose a novel iterative resource allocation (IRA)-data rate selection (DRS) algorithm. Unlike existing works, continuous power levels and discrete modulation schemes are considered. Numerical simulations demonstrate that IRA-DRS yields a sum-rate performance close to optimal, providing an excellent trade-off between computational complexity and performance. IRA-DRS benefits from multi-user diversity in terms of achievable sum-rate, number of iterations required for convergence, and degrees of freedom in choosing different combinations of modulation levels.

Contents

List of Figures

List of Tables

List of Algorithms

Acronyms

1G first generation

3GPP 3rd generation partnership project

4G fourth generation

5G fifth generation

AWGN additive white Gaussian noise

BER bit error rate

BPSK binary phase shift keying

CDMA code division multiple access

DRS data rate selection

FDMA frequency division multiple access

FTPC fractional transmit power control

IC interference cancellation

IoE Internet of Everything

IoT Internet of Things

IRA iterative resource allocation

LTE long-term evolution

MMSE minimum mean squared error

NOMA non-orthogonal multiple access

NP nondeterministic polynomial-time

OFDM orthogonal frequency division multiplexing

OFDMA orthogonal frequency division multiple access

OMA orthogonal multiple access

QAM quadrature amplitude modulation

QPSK quadrature phase shift keying

SC superposition coding

SCUS sub-optimal single-carrier user selection

SER symbol error rate

SIC successive interference cancellation

SNR signal-to-noise ratio

TDMA time division multiple access

UB-LDDP upper-bound Lagrangian duality and dynamic programming

Notation

$\mathbb{C}$	set of complex numbers
$\mathbb{I}^+$	set of positive integers
$\mathbb{R}$	set of real numbers
$\mathbf{A}_{(i)}$	updated values of elements in A after the i-th algorithm iteration
$\alpha_{j,s}$	power allocation factor of user j on subcarrier s
β_0	BER constraint
$\beta_{j,s}$	BER at user j on subcarrier s
$\delta_{j,s}$	binary allocation variable for user j on subcarrier s
$d_{j,s}$	Euclidean distance between symbols in $\mathcal{X}_{j,s}$
$d_{min,j,s}$	minimum Euclidean distance between symbols in user j's constellation layer after superposition
$d'_{j,s}$	difference between Euclidean distances of superposed constellations on subcarrier s
$\mathbb{E}\{\cdot\}$	statistical expectation
$E_{sym,j,s}$	average symbol energy of user j on subcarrier s
$E_{bit,j,s}$	average bit energy of user j on subcarrier s
$\varepsilon_{j,s}$	excess power allocated to user j on subcarrier s
$\mathcal{F}_s$	power correction factor on subcarrier s
$\hat{\mathcal{F}}_s$	power correction factor associated to $\hat{R}_{j,s}$, $\hat{R}_{k,s}$

$\mathcal{F}_s^*$	power correction factor associated to $R_{j,s}^*$, $R_{k,s}^*$
$g_{\{j,k\},s}$	channel gain ratio of users j and k on subcarrier s
$\mathcal{G}(M_j, M_k, \beta_0)$	optimal $g_{\{j,k\},s}$, for given M_j, M_k, β_0
$h_{j,s}$	channel response of user j on subcarrier s
$\lvert h_{j,s}\rvert$	channel gain of user j on subcarrier s
$H_{\{j,k\},s}$	channel gain gap of user j and user k on subcarrier s
$\mathbf{H}_s$	set of all possible channel gain gaps that fulfill the channel gain ratio requirement on subcarrier s
$\mathbf{J}$	set of possible NOMA partners for a given user
$\mathcal{K}$	set of active users
K	number of active users
$\mathcal{L}$	Lagrangian function
λ	Lagrange multiplier
$M_{j,s}$	modulation level of user j on subcarrier s
$\hat{M}_{j,s}$	real version of $M_{j,s}$
$\tilde{M}_{j,s}$	modulation level associated to $\tilde{R}_{j,s}$
$M_{j,s}^*$	modulation level associated to $R_{j,s}^*$
$\mathbf{M_U}$	vector of modulation levels of elements in $\mathbf{U}$
$\max\{\cdot\}$	maximal element
$N_0/2$	power spectral density of AWGN
$\mathcal{O}(\cdot)$	numerical complexity
$\overline{p}$	average transmit power per subcarrier
p_s	transmit power on subcarrier s
$\check{p}_s$	transmit power on subcarrier s after power correction
p_j	user j's transmit power
$p_{j,s}$	user j's transmit power on subcarrier s
p_{total}	total downlink transmit power

Δp_s	additional extra power required on subcarrier s
p_{unused}	unused power
$\mathbf{P}$	vector of transmit powers per subcarrier
$P_{j,s}(e)$	symbol error rate at user j on subcarrier s
$P_{j,s}^{(I)}(e)$	error probability in the in-phase dimension at user j on subcarrier s
$P_{j,s}^{(Q)}(e)$	error probability in the quadrature dimension at user j on subcarrier s
$P_{k,s}(correct_{x_{j,s}})$	probability of correctly decoding $x_{j,s}$ at user k on subcarrier s
$P_{k,s}(error_{x_{j,s}})$	probability of error during the decoding of $x_{j,s}$ at user k on subcarrier s
$P_{k,s}(e\|correct_{x_{j,s}})$	probability of error at user k on subcarrier s, conditioned on the correct decoding of $x_{j,s}$
$P_{k,s}(e\|error_{x_{j,s}})$	probability of error at user k on subcarrier s, conditioned on error during the decoding of $x_{j,s}$
$Q(\cdot)$	Q-function
$R_{j,s}$	data rate of user j on subcarrier s
$\hat{R}_{j,s}$	real version of $R_{j,s}$
$\tilde{R}_{j,s}$	round-up value of $\hat{R}_{j,s}$
$R_{j,s}^*$	increased data rate at user j on subcarrier s
$R_{j\to k,s}$	data rate at user k when decoding user j's signal on subcarrier s
$\mathbf{R_U}$	vector of data rates of elements in $\mathbf{U}$
$\mathcal{S}$	set of orthogonal subcarriers
S	number of orthogonal subcarriers
$x_{j,s}$	user j's transmit symbol on subcarrier s

$x_{\{j,k\},s}$	transmit supersymbol for users j and k on subcarrier s
$\mathbf{x}$	Euclidean vector of transmit symbol x
$\mathcal{X}_f$	transmit superconstellation on subcarrier s
$\mathcal{X}_{j,s}$	user j's transmit constellation on subcarrier s
$\mathbf{U}$	subcarrier allocation vector
$\mathbf{V}$	low-order users allocation vector
$\mathbf{W}$	high-order users allocation vector
W	size of the system bandwidth (Hertz)
$z_{j,s}$	received noise power at user j on subcarrier s
$\mathcal{CN}(a, \sigma^2)$	circularly symmetric complex Gaussian distribution with mean a and variance σ^2
$\|\|\cdot\|\|_2$	Euclidean norm of a vector
$\lceil\cdot\rceil$	round-up function
$\lfloor\cdot\rfloor$	round-down function
$\overline{\mathbf{A}}$	average of values in $\mathbf{A}$

Chapter 1

Introduction

Contents

1.1 Motivation

Over the last few years, the evolution of wireless networks has been driven by the need for the fifth generation (5G) of cellular systems to provide a 1000-fold increase in capacity with respect to fourth generation (4G) long-term evolution (LTE) networks [4, 5]. While the deployment of 5G cellular systems is ongoing, considerable research efforts are already being directed to beyond-5G wireless networks [6–8].

Beyond-5G wireless networks should support the intelligent information society of 2030, including the mobile Internet and the Internet of Everything (IoE) [9, 10]. These will be two major drivers for beyond-5G wireless networks that will enhance 5G vertical applications and enable pervasive computing scenarios. The information society of 2030 will be highly digitized, intelligence inspired, and globally data driven [1]. Meanwhile, an unprecedented proliferation of new IoE services, devices and connections is already occurring. Some examples of IoE applications include extended reality services, telemedicine, flying vehicles and connected autonomous systems [11].

It is envisioned that, among other requirements, beyond-5G systems in the 2030s will have to provide ten times the connectivity density of 5G, reaching up to 10^7 devices/km^2, with an area traffic capacity of up to 1 Gb/s/m^2, along with up to a 100-fold increase in network energy efficiency and a 10-fold increase in spectrum efficiency [1], as illustrated in Figure 1.1. Massive machine-type communications networks will be occupied by the traffic of billions of devices, posing a major challenge in the design of beyond-5G networks [12].

Since multiple access techniques allow multiple users to share communication resources, their importance is paramount in the design of wireless communications networks. From the first generation (1G) of cellular systems to 5G, diverse orthogonal multiple access (OMA) techniques have been utilized, namely frequency division multiple access (FDMA) [13], time division multiple access (TDMA) [14], code division multiple access (CDMA) [15] and orthogonal frequency division multiple access (OFDMA) [16]. However, the use of OMA schemes introduces additional challenges to wireless systems with a massive number of devices, given that the available orthogonal resources may not be sufficient to support this huge number of devices. Many of these connections only require sporadic transmissions with very diverse data rate and latency requirements. Therefore, allocating re-

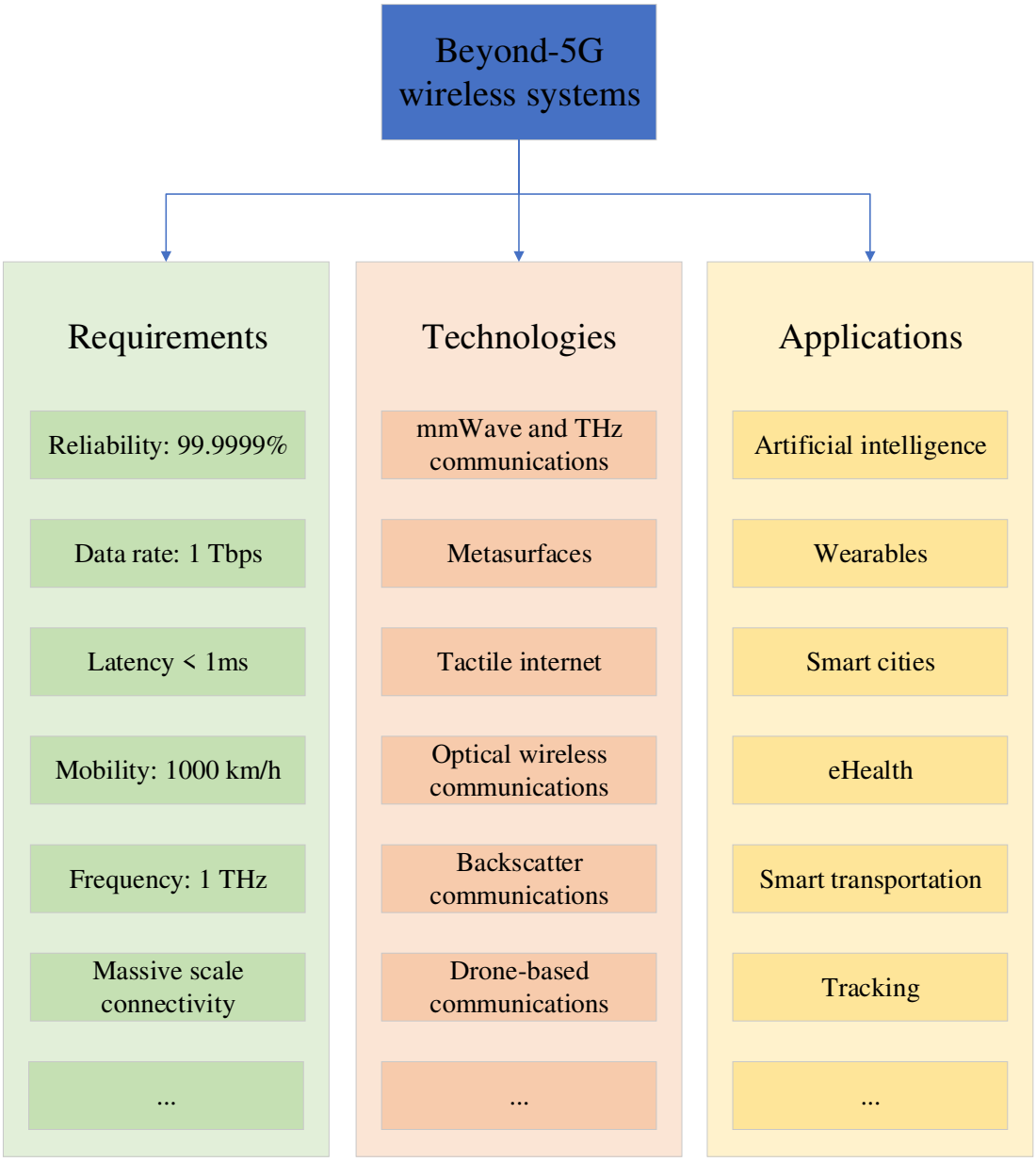


Figure 1.1: Typical scenarios and capabilities of beyond-5G wireless networks [1, 2].

source blocks to each one of these connections in an orthogonal manner is neither feasible nor efficient [17]. In contrast, non-orthogonal multiple access (NOMA) systems possess the ability to serve multiple devices in one resource block, thus addressing the large explosion of connected users forecast in the massive machine-type communications and IoE paradigms of beyond-5G networks. In particular,

NOMA introduces the following attractive advantages for beyond-5G networks in comparison with OMA schemes [18]:

- Improved spectral efficiency and energy efficiency [19–21], given that the same bandwidth can be shared among multiple users.
- Improved user fairness [19], since multiple users can be accommodated in the same resource block with guaranteed minimum data rate requirements.
- Low transmission latency [22], as NOMA allows for more flexible user scheduling schemes as well as grant-free transmissions.
- Higher cell-edge throughput, by flexibly changing the fraction of power allocated to cell-edge users [23] and/or by applying a cooperative relaying scheme [24] in order to support a certain quality of service.

One of the major benefits of NOMA is that it can be easily combined with other existing and emerging technologies [25]. This can contribute to further increasing the scalability, spectral efficiency and energy efficiency of future wireless communication networks. In [26], the authors provided an interesting survey on the combination of NOMA with emerging technologies such as massive multiple-input multiple-output and millimeter wave communications [27–31], cooperative communications [32], cognitive communications [33], physical layer security [34], energy harvesting [35], visible light communications [36] and mobile edge computing [37]. Moreover, NOMA can be realized in the downlink and the uplink [38], and in various domains such as power, code [39–41], space [42] or combinations of them [43–45]. Figure 1.2 presents a simple classification of NOMA techniques.

Power-domain NOMA has been regarded as the most promising NOMA scheme [46, 47], and therefore it is the focus of the work in this thesis. In power-domain NOMA, multiple access is achieved in a non-orthogonal manner by using superposition coding (SC) at the transmitter and successive interference cancellation

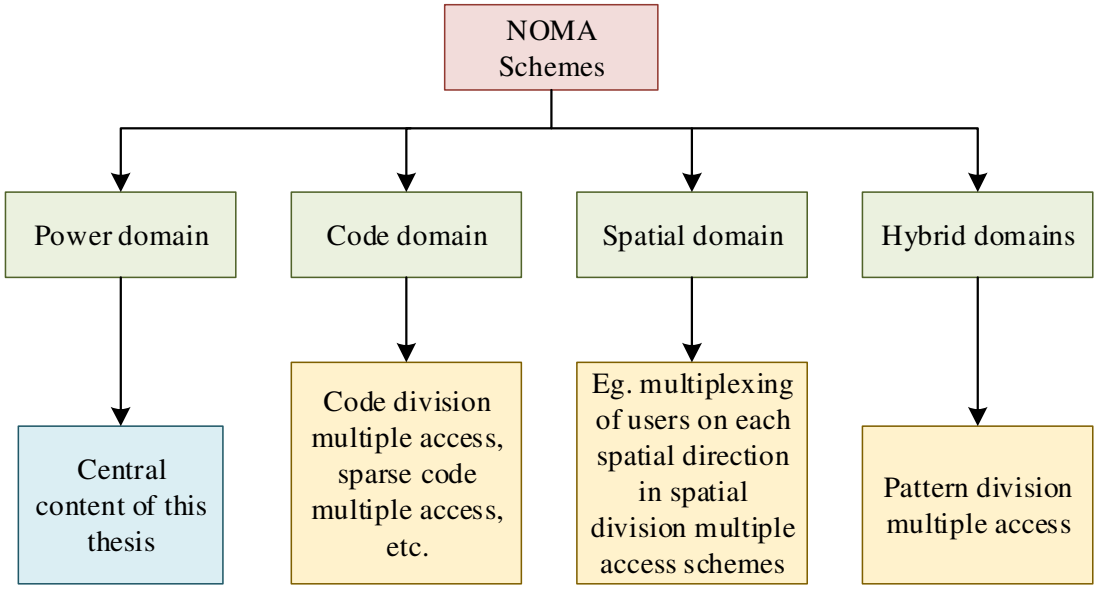


Figure 1.2: Classification of NOMA techniques [3].

(SIC) at the receiver [48]. Resource blocks can be allocated to multiple users in the same time slot through power-domain multiplexing. Hence, NOMA can provide service to a greater number of users than other OMA technologies. Further, compared to OMA, NOMA is able to offer a higher system throughput and accommodate more devices in major scenarios of beyond-5G wireless networks, such as massive machine-type communications, enhanced mobile broadband and ultra-reliable low latency communications [49].

1.2 Challenges

For the variety of reasons stated in Section 1.1, NOMA has received considerable attention as a potential multiple access technique for 5G systems over the last few years. However, up to this date, the relative performance gain of NOMA over OMA is still smaller than the cost of its implementation complexity. Although NOMA was considered as a study item in the 3rd generation partnership project (3GPP) for 5G new radio, it was decided not to continue with it as a work item.

Instead, it was decided to leave it for beyond-5G systems, where the proliferation of new use cases with an ultra-dense number of devices may be motivating for the use of NOMA [50].

Therefore, in order for NOMA to fulfill the challenges of beyond-5G scenarios with ultra dense devices, it is necessary to address some further research challenges. Previous works demonstrated that, in terms of achievable sum-rate, it is preferable to pair two NOMA users with very distinct channel conditions [51]. This result has since been applied to numerous researches. However, the application of this theoretical result might turn unrealistic in scenarios where some users have a very poor channel link. In a NOMA setting, these users would have to be assigned most of the NOMA power share in order to enable successful decoding at the receiver. Such power allocations are inefficient in terms of the sum-rate because most of the power is allocated to a user with a very weak channel condition.

Previous works such as [3, 51] also assume perfect SIC at the receivers. However, in practical wireless communications systems, the use of SIC can result in error propagation as successive layers of information are removed at the receivers. This results in decoding errors being carried over from one decoding stage to the next, greatly impacting the achievable BERs at the receivers. This can, in turn, make the application of NOMA unfeasible. Therefore, it is of paramount importance to understand the impact of performance optimization parameters, such as the transmit power and the power allocation factor, on the achievable BER in NOMA. Based on the need for considering more realistic scenarios in NOMA and on the impact of imperfect SIC on the system performance in terms of BER, new performance bounds on the BER of NOMA systems are presented in this thesis.

Moreover, optimal user pairing in NOMA –i.e. the selection of the optimal set of NOMA users to be multiplexed into the same radio resource– becomes specially

challenging for systems with a large number of users, such as IoE settings, because it requires a large channel state information acquisition and feedback overhead, as well as hugely computationally demanding optimization algorithms [50]. Therefore, there is a need to improve the practicality of implementation of user pairing algorithms. Most existing user pairing schemes in the literature considered solutions based on broadly classifying users according to their channel gains and randomly selecting users from a group according to a system optimization metric. Albeit of low complexity, these approaches yield a performance far from the optimal. In this thesis, knowledge about the BER performance bounds is applied to propose a user pairing algorithm of quasi-linear complexity.

In addition, the spectral efficiency of NOMA needs to be increased in order to provide larger gains than its OMA counterpart. Nevertheless, the capabilities of NOMA can be further extended through its combination with multicarrier schemes. The application of NOMA to OMA-based multicarrier schemes offers a a lower implementation complexity than purely NOMA-based systems, and therefore a better trade-off between performance and complexity. The key to achieve the full potential of multicarrier NOMA is resource allocation, which carries out an assignment of radio resources to users with a certain objective, for example maximizing the system throughput or minimizing the transmit power. Therefore, it is necessary to investigate how to reduce the computational complexity of resource allocation in multicarrier NOMA systems, while achieving a performance close to optimal.

Resource allocation in multicarrier NOMA systems requires the joint optimization of subcarrier allocation, power allocation and data rate assignment [52–54]. However, many resource allocation problem formulations in NOMA lead to a mixed-integer, nondeterministic polynomial-time (NP)-hard problem [55]. Hence, the optimal solution can only be found through exhaustive search, which

is a combinatorial optimization problem. One approach to select the best user set is to search over all possible combinations of users, then select the one that provides the maximum sum-rate on each subcarrier [56]. However, this causes prohibitive computational complexity. A more practical approach is to separate the problems of subcarrier allocation and power allocation, fix one of them and optimize the other [55], [57]. This leads to suboptimal but practical and efficient solutions. In summary, novel resource allocation schemes are needed in NOMA which are capable of providing a performance close to optimal in a practical and simplified manner. In this thesis, a novel resource allocation scheme that provides a performance close to optimal with a reduced computational complexity is proposed.

1.3 Contributions

Despite of the many potentials of NOMA, much work is still needed to further increase its performance and reduce its implementation complexity. This can be achieved to a certain extent through the application of NOMA to subcarrier-based schemes. Further, it is necessary to study NOMA under more realistic assumptions, such as imperfect SIC receivers and their impact on the BER performance.

The contributions presented in this thesis are two-fold. First, new performance bounds on the BER of NOMA systems are provided. And second, a novel resource allocation scheme is presented, which can achieve a performance close to optimal with low computational complexity. These are summarized below.

1.3.1 BER Performance Bounds in NOMA

First of all, theoretical BER expressions are presented for the BER in NOMA, assuming multi-layer, multi-level QAM. To the best of the author's knowledge,

this work represents the first attempt in developing such expressions, with previous literature focusing mainly on fixed-level modulation. The analytical BER expressions are used for deriving the optimal value of the power allocation factor for which the BER is minimized. The optimal power allocation factor value is also the value that minimizes the transmit power in BER-constrained NOMA. It is proved that the optimal value of the power allocation factor is dependent on the modulation levels assigned to each user.

The theoretical BER expressions are used for calculating the optimal channel gain ratios for a pair of NOMA users, i.e. the ratios of channel gains that maximize the achievable sum-rate for a given BER constraint. Unlike previous research in NOMA [51], this work demonstrates that, in NOMA systems with practical QAM, the channel gains of two NOMA users must be of approximately the same order of magnitude in order to guarantee that the effect of inter-user interference can be overcome at the receivers, specially when the modulation levels assigned to both NOMA users are similar.

However, the derived BER expressions are in terms of products of the Q-function. These expressions are neither easily differentiable nor easily invertible in their arguments, and it becomes complicated to apply them to developing a resource allocation strategy. Therefore, accurate BER approximations are presented in the form of exponential functions. The BER exponential approximations are used for finding numerical boundaries for the values of the channel gain ratios of NOMA users that fulfill BER constraints for different combinations of modulation levels; any pair of NOMA users whose channel gain ratio falls outside the numerical boundaries for a given modulation level cannot meet the individual BER constraints.

Publications Related to the Contributions on the BER in NOMA

1. E. Carmona Cejudo, H. Zhu and O. Alluhaibi, "On the Power Allocation and Constellation Selection in Downlink NOMA," *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, 2017, pp. 1-5.

2. E. Carmona Cejudo, H. Zhu, and J. Wang, "Resource Allocation in Multicarrier NOMA Systems Based on Optimal Channel Gain Ratios," submitted to *IEEE Transactions on Wireless Communications* (under review), 2020.

3. E. Carmona Cejudo, H. Zhu, and J. Wang, "Resource Allocation in BER-Constrained Multicarrier NOMA Systems," submitted to *2021 IEEE International Conference on Communications (ICC)*, 2020.

1.3.2 Resource Allocation in NOMA

In terms of the subcarrier allocation problem, the findings about optimal channel gain ratios and channel gain gaps are applied to propose a user pairing algorithm that achieves quasi-linear complexity with respect to the number of users. In contrast, the complexity of exhaustive search procedures is of the order of the squared number of users. Other sub-optimal user pairing schemes in the literature [55, 56] also achieve a much larger computational complexity than the algorithm presented in the thesis.

Further, BER exponential approximations are used for solving the problem of power and data rate allocation by applying a Lagrangian optimization method [58], with the objective of maximizing the sum-rate. The optimization result is applied to proposing a novel iterative resource allocation (IRA)-data rate selection (DRS) algorithm for BER-constrained multicarrier NOMA systems, with the objective of maximizing the sum-rate. Unlike existing works, continuous power levels and discrete modulation schemes are considered in order to achieve

a performance close to optimal with reduced computational complexity. Through numerical simulations, it is demonstrated that the proposed resource allocation scheme for multicarrier NOMA systems yields a performance close to optimal, and it outperforms other suboptimal schemes such as fractional transmit power control (FTPC) [56]. Therefore, IRA-DRS yields an excellent trade-off between sum-rate performance and computational complexity. Further, the proposed algorithm greatly benefits from multi-user diversity in terms of achievable sum-rate, number of iterations required for convergence, and degrees of freedom in choosing different combinations of modulation levels at each subcarrier that yield the same sum-rate.

Publications Related to the Contributions on Resource Allocation in NOMA

1. E. Carmona Cejudo, H. Zhu, J. Wang and O. Alluhaibi, "A Fast Algorithm for Resource Allocation in Downlink Multicarrier NOMA," *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, 2019, pp. 1-5.

2. E. Carmona Cejudo, H. Zhu, and J. Wang, "Resource Allocation in Downlink Multicarrier NOMA under a Fairness Constraint," *2020 IEEE 92nd Vehicular Technology Conference (VTC-Fall)*, 2020.

3. E. Carmona Cejudo, H. Zhu, and J. Wang, "Resource Allocation in Multicarrier NOMA Systems Based on Optimal Channel Gain Ratios," submitted to *IEEE Transactions on Wireless Communications* (under review), 2020.

4. E. Carmona Cejudo, H. Zhu, and J. Wang, "Resource Allocation in BER-Constrained Multicarrier NOMA Systems," submitted to *2021 IEEE International Conference on Communications (ICC)*, 2020.

1.4 Structure of the Thesis

The remainder of this thesis is organized as follows.

- Chapter 2 offers an overview of the basic principles of OMA and NOMA techniques, with a special emphasis on multicarrier-based systems, SC and SIC.

- Chapter 3 presents research on the performance bounds on the BER of NOMA systems. Analytical BER derivations are obtained, along with their exponential approximations, numerical values of optimal channel gain ratios in NOMA, and boundaries for the allowable channel gap conditions for the fulfillment of BER constraints in NOMA.

- Chapter 4 presents research on resource allocation in multicarrier NOMA systems, based on the optimal channel gain ratios that yield the maximum sum-rate for a given BER constraint. User pairing and power and data rate allocation schemes are proposed. Their close-to-optimal performance and low computational complexity are demonstrated through numerical simulations.

- Finally, Chapter 5 provides a summary of the work presented in this thesis and highlights its main conclusions. In addition, potential future research directions are discussed.

Chapter 2

Overview of Multiple Access Techniques

Contents

Multiple access techniques allow multiple users to share communication resources. Therefore, their importance is paramount in the design of modern wireless communication systems, and most specially in the paradigms of beyond-5G networks with ultra-dense number of users.

Over the past few decades, diverse multiple access techniques have been adopted by wireless communications systems from 1G to 4G, namely FDMA,

TDMA, CDMA, and OFDMA. All of these techniques belong to the category of OMA schemes, where the allocation of resources to users is done in an orthogonal manner, in the time, frequency or code domain, or their combinations. Therefore, the number of supported users in OMA schemes is limited by the number of orthogonal resources in the system [59].

In contrast, NOMA schemes allocate resources to users in a non-orthogonal manner, by multiplexing several users in the power, code, or space domain, or their combinations. Therefore, NOMA-based systems can provide service to a larger number of users than OMA-based systems, thus addressing the large explosion of connected users forecast in beyond-5G networks.

However, the benefits introduced by NOMA come at a high cost of increased implementation complexity. A good trade-off between system throughput and complexity can be achieved through the implementation of hybrid NOMA-OMA systems, such as OFDM-based multicarrier NOMA. In the following, a brief overview of the main OMA and NOMA technologies is presented, and the role of multicarrier NOMA in enabling beyond-5G systems is discussed.

The rest of this Chapter is organized as follows. The basic principles of OMA are summarized in Section 2.1, with a special emphasis in OFDMA systems. Section 2.2 introduces the theoretical principles of NOMA with SC and SIC and multicarrier NOMA systems.

2.1 Basic Principles of OMA

Conventional OMA schemes allocate orthogonal resources to users either in the time, frequency or code domain, or their combinations. In TDMA, an exclusive time slot is allocated to each user, and transmissions occupy the entire bandwidth. In FDMA, the spectrum is divided up into orthogonal channels, and each

frequency resource is allocated to a single user. In CDMA, channels are defined by code, and users are assigned with different codes that have low cross-correlation properties [60]. Figure 2.1 represents the orthogonal division of resource blocks in TDMA, FDMA and CDMA. The major benefit of orthogonality in OMA schemes is that multi-user detection can be achieved with low-complexity receivers [59], since inter-user interference is greatly reduced due to the orthogonality principle.

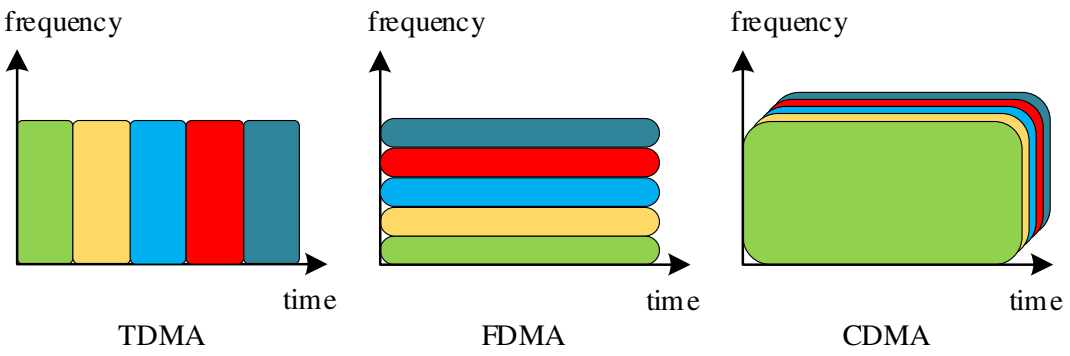


Figure 2.1: Orthogonal division of resource blocks in TDMA, FDMA and CDMA.

2.1.1 OFDMA

4G wireless systems adopted OFDMA, which is based on OFDM. In OFDM systems, the broadband channel is divided into a number of flat-fading narrowband channels.

In downlink OFDMA, the base station transmits data to a set of users whose channel conditions are time and frequency variant. Due to the scarcity of spectrum and power resources, these must be allocated most effectively at the transmitter, in order to optimize some system performance metric. This objective is achieved through resource allocation, which comprises three basic tasks in OFDMA systems [16]:

- Subcarrier allocation, consisting in allocating users to subcarriers in an

efficient manner, according to factors such as the channel conditions of users.

- Data rate allocation, consisting in varying the number of transmitted bits per OFDM symbol in each subcarrier, according to the objective performance metric and the instantaneous subchannel quality, while maintaining an acceptable quality of service.

- Power allocation, consisting in the effective distribution of power over subcarriers, so as to maintain the link quality and optimize the objective performance function.

It was shown in [61] that, in OFDMA systems with mutually independent subchannels among users, the best performance is achieved by allocating each subcarrier to the user with the best channel condition over that subcarrier, and by allocating transmit power to subcarriers by following the water-filling principle [62].

2.2 Basic Principles of NOMA

Power-domain NOMA has been regarded as the most promising NOMA scheme [46, 47], and therefore it is the focus of the work presented in this thesis.

Power domain NOMA –referred to as NOMA hereafter– transmits several users' signals on the same time, frequency and spatial resources by superposing them in the power domain through the use of SC techniques [63]. The throughput of each NOMA user is controlled through suitable adjustment of the power allocation factor and modulation level at the transmitter.

Signals of multiple users are separated at the receivers by using advanced multi-user detection and demodulation mechanisms based on the SIC technique.

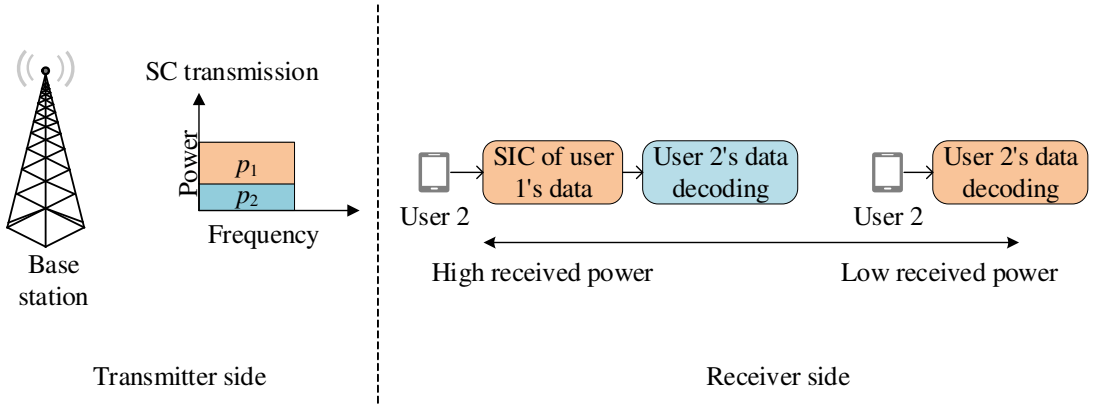


Figure 2.2: 2-user NOMA downlink scenario with SC at the transmitter and SIC at the receiver.

SIC receivers exploit differences in received signals strengths to iteratively decode each user's data stream [3, 64]. NOMA with SIC can significantly enhance the throughput of users with a poor channel condition, and thus the user fairness, compared to OMA [65]. Figure 2.2 represents a 2-user NOMA downlink scenario where SC is applied at the transmitter and SIC is employed at the receiver.

2.2.1 Superposition Coding

At the transmitter, the superposition of users' signals in NOMA can be achieved through the use of SC schemes [63]. SC was first proposed in [64] as a technique for simultaneously transmitting data to several receivers from a single transmitter [66, 67].

A SC scheme can be viewed as a multiplexing technique where a high throughput is achieved by simultaneously transmitting data in the form of superimposed layers. In order to make SC practical, information relevant to each user must be encoded at the transmitter, while applying a different layer-specific amplitude through power control. For example, in a two-user scenario, two point-to-point encoders at the transmitter map each user's input data stream to complex-valued sequences of the two-user superposed signal [3].

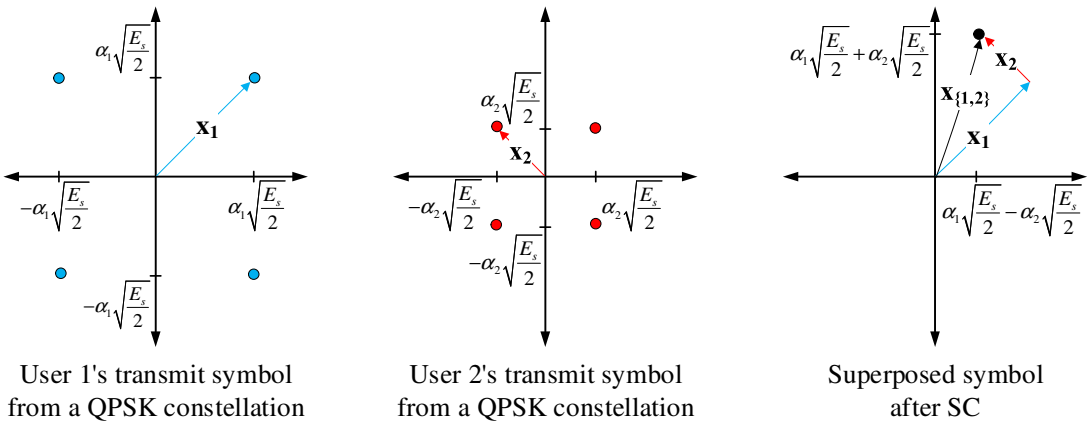


Figure 2.3: Example of SC encoding, where $E_s = 1$ and $\alpha_1 + \alpha_2 = 1$.

Figure 2.3 represents the superposition of two quadrature phase shift keying (QPSK)-modulated symbols after applying SC, where the average symbol energy E_s is normalized to one in both superposed layers as well as in the resulting superposed constellation. In Figure 2.3, more power is allocated to user 1 and, as a result, the Euclidean norm of vector $\mathbf{x_1}$ is larger than the Euclidean norm of vector $\mathbf{x_2}$. Further, the Euclidean vector of symbol $x_{\{1,2\}}$ is $\mathbf{x_{\{1,2\}}} = \mathbf{x_1} + \mathbf{x_2}$, and its Euclidean norm is calculated as $||\mathbf{x_{\{1,2\}}}||_2 = \sqrt{||\mathbf{x_1}||_2^2 + ||\mathbf{x_2}||_2^2}$. Note that boldface symbols denote vectors, and $||\cdot||_2$ denotes the Euclidean norm of a vector.

Each user's transmit constellation can be considered as an independent information layer. During superposition, a different power allocation factor is applied to each layer. After SC, the superposed contellation is formed by non-uniformly spaced signal points, where the position of signal points depends on the power allocation factor allocated to individual layers during superposition.

The throughput of each NOMA user is controlled through suitable adjustment of the power allocation factor and the modulation level at the transmitter. Further, every layer of information offers a different degree of protection against errors [68], [69], which can also be adjusted through the power allocation factor and the modulation level for each user.

2.2.2 Successive Interference Cancellation

User demultiplexing at the receiver side is ensured in NOMA through the allocation of an adequate power difference between paired users during transmission, and the application of SIC in the power-domain at the receivers, by exploiting the differences in signal strength among the data streams of interest. Typically, the user who enjoys a better channel condition is allocated with less power than the user with a poorer channel condition. However, this is not always the case [17], since the power level must be adjusted according to the target data rates at the users.

The order of SIC decoding depends exclusively on the magnitude of the users' channel gains [17,70], independently of power allocation. To achieve the capacity region, SIC needs to be carried out in increasing order of users' channel gain magnitudes [17]. Following the example in Figure 2.3, SIC decoding at user 2 is represented in Figure 2.4, where it is assumed that user 1 enjoys a poorer channel condition than user 2, so it comes first in the decoding order. Therefore, user 1 can recover its signal directly, without applying SIC, by disregarding user 2's signal as noise. User 2 decodes user 1's signal first, and then subtracts it from the received data stream. After decoding and removing user 1's signal, user 2 can decode its own signal.

The use of SIC can result in error propagation as successive layers of information from the incoming signal are removed at the receivers. However, the implementation of SC with SIC is feasible in the two-user case, through the choice appropriate parameters such as channel gain difference between users, modulation type and power allocation factor [71].

The high decoding complexity in NOMA is a limiting factor for practical system implementation, and it remains an open research challenge [50]. However, the extended processing capabilities of new generations of mobile devices together

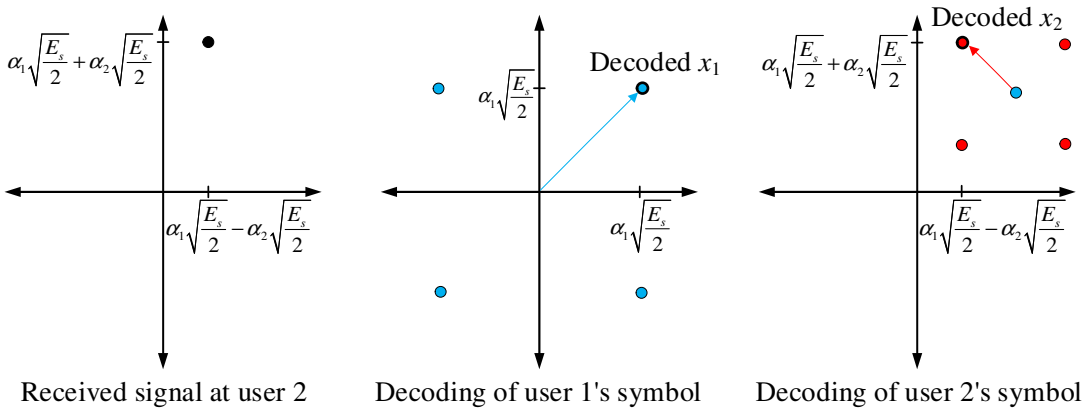


Figure 2.4: Example of SIC decoding at user 2, where $E_s = 1$ and $\alpha_1 + \alpha_2 = 1$.

with recent advances such as network-assisted interference cancellation and suppression techniques have made the use of SIC at the user equipment practical. For simpler terminals such as low-cost Internet of Things (IoT) devices [72], interference cancellation still remains challenging, but this fact can be overcome by adopting alternative solutions, such as letting devices treat interference from other users as noise during decoding, such that SIC does not need to be applied [17].

Power control in NOMA with SC and SIC

Power control is a critical feature of SC, since it directly impacts the achievable sum-rate and provides different data streams with different levels of protection against noise [63]. Unlike broadcast systems applications where information is grouped into data streams with different relative importance and where users may have similar channel conditions [81], in NOMA systems there are multiple users with large channel gain differences, and each user must be able to decode its own data stream. Therefore, power control in NOMA must guarantee that each user can overcome the inter-user interference introduced by SC, and decode its own data. This is achieved in NOMA by allocating users with a weaker channel condition with a greater share of power (i.e., with a larger power allocation factor).

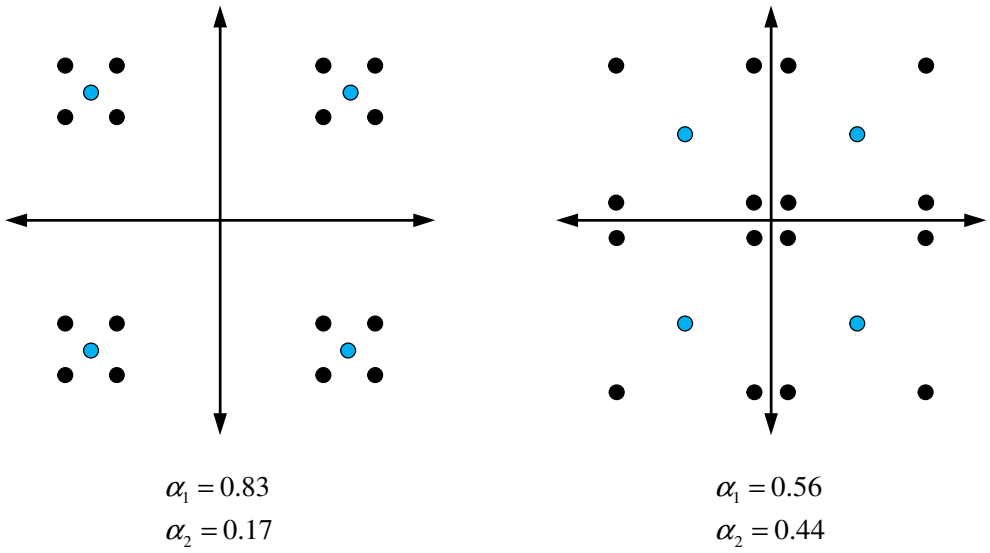


Figure 2.5: Effect of the power allocation factor on the inter-user interference in superposed 4-QAM + 4-QAM constellations.

Figure 2.5 represents the effect of SC power allocation in NOMA for the case where two 4-QAM symbol constellations are superposed. The diagram on the left represents the case where the lower-layer user is assigned with a 0.83% of the transmit power. The diagram on the right represents the situation where the lower-layer user is assigned with a 0.56% of the transmit power. During the first stage of SIC, a 4-QAM symbol is decoded. The axes represent the different decision region boundaries, with symbols situated next to the axes yielding the largest decoding error probability. Inter-user interference is larger for the diagram on the right, since more power is allocated to the upper constellation layer. It is observed that the Euclidean distances between the axis and adjacent symbols are relatively small, resulting in an increased probability of decoding error during the first stage of SIC, compared with the diagram on the left. Therefore, the power allocation factor in NOMA must be carefully selected in order to guarantee that users can overcome inter-user interference during decoding.

The capacity of NOMA with SC and SIC

From an information-theoretic perspective, NOMA with SC at the transmitter and SIC at the receiver is optimal in the sense of achieving the capacity region of the downlink broadcast channel [46], provided that each superposed layer is protected by a capacity-achieving code, and assuming optimal power allocation and perfect SIC [63].

Consider a downlink NOMA system with two users j and k, where $|h_j|^2 \leq |h_k|^2$, and therefore user j comes first in the SIC decoding order. User j can decode its own signal directly, disregarding user k's signal as noise [51]. Hence, it can theoretically achieve a maximum data rate of

$$R_j = \log_2 \left(1 + \frac{\alpha_j |h_j|^2 p}{\alpha_k |h_j|^2 p + N_0} \right). \tag{2.1}$$

User k decodes user j's signal first, with an achievable data rate of

$$R_{j \to k} = \log_2 \left(1 + \frac{\alpha_j |h_k|^2 p}{\alpha_k |h_k|^2 p + N_0} \right). \tag{2.2}$$

After decoding user j's signal, user k can subtract it from the receive signal, and it can decode its own symbols next. The theoretically maximum achievable data rate when decoding its own signal, when perfect SIC is achieved, is given by

$$R_{k,s} = \log_2 \left(1 + \alpha_k |h_k|^2 p \right). \tag{2.3}$$

The capacity of SC is bounded by the sum of individual rates of each information layer [63]. In SC, the distribution of the channel input may either be Gaussian, which is the typical assumption in theoretical studies, or any other type of constellation-constrained input. In any case, the channel capacity is a monotonically increasing function of the signal-to-interference noise ratio. How-

ever, in practical NOMA systems the achievable sum-rate is constrained by the use of constellation-constrained inputs with non-ideal channel codes, which necessarily yields lower data rates than those given by (2.1) and (2.3). Nevertheless, the benefits of NOMA with respect to OMA can be retained when employing non-ideal channel codes, through suitable resource allocation.

2.2.3 Multicarrier NOMA Systems

One of the main reasons why NOMA was initially proposed for inclusion in the 3GPP LTE-advanced standard [73] is due to its compatibility with OFDMA, which is based on OFDM. NOMA can be applied to a group of users on each OFDMA subcarrier without requiring any changes to LTE resource blocks [47], facilitating the integration of NOMA into OFDMA.

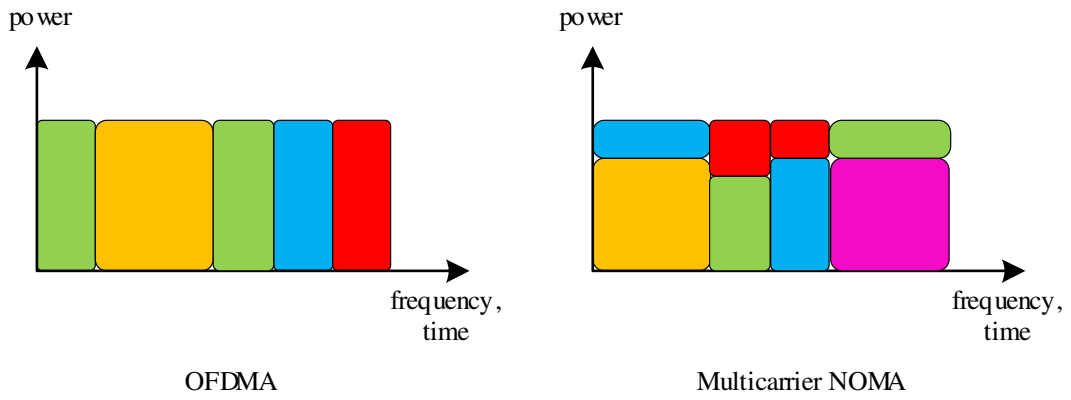


Figure 2.6: Orthogonal division of resource blocks in OFDMA and multicarrier NOMA.

Multicarrier NOMA systems [47] combine the benefits of OMA and NOMA. In multicarrier NOMA, users are divided into multiple pairs. Note that, in this thesis, only two-user NOMA is considered, due to the high implementation cost of SIC receivers [74], and therefore the terminology refers to user pairs rather than groups. However, the NOMA principle is also applicable to groups of more than two users.

In multicarrier NOMA, all users within a pair are served in the same resource block following the NOMA principle, and different resource blocks are orthogonal in the time and frequency domains. By employing multicarrier NOMA, the system complexity can be reduced with respect to a pure NOMA system where all users are multiplexed simultaneously. Therefore, multicarrier NOMA achieves a good trade-off between system performance and implementation complexity. Figure 2.6 represents the orthogonal division of resource blocks in OFDMA and in multicarrier NOMA.

In multicarrier NOMA, although different pairs of users are allocated to different resource blocks, the same user can be included into several different user pairs, and therefore be served in several resource blocks simultaneously. Intra-pair interference is mitigated by applying the NOMA principle, whereas inter-pair interference is avoided due to the orthogonality among subcarriers. As a result, the system can be overloaded, i.e. it can provide service to a larger number of users than the number of subcarriers, a requirement for enabling massive connectivity [47].

Appropriate resource allocation, which includes user pairing and power allocation, plays a critical role in the performance of multicarrier NOMA systems [52], [59]. While user pairing can reduce the computational complexity of executing SIC and improve user fairness, power allocation directly impacts the achievable throughput. Further, dynamic resource allocation allows for the real-time balancing of the system resources through adaptive variation of the system parameters such as the transmit power [75], and it can be applied to multicarrier NOMA to further improve the system performance. In addition, resource allocation improves the capability of supporting diverse applications [59] in multicarrier NOMA systems.

However, the joint optimization of user pairing and power allocation is an

NP-hard problem requiring of exhaustive search over all user pairs [3]. To avoid an excessively large computational complexity in resource allocation, there must exist an optimization criterion such that a trade-off between system performance and complexity can be achieved [52].

2.3 Conclusions

Multiple access techniques allow multiple users to share limited communication resources. Therefore, their design is critical in modern wireless communication systems, and specially in beyond-5G paradigms with ultra-dense number of users.

Wireless communications systems from 1G to 4G have adopted OMA techniques, where orthogonal resources are allocated to users in the time, frequency or code domain, or their combinations. OMA systems rely in the orthogonality principle to reduce inter-user interference, and multi-user detection can be achieved with low-complexity receivers. However, the number of supported users in OMA systems is limited by the number of orthogonal resources. NOMA systems can provide service to a larger number of users than OMA-based systems, thanks to the non-orthogonality principle, and have the potential to address the ultra-high densities of users forecast for beyond-5G paradigms. Further, NOMA with SC and SIC is capacity-achieving when employing ideal channel codes and assuming perfect SIC. When employing non-ideal channel codes and imperfect SIC, NOMA can still achieve larger data rates than OMA techniques.

However, the advantages of NOMA with respect to OMA come at a cost of increased implementation complexity. Moreover, power control in NOMA must be carefully designed in order to balance user fairness and achievable sum-rate, whilst limiting the effect of inter-user interference. Unsuitable power allocation can make signal decoding in NOMA unfeasible, due to SIC error propagation.

By combining NOMA with multicarrier schemes, a better trade-off between performance and implementation complexity can be achieved. Multicarrier NOMA can provide service to several users within a resource block, while maintaining the orthogonality among different resource blocks in the time and frequency domains. As a result, multicarrier NOMA systems can provide service to a number of users larger than the number of resource blocks. Appropriate resource allocation plays a critical role in the performance of multicarrier NOMA systems, and it must be optimized according to certain criteria such as throughput or user fairness.

Chapter 3

New Performance Bounds on the BER of NOMA Systems

Contents

3.1 Introduction

3.1.1 Motivation

Due to the potential advantages that NOMA introduces over OMA for beyond-5G networks, as presented in Section 1.1, the performance of NOMA systems has attracted extensive literature over the past few years. However, many of these works addressed research problems based on the perspective of the channel capacity and the system capacity, based on the Shannon formula [3,51]. In order for NOMA to fulfill the challenges of practical scenarios in beyond-5G systems, it is necessary to address some further research challenges.

Previous researches demonstrated that, in terms of achievable sum-rate, it is preferable to pair two NOMA users with very distinct channel conditions [51]. This result has since been applied to numerous researches. However, the application of this theoretical result might turn unrealistic in scenarios where some users have a very poor channel link. In a NOMA setting with SC and SIC, these users would have to be assigned most of the NOMA power share in order to enable successful decoding at the receiver. Such power allocations are inefficient in terms of the sum-rate because most of the power is allocated to a user with a very weak channel condition. In contrast, the user that enjoys a better channel condition and contributes most to the achievable sum rate, is allocated with a tiny fraction of the power allocation factor. Moreover, users with a very poor

channel condition might not be able to meet BER and quality of service constraints. Therefore, it is critical to understand what the limit is in the channel gain difference of two NOMA users in terms of BER and sum-rate, and how a certain ratio of user channel gains affects the achievable BER.

Previous works such as [3, 51] also assume perfect SIC at the receivers. However, in practical wireless communications systems, the use of SIC can result in error propagation as successive layers of information are removed at the receivers. This results in decoding errors being carried over from one decoding stage to the next, greatly impacting the achievable BERs at the receivers. This can, in turn, make the application of NOMA unfeasible. Therefore, it is of paramount importance to understand the impact of performance optimization parameters, such as the transmit power and the power allocation factor, on the achievable BER in NOMA.

Recently, some efforts have been devoted to studying the BER performance of NOMA systems [76, 77]. However, many of these works assumed fixed and equal modulation levels at the users. This is not applicable to many realistic scenarios where users have different data rate requirements. Therefore, it is critical to gain more knowledge about the attainable BER in practical NOMA systems, eg. in a scenario where users use different modulation levels, and to study the optimal adjustment of the power allocation factor and modulation level in BER-constrained NOMA systems.

3.1.2 Contributions

In this Chapter, several contributions are made in terms of proving the complex interplay between critical parameters in NOMA and the achievable BER, and new performance bounds on the BER of NOMA systems are provided.

First of all, theoretical BER expressions are presented for the BER in NOMA,

assuming multi-layer, multi-level QAM. To the best of the author's knowledge, this work represents the first attempt in developing such expressions, with previous literature focusing mainly on fixed-level modulation. It is proved that, for the user who comes second in the SIC decoding order –i.e. the user in the higher SC layer–, there exists a certain value of the power allocation factor for which its individual achievable BER is minimized. From the higher-layer user's perspective, the value of the power allocation factor must be high enough as to increase the power level of its own received signal, but it must also be low enough as to ensure successful decoding and removal of the lower-layer data stream. The optimal power allocation factor value is also the value that minimizes the transmit power in BER-constrained NOMA. It is proved that the optimal value of the power allocation factor is dependent on the modulation levels assigned to each user. Further, it is proved that, as the modulation level assigned to the lower-layer user increases, the value of the optimal power allocation factor rises sharply.

The theoretical BER expressions are used for calculating the optimal channel gain ratios for a pair of NOMA users, i.e. the ratios of channel gains that maximize the achievable sum-rate for a given BER constraint. Unlike previous research in NOMA [51], this work demonstrates that, in NOMA systems with practical QAM, the channel gains of two NOMA users must be of approximately the same order of magnitude in order to guarantee that the effect of inter-user interference can be overcome at the receivers, specially when the modulation levels assigned to both NOMA users are similar.

However, the derived BER expressions are in terms of products of the Q-function. These expressions are neither easily differentiable nor easily invertible in their arguments, and it becomes complicated to apply them to developing a resource allocation strategy. Therefore, accurate BER approximations are presented in the form of exponential functions. Last, the BER exponential approxi-

mations are used for finding numerical boundaries for the values of the channel gain ratios of NOMA users that fulfill BER constraints for different combinations of modulation levels; any pair of NOMA users whose channel gain ratio falls outside the numerical boundaries for a given modulation level cannot meet the individual BER constraints.

3.1.3 Structure

The rest of this Chapter is organized as follows. In Section 3.2, a relevant literature review is provided. In Section 3.3, the system model is presented. In Sections 3.4 and 3.5, BER expressions are presented, and these are used for deriving the optimal power allocation factors that minimize individual BERs, as well as the optimal channel gain ratio for different modulation levels. In Sections 3.6 and 3.7, exponential BER approximations are presented, and these are used for deriving the values of the channel gain gap conditions that guarantee the fulfillment of the BER constraint at both users. Numerical results are given in Section 3.8, and conclusions in Section 3.9.

3.2 Literature Review

The attainable individual BERs in NOMA are dependent on the power allocation factor at each user. It was proved in [51] that, by pairing NOMA users with very different channel qualities and allocating a significantly larger power allocation factor to the user with a poorer channel gain, NOMA is likely to outperform OMA techniques. When the power allocation factor for the user with a poorer channel gain decreases, it is still probable that NOMA will achieve a higher data rate than conventional OMA, at a cost of a significant performance degradation.

Unlike previous researches in SC [63, 67], constellation design in NOMA must

guarantee the decoding at each user through the combination of first, a suitable power allocation factor, and second, an appropriate modulation scheme [78, 79]. Some previous researches in SC considered the superposition of fixed constellations. In [63], the authors studied the superposition of a large number of binary phase shift keying (BPSK) signals under equal power allocation factors. In [80], the superposition of BPSK signals with unequal power allocation factors was investigated, whereas [67] considered SC with QPSK signals where a combination of unequal power allocation factor and phase shift was applied to superposed constellations.

Furthermore, the concept of SC in NOMA is also related to that of hierarchical modulations in broadcast systems, where information is grouped into data streams with different relative importance and different levels of protection against noise. The superposed constellation points are designed based on the proportional number of bits allocated to each data stream [81]. Previous research in hierarchical modulations in broadcast systems assumed systems models where all users have similar channel conditions. For example, in [82] and [68], exact BER expressions of hierarchical QPSK/M-QAM constellations were derived for 1-user systems. In NOMA systems, in contrast, there are multiple users with large differences in channel conditions. Independent symbols intended for different users are superposed in NOMA, and each user must be able to decode information contained within a certain layer through SIC. In addition, the use of a SIC receiver at the user allocated to the top superconstellation layer means that a decoding error at the low layer causes an error propagation when the top layer is decoded. Therefore, in NOMA systems, the BER at each user is indirectly influenced by the other user's channel conditions through the setting of the power allocation factor.

Some recent researches have considered the achievable BER in NOMA sys-

tems. However, most existing studies considered single-carrier NOMA and assumed fixed modulation levels. In [83], analytical BER expressions were derived for two-user NOMA with QPSK + BPSK constellations, under Rayleigh fading channels, and it was noted that some values of the power allocation factor might cause some BER performance degradation, but no insight was given on how to select the power allocation factor to optimize the BER. In [76], a BER analysis and study of the impact of power allocation in the BER was also carried out for fixed QPSK + BPSK, in two-user uplink NOMA. The work in [84] proposed a user grouping scheme for NOMA, and system simulations were used to compute the average BER over Rayleigh fading channels, under fixed QPSK modulation, but the effect of the power allocation factor on the BER was not studied. In [77], analytical BER expressions were also derived for the uplink, under QPSK + QPSK constellations over an AWGN channel. In [85], the BER performance of downlink NOMA over Nakagami-m flat fading channels was studied, for two-user and three-user NOMA. Exact BER expressions were derived, and these were used to evaluate the optimal power allocation under fairness and average BER constraints. However, only QPSK modulation and single-carrier NOMA were assumed. A more thorough analysis was presented in [86], where the effect of the BER on NOMA-enabled visible light communication systems was studied, focusing mainly on the BER in M-ary phase shift keying. However, this work assumed fixed allocation factor, and did not consider the BER in NOMA users with different QAM modulation levels. Although most of these works demonstrated that the channel gain gap of NOMA users has an impact on the BER, this was not analytically evaluated, and did not take into account the effect of varying the power allocation factor.

The work in [87] considered the optimization of the minimum BER in multi-carrier NOMA systems. The proposed scheme had a high computational comple-

xity and, although it attained a good BER performance, this was at the expense of a reduction in the achievable sum-rate. Further, although generalized QAM schemes were considered, the same level modulation was assumed at each user. Only perfect SIC was considered, and it was assumed that, for all users, their individual BER degenerates into the BER of regular QAM in OMA systems, which is an over-simplistic assumption. Therefore, these contributions cannot be easily translated to practical systems, where perfect SIC is unfeasible, and where more flexibility is required in assigning users with different modulation levels.

Therefore, most existing works did not thoroughly study the relationship between power allocation, user channel gains, and their impact on the individual BER in NOMA. Further, most works assumed fixed modulation levels at the NOMA users. It is important to gain further knowledge about the attainable BER in practical NOMA systems, eg. in a scenario where NOMA users use different modulation levels. It is also of paramount importance to consider the effect of imperfect SIC on the BER performance of NOMA, and the relationship between the imperfect SIC effect and the channel gains of NOMA users.

3.3 System Model

In this Chapter, a single-carrier downlink NOMA scenario with K users is considered, where it is assumed that only user j and user k are paired together to perform NOMA; user k enjoys a better channel condition than user j, and a different power level is allocated to each user in the power domain during transmission. It is assumed that the downlink channel gains are perfectly estimated by both users through pilot channels and perfect channel state information. Studying the impact of channel estimation error is not within the scope of this thesis, and [88–91] provide insight into its impact in NOMA systems.

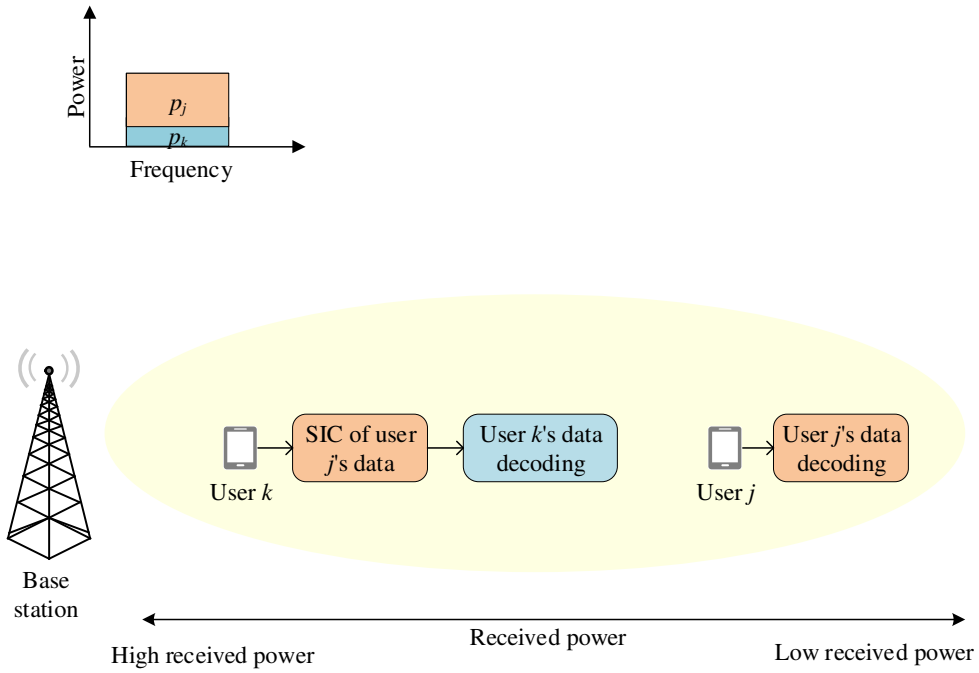


Figure 3.1: Two-user downlink NOMA scenario with power-domain superposition coding at the transmitter and successive interference cancellation at the receiver.

The system model under consideration is illustrated in Figure 3.1. At the transmitter, the multiplexing of both users' signals is achieved through the use of SC [63]. In the system model under consideration, the user who enjoys a better channel condition is allocated with less power than the user with a poorer channel condition. User demultiplexing is ensured through the allocation of an adequate power difference between users during transmission, and the application of SIC in the power-domain at the receiver of user k. User j comes first in the decoding order, and it does not need to apply SIC since it can decode its own signal directly.

The communication channel between the base station and the users is assumed to be a Rayleigh fading channel with additive white Gaussian noise (AWGN), with double-sided spectral density $N_0/2$. The channel response of user j is given by h_j, with $\mathbb{E}\{|h_j|^2\} = 1$. Similarly, the channel response of user k is given by h_k, with $\mathbb{E}\{|h_k|^2\} = 1$ It is assumed that $|h_j|$ and $|h_k|$ are independent and Rayleigh-distributed. Further, it is assumed that $|h_j|^2 \leq |h_k|^2$. The base station transmits

a signal of the form

$$x_{\{j,k\}} = \sqrt{p_j}x_j + \sqrt{p_k}x_k, \tag{3.1}$$

where $x_j \in \mathbb{C}$ and $x_k \in \mathbb{C}$ denote user j's and user k's transmit symbols, respectively. p_j and p_k are the transmit powers assigned to user j and user k, respectively. Further, $\mathbb{E}\{|x_j|^2\} = \mathbb{E}\{|x_k|^2\} = 1$, $p_j = \alpha_j p$ and $p_k = \alpha_k p$, where p is the total transmit power and α_j, α_k are the power allocation factors of user j and user k, respectively. The conditions $p_j + p_k = p$ and $\alpha_j + \alpha_k = 1$ hold. Hence, the received signals at user j and user k can be expressed as

$$y_j = \sqrt{p_j}h_j x_j + \sqrt{p_k}h_j x_k + z_j, \tag{3.2}$$

$$y_k = \sqrt{p_k}h_k x_k + \sqrt{p_j}h_k x_j + z_k, \tag{3.3}$$

respectively, where $z_j \sim \mathcal{CN}(0, \sigma_{z_j}^2)$, $z_k \sim \mathcal{CN}(0, \sigma_{z_k}^2)$ denote the AWGN.

The base station feeds forward information about modulation level assignment to both users through downlink control signaling. In addition, information about user j's assigned power and modulation level is also fed forward to user k, thus enabling SIC decoding at the receiver.

User j comes first in the SIC decoding order since $|h_j|^2 \leq |h_k|^2$, and it can decode its own signal directly, disregarding user k's signal as noise [51]. User k decodes user j's signal first, subtracts it from the receive signal, and decodes its own symbols next.

It is assumed that user j, uses maximum-likelihood detection with hard decision. At user k, the maximum-likelihood method is applied to decode user j's data stream. Then, user k uses SIC to remove interference from user j, where imperfect SIC is assumed, i.e. errors in detecting user j's signals are carried over to the second stage of detection, thus affecting the probability of correct detection of user k's own symbols. After removing user j's symbols, user k uses maximum-

likelihood detection with hard decision to decode its own data stream [92].

The instantaneous bit error rate (BER) at user j is represented as β_j, and that at user k is represented as β_k. β_0 denotes the individual BER constraint, which is assumed to be equal at both user j and user k.

3.4 BER Derivations

Below, a BER analysis is carried out for multi-layer, multi-level modulation in NOMA. Apart from proving the impact of the power allocation factor on the BER, this analysis also yields the optimal channel gain ratio in NOMA, i.e. the ratio of channel gains of the pair of NOMA users that maximize the sum-rate for a given BER constraint. Further, expressions (2.1) and (2.3) are the theoretical limits to the achievable data rates in NOMA under continuous modulation levels; the BER expressions derived below relate the achievable data rates in NOMA to the modulation levels of users and their BER constraint.

Note that the BER analysis below is carried out for the AWGN channel. Assuming that the instantaneous channel fading factors of the users are known at the receivers, the AWGN BER expressions are extensible to the instantaneous BER in Rayleigh fading channels, by applying a suitable channel fading scaling factor, according to the instantaneous channel gains of the users.

Consider the situation where two NOMA users j and k are multiplexed in the power domain. The base station transmits a superposed signal $x = x_{\{j,k\}}$ as given in (3.1). Let x_j be the symbol intended for user j, from a signal constellation $\mathcal{X}_j$ of M_j symbols, and let x_k be user k's symbol, from a constellation $\mathcal{X}_k$ of M_k symbols. The *supersymbol* $x = x_{\{j,k\}}$ is the result of superposing symbols x_j and x_k, where x_j belongs to the lower layer of the superposed constellation, and x_k to the upper layer. Further, supersymbol x belongs to a *superconstellation* (i.e.

a superposed constellation) $\mathcal{X}$ of size $M_j M_k$. The concept of superconstellations in NOMA is similar to that of hierarchical constellations in digital broadcasting systems [78, 82], where information is grouped into data streams with different relative importance. The location of the superconstellation points in NOMA is determined by the power allocation factors α_j and α_k. The average transmit symbol energy for the resulting superconstellation is normalized to $E_{sym} = 1$. As an example, Figure 3.2 represents the constellations associated to two superposed NOMA users, where user j's and user k's symbols belong to a 4-QAM constellation. The resulting superconstellation is formed by 16 supersymbols, whose Euclidean distances depend on α_j and α_k.

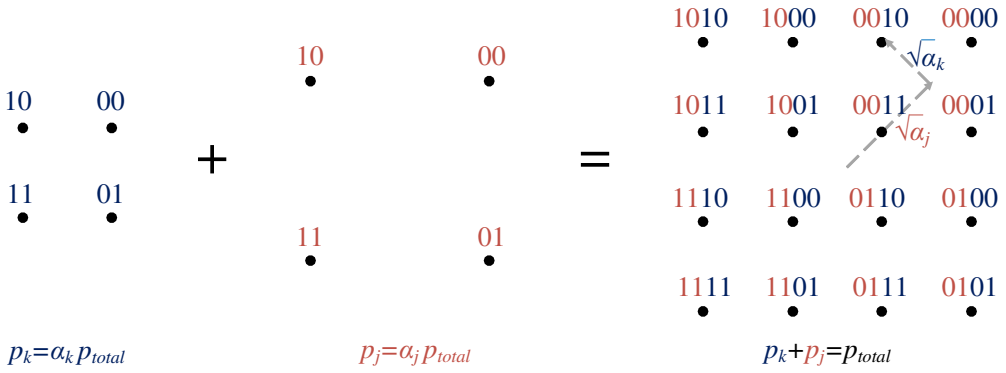


Figure 3.2: Conventional NOMA superposed 4-QAM+4-QAM constellation at transmitter with power allocation factor $\alpha_k = 0.2$.

To determine the symbol error rate (SER) at each user, the layout and number of superconstellation points must be considered, as well as the minimum Euclidean distances among them. The SER at user k is conditioned on the incoming interference from user j's signal, which can either reduce or increase user k's minimum Euclidean distance. User k uses an SIC receiver, so an error in decoding user j's symbol means that the decoding of user k's own signal is also unsuccessful, due to SIC error propagation. Hence, the total SER at user k, $P_k(e)$, is affected by the SER that user k achieves when decoding user j's transmit

symbols during the first stage of SIC. Therefore,

$$P_k(e) = P_k(e|correct_{x_j})P_k(correct_{x_j}) + P_k(e|error_{x_j})P_k(error_{x_j}), \tag{3.4}$$

where $P_k(correct_{x_j})$ is the probability that user j's symbols are correctly detected at user k, $P_k(error_{x_j})$ is the probability that an error occurs whilst detecting user j's symbols at user k, $P_k(e|correct_{x_j})$ is the probability of error at user k under the condition that no error occurred while detecting user j's symbols, and $P_k(e|error_{x_j})$ is the probability of error under the condition that an error occurred whilst detecting user j's symbols. Therefore, the term $P_k(e|error_{x_j})P_k(error_{x_j})$ models the SIC propagation error at user k.

Below, closed-form expressions are derived for $P_j(e)$ and $P_k(e)$. It is assumed that symbol errors in decoding are caused by incoming interference from the nearest neighbor, and that Gray coding is independently applied to each user's layer. Therefore, for small error rates, the BER at user j and user k can be calculated from the SER as $\beta_j \approx P_j(e)/\log_2 M_j$ and $\beta_k \approx P_k(e)/\log_2 M_k$, respectively, even in the event of SIC error propagation [62, 68]. It is assumed that the lower level constellation is assigned to user j in the lower layer, and that the average transmit power is always normalized to one for the NOMA superconstellation, i.e. $p = p_j + p_k = 1$. Note that, throughout the text, $E_{sym,j} = 1$ and $E_{sym,k} = 1$ are respectively user j's and user k's normalized symbol energies before NOMA power allocation and superposition. $E_{bit,j}$ and $E_{bit,k}$ refer to the average bit energy of user j and user k, respectively, where

$$E_{bit,j} = \frac{E_{sym,j}}{\log_2 M_j} = \frac{1}{\log_2 M_j}, \tag{3.5}$$

$$E_{bit,k} = \frac{E_{sym,k}}{\log_2 M_k} = \frac{1}{\log_2 M_k}. \tag{3.6}$$

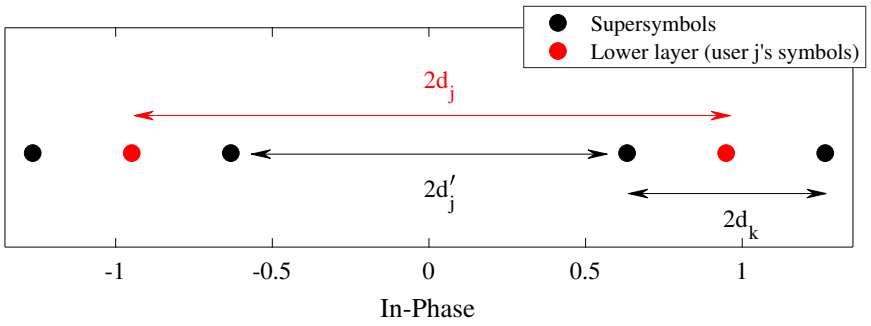


Figure 3.3: BPSK + BPSK superconstellation, where the average superconstellation transmit symbol energy is normalized to one.

3.4.1 BPSK + BPSK

Assume that both user j's and user k's signals are BPSK modulated. The resulting superconstellation is made up of four symbols of different amplitude in one dimension only, as represented in Figure 3.3. Let d_j and d_k represent the Euclidean distances of the layers belonging to user j's and user k's symbols, respectively, after power scaling by factors α_j and α_k. Further, let d'_j represent the largest Euclidean distance between symbols in the superconstellation, as represented in Figure 3.3. The set of Euclidean distances between superconstellation symbols is given by

$$d'_j = d_j - d_k, \tag{3.7}$$

$$d_j = \sqrt{\alpha_j E_{bit,j}}, \tag{3.8}$$

$$d_k = \sqrt{\alpha_k E_{bit,k}}. \tag{3.9}$$

User j's SER

User j's SER is similar to that of a conventional BPSK constellation (see [93]), but the effect of interference from user k's symbols must be taken into account.

In the best case, symbols in the superconstellation are at a distance $d'_j + 2d_k$ from the decision region boundary. In the worst case, the distance is reduced to d'_j. By averaging out the effect of both possibilities, the SER at the receiver results

$$P_j(e) = 0.5Q\left(\sqrt{\frac{2}{N_0}}d'_j\right) + 0.5Q\left(\sqrt{\frac{2}{N_0}}(d'_j + 2d_k)\right), \tag{3.10}$$

where $Q(x) = 1/\sqrt{2\pi}\int_x^\infty \exp\left(-u^2/2\right) \mathrm{d}u$ is known as the Q-function [93].

The error rate due to interference from user k's symbols at user j is given by

$$P_{k\to j}(e) = P_j(e) - Q\left(\sqrt{\frac{2}{N_0}}d_j\right), \tag{3.11}$$

where the term $Q\left(\sqrt{(2/N_0)}d_j\right)$ is equivalent to the SER of a BPSK constellation with a transmit power of $\alpha_j p$.

User k's SER

During the first stage of SIC decoding, user k decodes and removes user j's symbols from the received data stream. Errors in SIC are caused by symbols received on the wrong side of the lower layer decision boundaries. Therefore, the SER due to SIC errors at user k is given by

$$P_k(e|error_{x_j})P_k(error_{x_j}) = 0.5Q\left(\sqrt{\frac{2}{N_0}}d'_j\right) + 0.5Q\left(\sqrt{\frac{2}{N_0}}(d'_j + 2d_k)\right), \tag{3.12}$$

where it is assumed that $P_k(e|error_{x_j}) = 1$, i.e. an error in decoding the lower layer symbols at user k necessarily causes an error when user k decodes its own symbols during the second stage of SIC. If SIC is successful, user k decodes its own symbols from a BPSK constellation of transmit power $\alpha_k p$, yielding an error rate

$$P_k(e|correct_{x_j}) = Q\left(\sqrt{\frac{2}{N_0}}d_k\right). \tag{3.13}$$

Thus, from (3.4), (3.12) and (3.13), the total SER at user k results

$$P_k(e) = Q\left(\sqrt{\frac{2}{N_0}}d_k\right)\left(1 - 0.5Q\left(\sqrt{\frac{2}{N_0}}d'_j\right) - 0.5Q\left(\sqrt{\frac{2}{N_0}}(d'_j + 2d_k)\right)\right) + 0.5Q\left(\sqrt{\frac{2}{N_0}}d'_j\right) + 0.5Q\left(\sqrt{\frac{2}{N_0}}(d'_j + 2d_k)\right). \tag{3.14}$$

Power Allocation Factor Condition

In order to ensure that the BER constraints are simultaneously met at user j and user k, the selected power allocation factor must ensure a manageable level of inter-user interference at user j and successful SIC decoding at user k. Therefore, the Euclidean distance among symbols in the lower superconstellation layer must be equal or larger than that among symbols in the upper layer, i.e.

$$d'_j \geq d_k. \tag{3.15}$$

From (3.7)–(3.9), the values of α_k that minimize the BER fulfill the condition

$$\alpha_k \leq 0.2. \tag{3.16}$$

For the value $\alpha_k = 0.2$, the distances among all the superconstellation symbols are identical, and the superconstellation is identical to that of single-layer 4-pulse amplitude modulation (4-PAM). The effect of varying the value of α_k on the inter-user interference in BPSK + BPSK is represented in Figure 3.4. It can be observed that, for values of α_k that do not fulfill 3.16, there is a high error probability when decoding the symbols placed next to the origin.

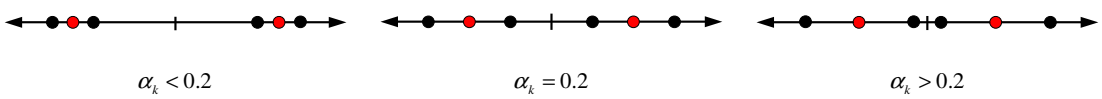


Figure 3.4: Effect of α_k on the inter-user interference in BPSK + BPSK, where the average superconstellation transmit symbol energy is normalized to one in all cases.

3.4.2 M_k-QAM + BPSK

In this case where user k is assigned with a square M_k-QAM constellation and user j with BPSK, the resulting superconstellation is not symmetrical in the in-phase and quadrature dimensions. User j's symbols are one-dimensional, whereas user k's layer is two-dimensional. User j will decode symbols from a BPSK constellation, with two decision regions contained within the in-phase dimension. User k will decode user j's symbols from a BPSK constellation first, and then it will decode its own symbols from a M_k-QAM constellation, with M_k decision regions. In general, for any M_k-QAM + BPSK constellation, the minimum Euclidean distance between user k's symbols is $d_{min,k} = 2d_k$, whereas for user j's symbols it is $d_{min,j} = 2(d'_j + 2id_k)$, as represented in Figure 3.5 for a 16-QAM + BPSK constellation. Note that, due to interference from user k, the minimum distance between user j's symbols in adjacent quadrants can be reduced or increased with respect to that in the originating constellation, i.e. $2d_j$. The following relation can be established among the set of distances:

$$d'_j = d_j - \left(\sqrt{M_k} - 1\right) d_k, \tag{3.17}$$

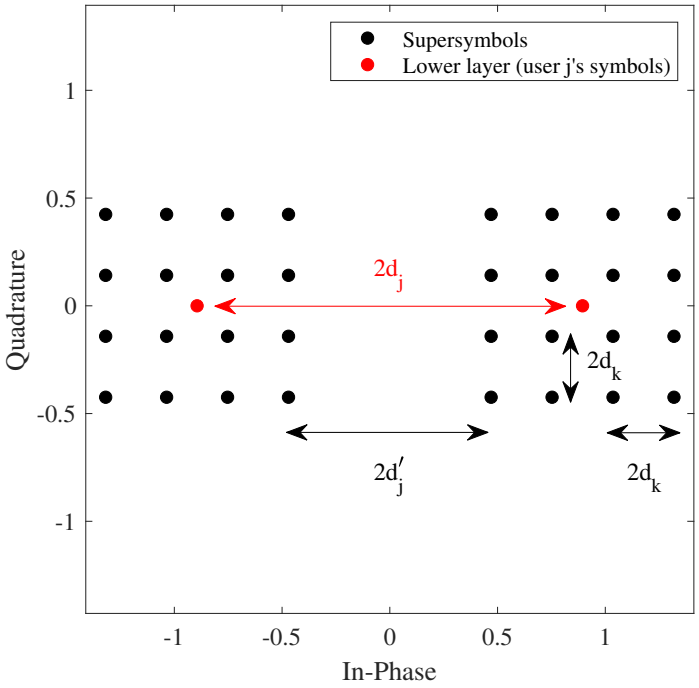


Figure 3.5: 16-QAM + BPSK superconstellation, where the average superconstellation transmit symbol energy is normalized to one.

where

$$d_j = \sqrt{\alpha_j E_{bit,j}} = \sqrt{\alpha_j E_{sym,j}}, \text{ and} \tag{3.18}$$

$$d_k = \sqrt{\frac{1.5 \log_2 M_k}{M_k - 1} \alpha_k E_{bit,k}} = \sqrt{\frac{1.5}{M_k - 1} \alpha_k E_{sym,k}}. \tag{3.19}$$

User j's SER

In this case, $P_j(e)$ is that of a BPSK constellation where $d_{min,j}$ varies depending on the interference introduced by user k's symbols. In the worst case, when the transmit NOMA symbol is adjacent to the imaginary axis, $d_{min,j} = 2d'_j$. In the best case, when the NOMA constellation points are farthest from the origin, $d_{min,j} = 2(d'_j + 2d_k + \ldots + 2d_k)$. By averaging out the corresponding symbol error

probabilities, the following expression is derived:

$$P_j(e) = P_j^{(I)}(e) = \frac{1}{\sqrt{M_k}} \sum_{i=0}^{\sqrt{M_k}-1} Q\left(2\frac{d'_j + 2id_k}{\sqrt{2N_0}}\right), \tag{3.20}$$

and the level of errors due to interference from user k's symbols can be calculated by replacing $P_j(e)$ into (3.11).

User k's SER

During the first stage of SIC decoding, user k decodes and removes user j's symbols from the data stream. Errors in SIC are caused by symbols received on the wrong side of the lower layer decision boundaries. Therefore, the SER due to SIC errors at user k is given by

$$P_k(e|error_{x_j})P_k(error_{x_j}) = \frac{1}{\sqrt{M_k}} \sum_{i=0}^{\sqrt{M_k}-1} Q\left(2\frac{d'_j + 2id_k}{\sqrt{2N_0}}\right), \tag{3.21}$$

where it is assumed that $P_k(e|error_{x_j}) = 1$, i.e. an error in decoding the lower layer symbols at user k necessarily causes an error when user k decodes its own symbols.

If SIC is successful, user k decodes its own symbols from a M_k-QAM constellation of transmit power $\alpha_k p$, which is symmetrical in the in-phase and quadrature dimensions. This yields [62, 93]

$$P_k^{(I)}(e|correct_{x_j}) = P_k^{(Q)}(e|correct_{x_j}) = 2\left(1 - \frac{1}{\sqrt{M_k}}\right) Q\left(\sqrt{\frac{2}{N_0}d_k}\right), \tag{3.22}$$

where $P_k^{(I)}(e|correct_{x_j})$ and $P_k^{(Q)}(e|correct_{x_j})$ are the error probabilities at user k when decoding its own symbols after SIC in the in-phase and quadrature dimensions, respectively. Therefore, from (3.4), the error probability at user k in

the in-phase dimension is given by

$$P_k^{(I)}(e) \lessapprox 2\left(1-\frac{1}{\sqrt{M_k}}\right) Q\left(\sqrt{\frac{2}{N_0}d_k}\right)\left(1-\frac{1}{\sqrt{M_k}}\sum_{i=0}^{\sqrt{M_k}-1} Q\left(2\frac{d'_j+2id_k}{\sqrt{2N_0}}\right)\right) + \frac{1}{\sqrt{M_k}}\sum_{i=0}^{\sqrt{M_k}-1} Q\left(2\frac{d'_j+2id_k}{\sqrt{2N_0}}\right). \tag{3.23}$$

In the quadrature dimension, the error probability at user k is given by

$$P_k^{(Q)}(e) = 2\left(1-\frac{1}{\sqrt{M_k}}\right) Q\left(\sqrt{\frac{2}{N_0}d_k}\right), \tag{3.24}$$

and the overall SER at user k for M_k-QAM + BPSK is found by replacing (3.23) and (3.24) into [93]

$$P_k(e) = 1-\left(1-P_k^{(I)}(e)\right)\left(1-P_k^{(Q)}(e)\right). \tag{3.25}$$

Power Allocation Factor Condition

In order to ensure that the BER constraints are simultaneously met at user j and user k, and that SIC can be successfully applied at user k, the Euclidean distances among symbols in user j's constellation layer need to be equal or larger than those among symbols in user k's constellation layer, i.e. condition (3.15) must hold. Therefore, the values of α_k that minimize the BER are analytically derived by substituting (3.17)–(3.19) into

$$\alpha_k \leq \frac{M_k-1}{2.5M_k-1}. \tag{3.26}$$

For the value $\alpha_k = (M_k-1)/(2.5M_k-1)$, the distances among all the super-constellation symbols are identical. The effect of varying the value of α_k on the

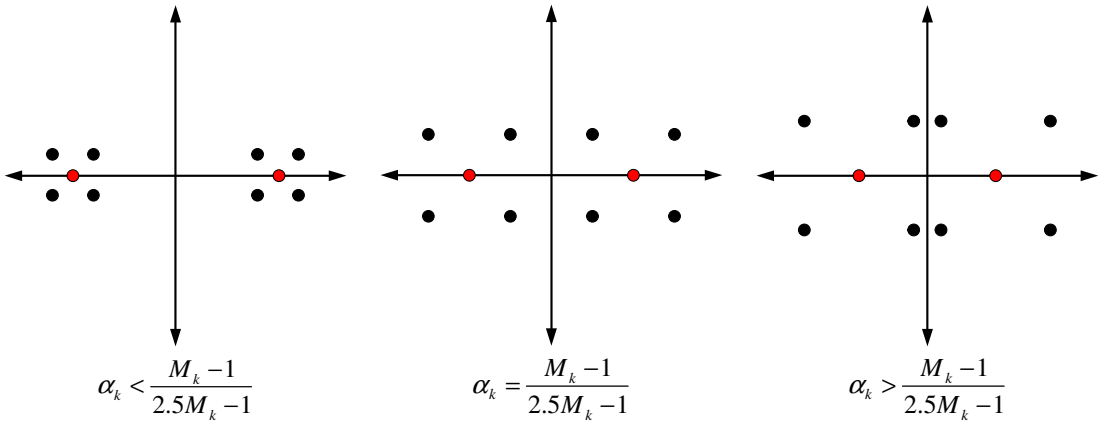


Figure 3.6: Effect of α_k on the inter-user interference in M_k-QAM + BPSK, where the average superconstellation transmit symbol energy is normalized to one in all cases.

inter-user interference in M_k-QAM + BPSK is represented in Figure 3.6. It can be observed that, for values of α_k that do not fulfill 3.26, there is a high error probability when decoding the symbols placed next to the origin.

3.4.3 M_k-QAM + M_j-QAM

Assume that user j is assigned with square M_j-QAM, and user k with square M_k-QAM. The superconstellation in this case is symmetrical in both dimensions, hence the in-phase and quadrature components have the same error probability. See, for a reference, the superconstellation points and Euclidean distances given in Figure 3.7, for a 16-QAM + 4-QAM superconstellation. The Euclidean distances expressions given in (3.17) and (3.19) still hold, whereas

$$d_j = \sqrt{\frac{1.5 \log_2 M_j}{M_j - 1} \alpha_j E_{bit,j}} = \sqrt{\frac{1.5}{M_j - 1} \alpha_j E_{sym,j}}. \tag{3.27}$$

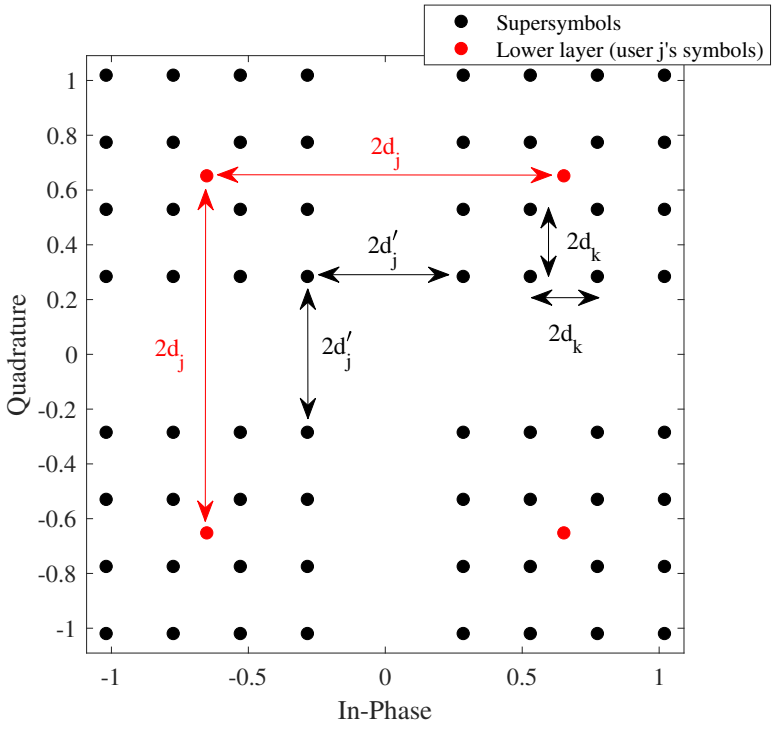


Figure 3.7: 16-QAM + 4-QAM superconstellation, where the average superconstellation transmit symbol energy is normalized to one.

User j's SER

For clarity, the in-phase dimension is considered first. In the lower superconstellation layer, there are $\sqrt{M_j}$ decision regions from which user j decodes its symbols. There are two outer decision regions which are limited by only one decision boundary. In addition, there are $\sqrt{M_j}-2$ decision regions enclosed within two decision boundaries. Enclosed within each decision region is a M_k-QAM constellation, which represents user k's symbols. In terms of decoding at user j, each upper constellation point introduces a certain amount of interference, due to the varying distance to user j's decision boundary. This results in a certain error rate during decoding. Averaging out the corresponding error rate per branch yields

$$P_j^{(I)}(e) = P_j^{(Q)}(e) = \frac{2\sqrt{M_j}-2}{\sqrt{M_j M_k}} \sum_{i=0}^{\sqrt{M_k}-1} Q\left(2\frac{d'_j + 2id_k}{\sqrt{2N_0}}\right), \tag{3.28}$$

due to in-phase and quadrature symmetry. The joint SER is given by

$$P_j(e) = 1 - \left[1 - P_j^{(I)}(e)\right]^2 = 1 - \left[1 - P_j^{(Q)}(e)\right]^2. \tag{3.29}$$

Substituting (3.28) into (3.29) yields

$$P_j(e) = 1 - \left[1 - \frac{2\sqrt{M_j} - 2}{\sqrt{M_j M_k}} \sum_{i=0}^{\sqrt{M_k}-1} Q\left(2\frac{d'_j + 2id_k}{\sqrt{2N_0}}\right)\right]^2. \tag{3.30}$$

The level of error due to interference from user k's symbols is given by

$$P_{k\to j}(e) = P_j(e) - 2Q\left(\sqrt{\frac{2}{N_0}}d_j\right), \tag{3.31}$$

where $2Q(\sqrt{(2/N_0)}d_j)$ is the SER of an M_j-QAM constellation with transmit power $\alpha_j p$.

User k's SER

During the first stage of SIC decoding, user k removes user j's symbols first. Errors in SIC are caused by symbols received on the wrong side of the lower layer decision boundaries. Therefore, the SER due to SIC errors in the in-phase dimension is given by

$$P_k^{(I)}(e|error_{x_j})P_k^{(I)}(error_{x_j}) = \frac{2\sqrt{M_j} - 2}{\sqrt{M_j M_k}} \sum_{i=0}^{\sqrt{M_k}-1} Q\left(2\frac{d'_j + 2id_k}{\sqrt{2N_0}}\right), \tag{3.32}$$

where it is assumed that $P_k^{(I)}(e|error_{x_j}) = 1$, i.e., an error in decoding user j's symbols at user k necessarily causes an error when user k decodes its own symbols. If SIC is successful, user k decodes its own symbols from an M_k-QAM

constellation, yielding [75]

$$P_k^{(I)}(e|correct_{x_j}) = 2\left(1-\frac{1}{\sqrt{M_k}}\right)Q\left(\sqrt{\frac{2}{N_0}}d_k\right). \tag{3.33}$$

Thus, from (3.4),

$$\begin{aligned} P_k^{(I)}(e) \lessapprox & \\ & 2\left(1-\frac{1}{\sqrt{M_k}}\right)Q\left(\sqrt{\frac{2}{N_0}}d_k\right)\left(1-\frac{2\sqrt{M_j}-2}{\sqrt{M_jM_k}}\sum_{i=0}^{\sqrt{M_k}-1}Q\left(2\frac{d_j'+2id_k}{\sqrt{2N_0}}\right)\right) \\ & +\frac{2\sqrt{M_j}-2}{\sqrt{M_jM_k}}\sum_{i=0}^{\sqrt{M_k}-1}Q\left(2\frac{d_j'+2id_k}{\sqrt{2N_0}}\right). \end{aligned} \tag{3.34}$$

As the SER on each dimension is identical, $P_k^{(I)}(e) = P_k^{(Q)}(e)$, and the joint SER can be calculated by substituting (3.34) into

$$P_k(e) = 1-\left[1-P_k^{(I)}(e)\right]^2 = 1-\left[1-P_k^{(Q)}(e)\right]^2, \tag{3.35}$$

which yields

$$\begin{aligned} P_k(e) \lessapprox 1- & \\ & \left[1-2\left(1-\frac{1}{\sqrt{M_k}}\right)Q\left(\sqrt{\frac{2}{N_0}}d_k\right)\left(1-\frac{2\sqrt{M_j}-2}{\sqrt{M_jM_k}}\sum_{i=0}^{\sqrt{M_k}-1}Q\left(2\frac{d_j'+2id_k}{\sqrt{2N_0}}\right)\right)\right. \\ & \left.-\frac{2\sqrt{M_j}-2}{\sqrt{M_jM_k}}\sum_{i=0}^{\sqrt{M_k}-1}Q\left(2\frac{d_j'+2id_k}{\sqrt{2N_0}}\right)\right]^2. \end{aligned} \tag{3.36}$$

Power Allocation Factor Condition

In order to ensure a manageable level of inter-user interference at user j and successful SIC decoding at user k, the distance among symbols in the lower superconstellation layer must be equal or greater than that among symbols in the upper layer, i.e. $d'_j \geq d_k$. From (3.17), (3.19) and (3.27), the optimal values of α_k that minimize the BER are analytically derived below in terms of M_j and M_k as

$$\alpha_k \leq \frac{M_k - 1}{M_j M_k - 1}. \tag{3.37}$$

Equivalently, the value of α_k for which the BER constraint can be met with minimum transmit power always fulfills (3.37), as otherwise interference in the lower layer would increase, resulting in a larger SIC error rate at user k due to larger inter-user interference. Under the assumption that the lower modulation level is always assigned to user j in the lower superconstellation layer, condition (3.37) guarantees that the effect of inter-user interference can be overcome.

For the value $\alpha_k = (M_k - 1)/(M_j M_k - 1)$, the distances among all the superconstellation symbols are identical. The effect of varying the value of α_k on the inter-user interference in M_k-QAM + M_j-QAM is represented in Figure 3.8. It can be observed that, for values of α_k that do not fulfill 3.37, there is a high error probability when decoding the symbols placed next to the axes, due to inter-user interference.

For 4-QAM + 4-QAM, the value of α_k obtained from (3.37) is identical to the value obtained in (3.16), i.e. $\alpha_k \leq 0.2$, since (3.17), (3.19) and (3.27) yield the same values of d_j, d_k and d'_j as (3.7)–(3.9) for modulation levels $M_j = M_k = 2$.

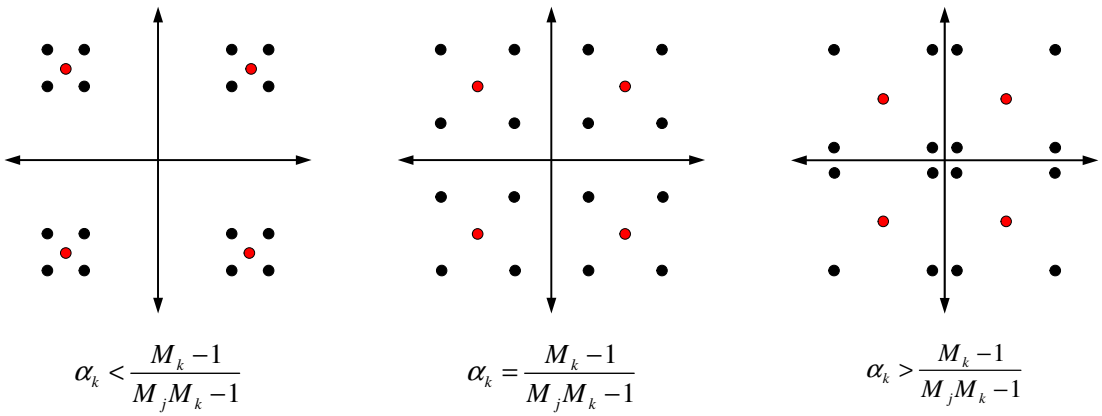


Figure 3.8: Effect of α_k on the inter-user interference in M_k-QAM + M_j-QAM, where the average superconstellation transmit symbol energy is normalized to one in all cases.

3.5 Optimal Channel Gain Ratios

Having prior knowledge about the optimal ratio between the channel gains of the two NOMA users is useful for simplifying the problem of user pairing in NOMA. Therefore, a channel gain ratio $g_{j,k}$ for user j and user k is defined as follows,

$$g_{\{j,k\}} = \frac{|h_k|^2}{|h_j|^2}. \tag{3.38}$$

Let $\mathcal{G}(M_j, M_k, \beta_0)$ denote the optimal channel gain ratio for two NOMA users, assuming modulation levels of M_j and M_k, and a BER constraint β_0. A pair of users $\{j, k\}$ is an optimal pair if

$$g_{j,k} = \mathcal{G}(M_j, M_k, \beta_0). \tag{3.39}$$

Numerical results in Table 3.1 show optimal channel gain ratio $\mathcal{G}(M_j, M_k, \beta_0)$ values calculated from the numerical evaluation of the analytical expressions (3.30) and (3.36), conditioned on the channel fading factors of users j and k, i.e. $|h_j|$ and $|h_k|$, and assuming that the BER constraint and expression (3.37)

Optimal channel gain ratio $\mathcal{G}(M_j, M_k, \beta_0)$				
	$\beta_0 = 10^{-3}$	$\beta_0 = 10^{-4}$	$\beta_0 = 10^{-5}$	$\beta_0 = 10^{-6}$
4-QAM + 4-QAM	1.548	1.342	1.264	1.206
16-QAM + 4-QAM	1.760	1.456	1.324	1.253
64-QAM + 4-QAM	2.225	1.686	1.467	1.343
256-QAM + 4-QAM	3.167	2.060	1.685	1.500
16-QAM + 16-QAM	2.113	1.640	1.443	1.342
64-QAM + 16-QAM	2.732	1.886	1.604	1.342

Table 3.1: Optimal channel gain ratio $\mathcal{G}(M_j, M_k, \beta_0)$ for M_k-QAM + M_j-QAM.

are marginally met, i.e.

$$\beta_j = \beta_k = \beta_0, \tag{3.40}$$

and

$$\alpha_k = (M_k - 1)/(M_j M_k - 1). \tag{3.41}$$

The results from Table 3.1 can be interpreted as follows. Take, for example, a BER constraint $\beta_0 = 10^{-3}$ and a 4-QAM + 4-QAM superconstellation, where $\mathcal{G}(4, 4, 10^{-3}) = 1.548$. Any pair of NOMA users j and k that fulfill $g_{j,k} = 1.548$ can simultaneously meet β_0, provided that the transmit power is large enough. Moreover, if user k fulfills β_0 for a given transmit power and modulation level, any user j with a channel gain $|h_j^s|^2 \geq |h_k^s|^2/g_{\{j,k\}}$ is guaranteed to fulfill β_0 too. In general, for any given BER constraint β_0, a larger difference between M_j and M_k requires a larger channel gain ratio between users j and k.

In [51], it was proved that the achievable data rate in NOMA is enlarged by pairing users with more distinctive channel gains, which is consistent with the results in Table 3.1. However, Table 3.1 also shows that there exists a limit on how distinctive the channel gains of two NOMA users can be in order for users with a poor channel condition to fulfill their individual BER constraint for a given modulation level. This fact is applied in Chapter 4, where the values in Table 3.1 are used to simplify the problem of user pairing.

3.6 BER Exponential Approximations

As the levels of M_j and M_k increase, the SER expressions (3.30) and (3.36) become more complicated, with additional Q-function terms. These expressions are neither easily invertible nor easily differentiable in their arguments, and therefore cannot be used for adaptive rate and modulation design. Therefore, instantaneous BER approximations are introduced with only one Q-function term. These approximations are used for finding the analytical relationship between the channel gain ratio of a pair of NOMA users and their achievable data rates. This knowledge is later applied in Chapter 4 to the development of the subcarrier, power and data rate allocation schemes.

Assume that the Euclidean distances condition given in (3.15) and the power allocation factor condition given in (3.37) are marginally met for the user pair $\{j, k\}$, i.e.

$$d'_j = d_k, \tag{3.42}$$

and

$$\alpha_k = \frac{M_k - 1}{M_j M_k - 1}. \tag{3.43}$$

In this case, it can be demonstrated that

$$Q\left(\sqrt{\frac{2}{N_0}} d_k\right) \lesssim \sum_{i=0}^{\sqrt{M_k}-1} Q\left(2\frac{d_k + 2id_k}{\sqrt{2N_0}}\right). \tag{3.44}$$

Therefore, according to (3.30), $P_j(e)$ is approximately proportional to the term $Q(\sqrt{2/N_0} d_{k,s})$, i.e.

$$P_j(e) \underset{\sim}{\propto} Q\left(\sqrt{\frac{2}{N_0}} d_k\right). \tag{3.45}$$

Further, the argument in all $Q(\cdot)$ terms in (3.36) is $\sqrt{2N_0}d_k$. Therefore,

$$P_k(e) \lessapprox 1- \left[1-2\left(1-\frac{1}{\sqrt{M_k}}\right)Q\left(\sqrt{\frac{2}{N_0}}d_k\right)\left(1-\frac{2\sqrt{M_j}-2}{\sqrt{M_jM_k}}Q\left(\sqrt{\frac{2}{N_0}}d_k\right)\right) - \frac{2\sqrt{M_j}-2}{\sqrt{M_jM_k}}Q\left(\sqrt{\frac{2}{N_0}}d_k\right)\right]^2. \tag{3.46}$$

After some algebraic manipulations, it is demonstrated that $P_k(e)$ is a quadratic polynomial function of the form

$$P_k(e) = aQ\left(\sqrt{\frac{2}{N_0}}d_k\right) + bQ^2\left(\sqrt{\frac{2}{N_0}}d_k\right) + cQ^4\left(\sqrt{\frac{2}{N_0}}d_k\right), \tag{3.47}$$

where a, b and c are some given polynomial coefficients in terms of M_j and M_k. For sufficiently small BER values, eg. for BERs smaller than 10^{-1},

$$Q\left(\sqrt{\frac{2}{N_0}}d_k\right) \ll Q^2\left(\sqrt{\frac{2}{N_0}}d_k\right) \ll Q^4\left(\sqrt{\frac{2}{N_0}}d_k\right). \tag{3.48}$$

Hence, from (3.47), it is demonstrated that $P_k(e)$ is also approximately proportional to the term $Q(\sqrt{2/N_0}d_k)$, i.e.

$$P_k(e) \underset{\sim}{\propto} Q\left(\sqrt{\frac{2}{N_0}}d_k\right). \tag{3.49}$$

Based on (3.45) and (3.49), and following eq. (9) in [75], it is found that the BERs at user j and user k in the AWGN channel can be approximated as

$$\beta_j \approx 0.14\frac{1.6+\sqrt{0.18\log_2 M_j \log_2 M_k}}{\sqrt{M_jM_k}}\exp\left(\frac{-1.54T_{sym}p}{N_0(M_jM_k-1)}\right), \tag{3.50}$$

and

$$\beta_k \approx 0.77 \frac{1.6 + \sqrt{0.18 \log_2 M_j \log_2 M_k}}{\sqrt{M_j M_k}} \exp\left(\frac{-1.54 T_{sym} p}{N_0 (M_j M_k - 1)}\right), \tag{3.51}$$

respectively, where T_{sym} is the symbol duration in seconds. Therefore, the instantaneous BERs at user j and user k under fading can be approximated as

$$\beta_j \approx 0.14 \frac{1.6 + \sqrt{0.18 \log_2 M_j \log_2 M_k}}{\sqrt{M_j M_k}} \exp\left(\frac{-1.54 |h_j|^2 T_{sym} p}{N_0 (M_j M_k - 1)}\right), \tag{3.52}$$

and

$$\beta_k \approx 0.77 \frac{1.6 + \sqrt{0.18 \log_2 M_j \log_2 M_k}}{\sqrt{M_j M_k}} \exp\left(\frac{-1.54 |h_k|^2 T_{sym} p}{N_0 (M_j M_k - 1)}\right), \tag{3.53}$$

respectively. The validity of the BER exponential approximations is demonstrated in Section 3.8 for the AWGN channel.

Expressions 3.50–3.53 demonstrate that it is possible to express the BER of NOMA users in an exponential form. These expressions are easily invertible and differentiable in their arguments, and can therefore be applied to the problem of resource allocation in NOMA under BER constraints. Similarly to exponential approximations to the BER in single-layer modulation [75], the value of the exponential function increases for higher modulation levels. However, the exponential approximations 3.50–3.53 are more complex than those given in [75] for single-layer modulation, as additional terms are added to 3.50–3.53 in order to model the effect of inter-user interference in NOMA.

3.7 Channel Gain Gap Condition

Next, expressions (3.52) and (3.53) are used for deriving a channel gain gap condition between NOMA users that maximizes the sum-rate whilst ensuring that the inter-user interference remains manageable for any given modulation level and BER constraint. For user k, the channel gain gap condition guarantees that the first stage of SIC can be carried out with an equal or lower probability of error than that of the last decoding stage, thus minimizing the probability of error propagation. For user j, it is ensured that interference from user k's symbols is low enough such that user j can decode its own data successfully.

Assume that the BER constraint β_0 is marginally met at both user j and user k, and that condition (3.43) is also met. In this case, (3.52) and (3.53) can be written in terms of R_j and R_k as

$$\exp\left(\frac{c|h_j|^2 p}{2^{R_j+R_k}-1}\right) = 7.14\beta_0 \cdot \frac{2^{0.5(R_j+R_k)}}{1.6+\sqrt{0.18R_jR_k}}, \tag{3.54}$$

and

$$\exp\left(\frac{c|h_k|^2 p}{2^{R_j+R_k}-1}\right) = 1.30\beta_0 \cdot \frac{2^{0.5(R_j+R_k)}}{1.6+\sqrt{0.18R_jR_k}}, \tag{3.55}$$

respectively, where $c = -1.54T_{sym}/N_0$. After some algebraic manipulations, the channel gains at user j and user k can be expressed in terms of R_j, R_k and β_0 as

$$|h_{u_j}|^2 = \ln\left(2.86\beta_0 \frac{2^{0.5(R_j+R_k)}}{1.6+\sqrt{0.18R_jR_k}}\right)\frac{2^{R_j+R_k}-1}{cp}, \tag{3.56}$$

and

$$|h_{u_k}|^2 = \ln\left(1.11\beta_0 \frac{2^{0.5(R_j+R_k)}}{1.6+\sqrt{0.18R_jR_k}}\right)\frac{2^{R_j+R_k}-1}{cp}, \tag{3.57}$$

respectively. It is clear that, for a fixed p and β_0, a larger data rate at either user j or user k requires larger a channel gain at both user j and user k. Further,

dividing (3.55) over (3.54) yields

$$\exp\left(\frac{cp\left(|h_k|^2-|h_j|^2\right)}{2^{R_j+R_k}-1}\right)=0.18. \tag{3.58}$$

Therefore,

$$H_{\{j,k\}}=\ln 0.18\cdot\frac{2^{R_j+R_k}-1}{cp}. \tag{3.59}$$

From (3.38), by writing $|h_j|^2$ as $|h_k|^2/g_{\{j,k\}}$,

$$H_{\{j,k\}}=|h_k|^2-|h_j|^2=|h_k|^2\left(1-\frac{1}{g_{\{j,k\}}}\right), \tag{3.60}$$

where $H_{\{j,k\}}$ is the channel gain gap of user j and user k.

From (3.59), it is clear that, for a fixed transmit power p, the sum-rate increases with an increasing value of the channel gain gap $H_{\{j,k\}}$. However, according to (3.60), $H_{\{j,k\}}$ is also upper-bounded by the channel gain ratio $g_{\{j,k\}}$, which must be small enough to guarantee that the BER constraint β_0 is simultaneously met at user j and user k for M_j and M_k, i.e.

$$g_{\{j,k\}}\leq\mathcal{G}(M_j,M_k,\beta_0). \tag{3.61}$$

In addition, $g_{\{j,k\}}$ must be large enough to maximize $H_{\{j,k\}}$ and, consequently, the achievable sum-rate, R_j+R_k. Consider the scenario where the BER constraint is marginally met at user k, and user j is selected as a partner to user k, such that the BER constraint is met by a larger margin. In this case, $P_j(e)\leq P_k(e)$. From (3.45) and (3.49), this condition translates into $d'_j|h_j|\geq d_k|h_k|$. By substituting (3.17), (3.19) and (3.27), this yields

$$g_{\{j,k\}}\geq 1,\forall\beta_0,M_j,M_k. \tag{3.62}$$

Therefore, from (3.60), $H_{\{j,k\}} \geq 0$. Hence, from (3.60), any two users j and k can be paired if and only if their individual channel condition gains jointly fulfill the channel gain ratio condition:

$$1 \leq g_{\{j,k\}} \leq \mathcal{G}(M_j, M_k, \beta_0), \tag{3.63}$$

for a BER constraint of β_0 and modulation levels M_j and M_k. The channel gain ratio condition (3.63) is applied in Chapter 4 for developing a subcarrier allocation algorithm with quasi-linear complexity.

3.8 Numerical Results

3.8.1 Impact of the Power Allocation Factor on the BER

Below, numerical results show the impact of varying the power allocation factor on the BER at both user j and user k. Note that, in all cases, the curves represent the BER in the AWGN channel; these curves are equivalent to those of the instantaneous BER under fading for $|h_j|= 1$ and $|h_k|= 1$.

BPSK + BPSK

Figure 3.9 represents the achievable BER at user j for BPSK + BPSK modulation, under a received SNR of 8dB and varying values of the power allocation factor. The BER at user j can be reduced only by increasing the share of power allocated to user j, α_j.

Figure 3.10 represents the BER at user k for BPSK + BPSK modulation, under a received SNR ratio of 15dB and varying values of the power allocation factor. There is an optimal value of the power allocation factor for which the BER is minimized at user k.

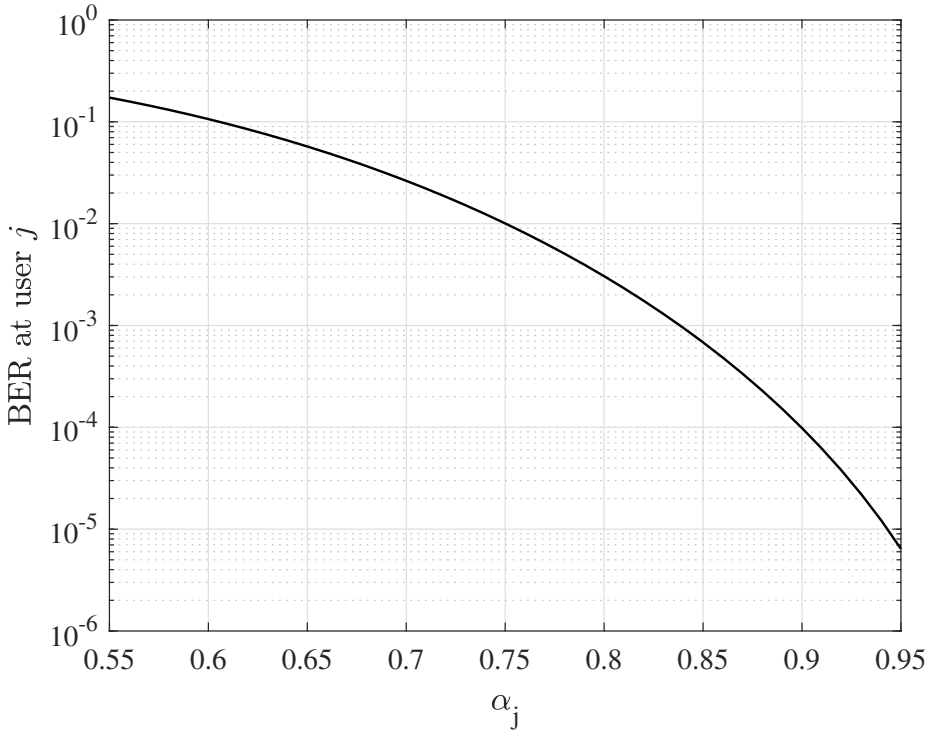


Figure 3.9: BER at user j for BPSK + BPSK under a received SNR of 8dB, for different values of the power allocation factor.

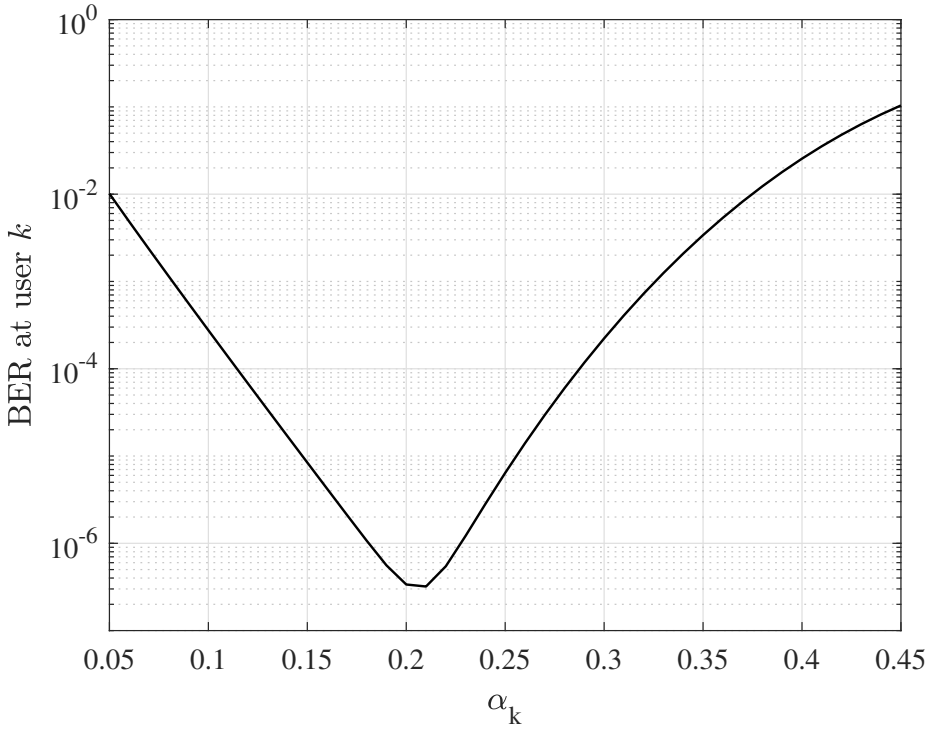


Figure 3.10: BER at user k for BPSK + BPSK under a received SNR of 20dB, for different values of the power allocation factor.

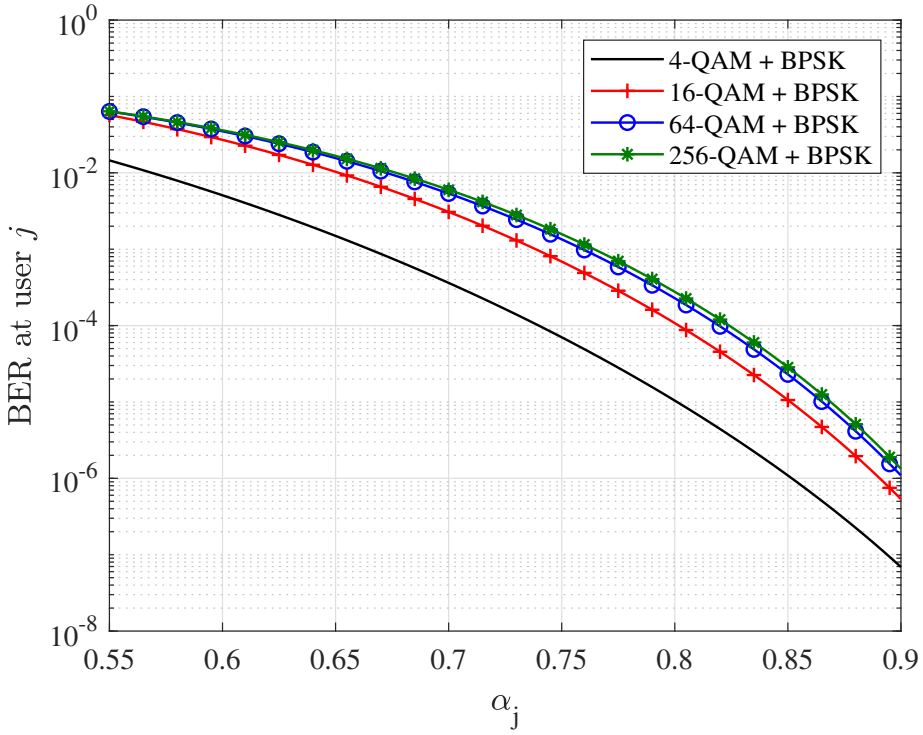


Figure 3.11: BER at user j for M_k-QAM + BPSK under a received SNR of 14dB, for different values of the power allocation factor.

At user j, according to Figure 3.9, the BER decreases when more transmit power is allocated to user j's symbols. At user k, according to Figure 3.10, the BER improves initially with an increasing power allocation factor too. However, due to the effect of inter-user interference, there is an optimal value of the power allocation factor for which the lowest BER is achieved. If the power allocation factor at user k increases, the BER deteriorates due to an increasing probability of error during the first stage of SIC. The lowest joint BER is achieved when the condition (3.16) is marginally met, i.e. when $\alpha_k = 0.2$. Further, condition (3.16) ensures that the BER constraint can be simultaneously met at user j and user k with minimum transmit power.

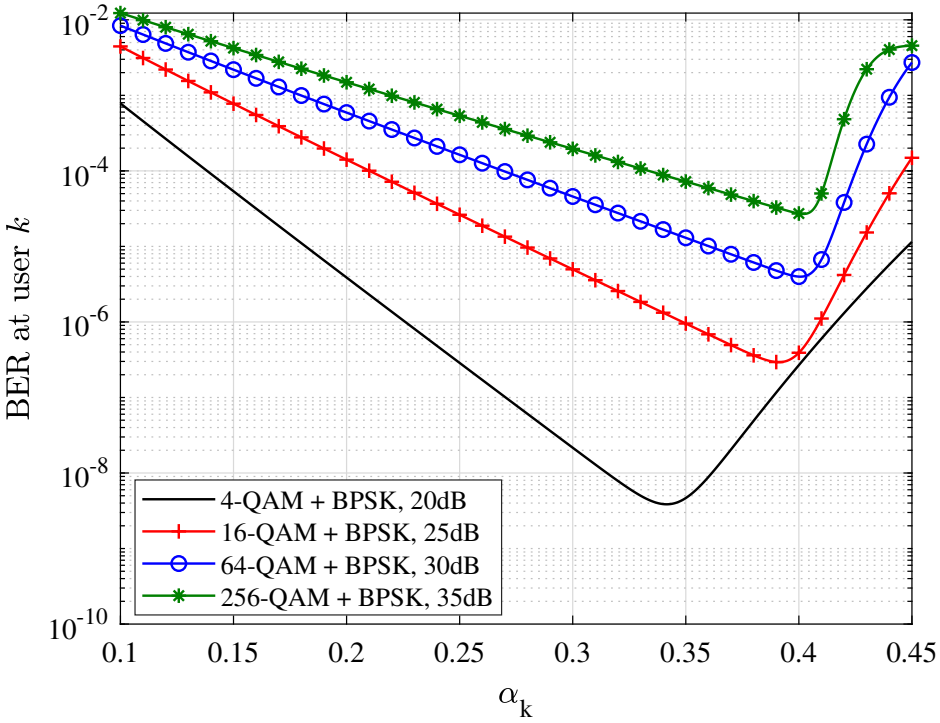


Figure 3.12: BER at user k for M_k-QAM + BPSK, for different values of the power allocation factor and received SNR.

M_k-QAM + BPSK

Figure 3.11 represents the BER at user j for M_k-QAM + BPSK, under a received SNR ratio of 16dB and varying values of the power allocation factor. It is observed that, for an increasing level of M_k, a higher share of power is required at user j in order to achieve a given BER value, due to an increased level of interference from user k's symbols.

Figure 3.12 shows the BER at user k for M_k-QAM + BPSK, for various levels of M_k and power allocation factor. It is observed that the minimum BER in each case is achieved for a value of the power allocation factor that increases with the modulation level M_k.

According to Figure 3.12, the minimum BER at user k is achieved for a value of α_k that increases with the modulation level M_k. Condition (3.26) guarantees that the effect of inter-user interference can be overcome at user k, and that the

BER constraint can be met with minimum transmit power.

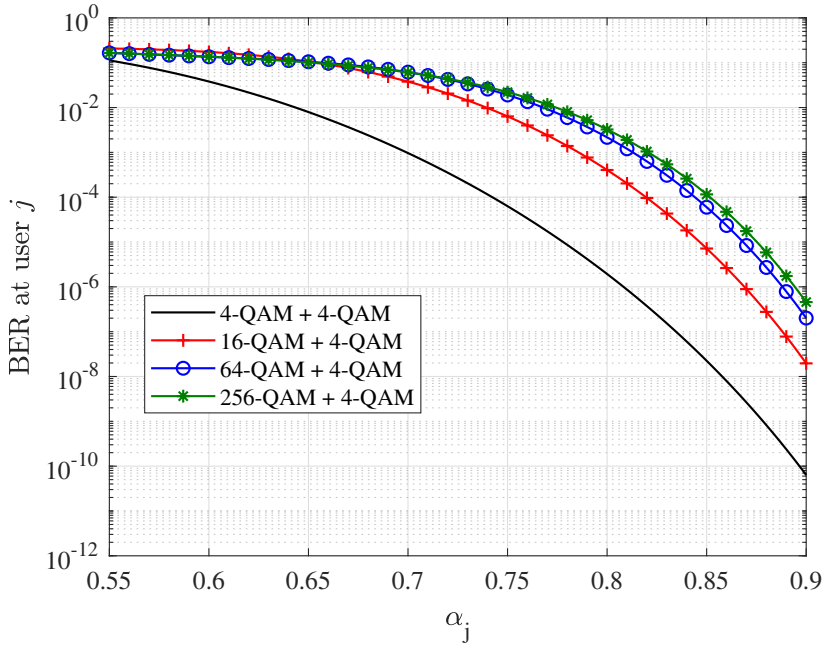
M_k-QAM + M_j-QAM

Figure 3.13 represents the BER at user j for M_k-QAM + 4-QAM and M_k-QAM + 16-QAM, under received SNRs of 20dB and 33dB respectively, for varying values of the power allocation factor. As M_j increases, the power share required at user j to overcome the effect of interference from user k's symbols rises sharply. This fact implies that, for two superconstellations of the same level (for example, 64-QAM + 4-QAM and 16-QAM + 16-QAM), a larger constellation level in the lower layer of the superconstellation requires a much better channel condition at user k, since a very large share of power is allocated to user j in order to overcome the effect of inter-user interference at both user j and user k.

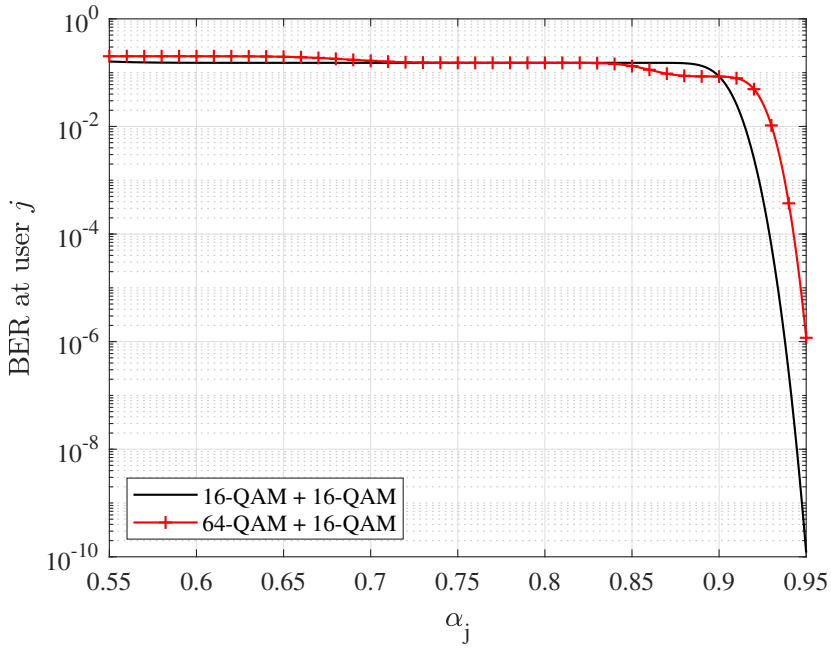
By comparing Figure 3.13a and Figure 3.13b, the required channel condition at user j to fulfill its BER constraint is significantly larger for a higher level of M_j. Therefore, in this case, the channel gap between the two NOMA users is significantly lower than in the case of superconstellations with a lower M_j level.

Figure 3.14 shows the BER at user k, for various levels of M_k and received SNR, for fixed $M_j = 4$ and $M_j = 16$. The effect of increasing M_j that was presented in Figure 3.13 for user j is also observed at user k. Figures 3.14a and 3.14b demonstrate that a smaller α_k is required when the level of M_j increases, in order to guarantee that user k can successfully carry out SIC. However, in order to guarantee that the BER constraint is met at user k, the value of α_k must increase when the level M_k increases.

Figure 3.13b and Figure 3.14b are good representations of the effect of the inter-user interference in the BER, which becomes more critical for higher modulation levels. The curves show dips in the BER for both user j and user k, for several values of the power allocation factor. At these values, there is a reduction

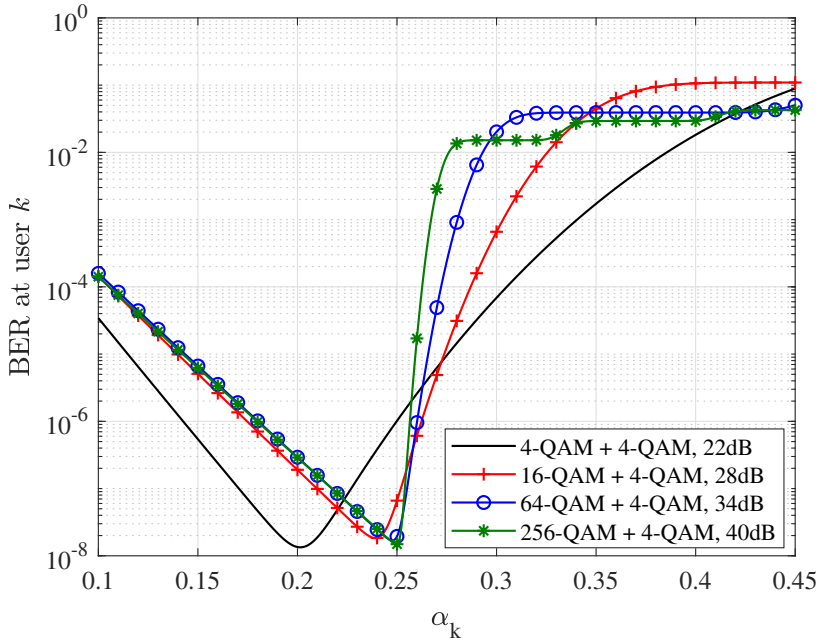


(a) BER at user j for M_k-QAM + 4-QAM under a received SNR of 20dB, for different values of the power allocation factor.

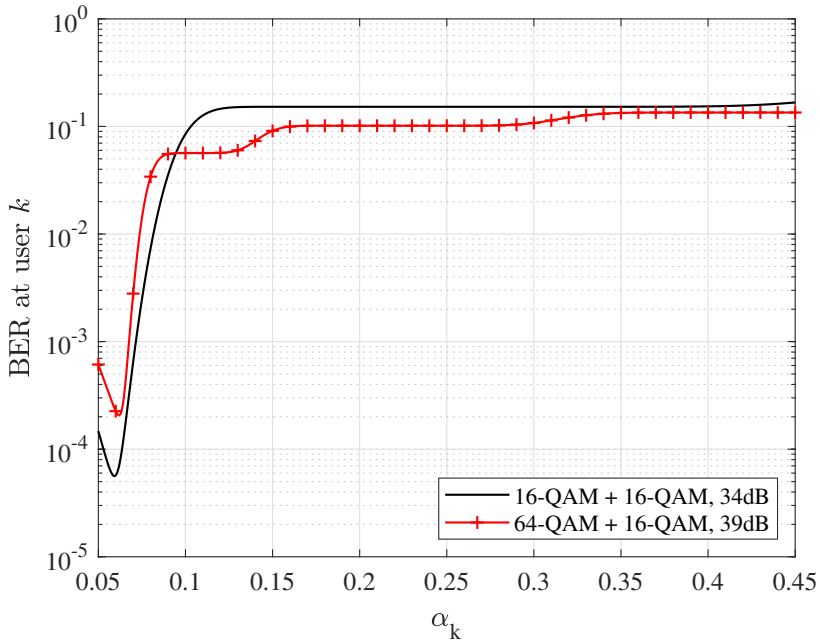


(b) BER at user j for M_k-QAM + 16-QAM under a received SNR of 33dB, for different values of the power allocation factor.

Figure 3.13: BER at user j for M_k-QAM + M_j-QAM.



(a) BER at user k for M_k-QAM + 4-QAM, under different values of received SNR, for different values of the power allocation factor.



(b) BER at user k for M_k-QAM + 16-QAM, under different values of received SNR, for different values of the power allocation factor.

Figure 3.14: BER at user k for M_k-QAM + M_j-QAM.

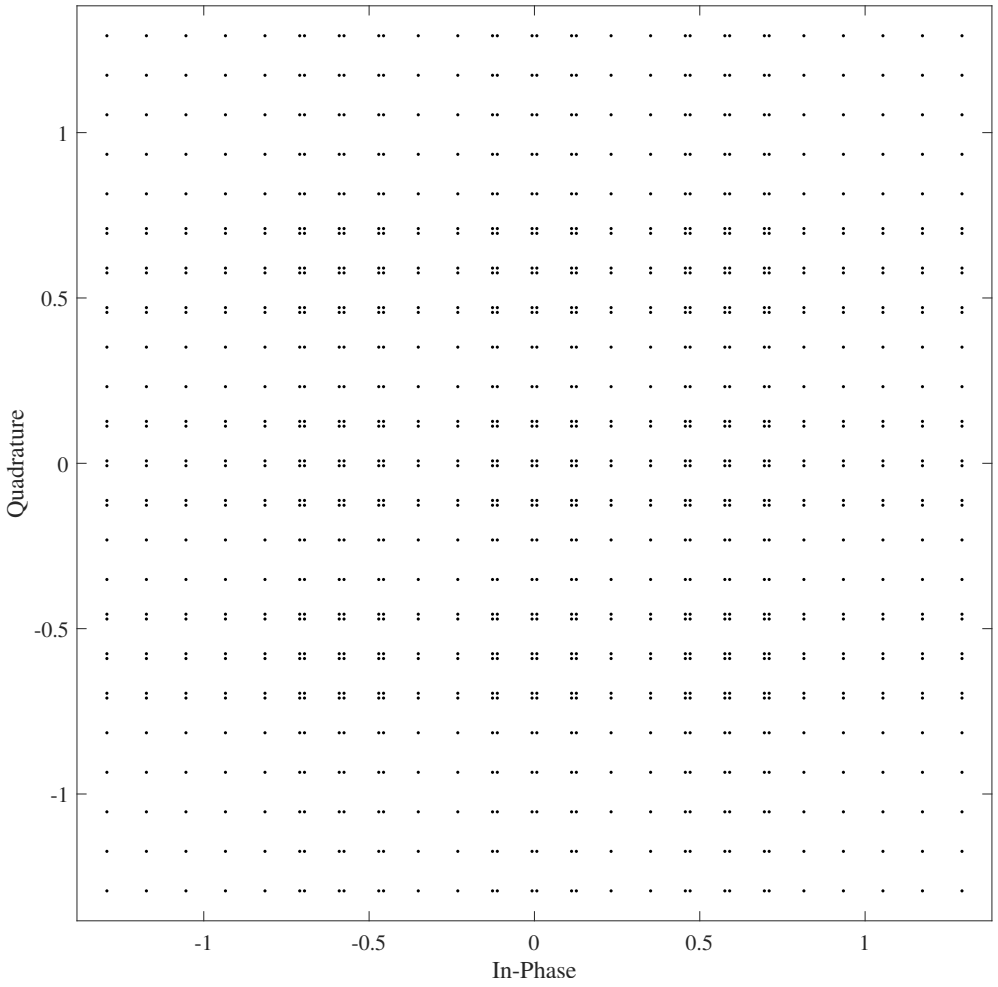


Figure 3.15: 64-QAM + 16-QAM superconstellation for $\alpha_k = 0.15$.

of the number of superconstellation symbols that cannot be correctly decoded due to inter-user interference. This effect is clearly illustrated in Figure 3.15 and Figure 3.16, where it can be observed a reduction of over 50% in the number of symbols for which inter-user interference is too large. In Figure 3.13b and Figure 3.14b this reduction translates into a dip in the BER curves.

3.8.2 Validity of the BER Exponential Approximations

The accuracy of expressions (3.52) and (3.53) is shown in Figures 3.17–3.20 for some modulation levels and for varying SNR. Note that, in all cases, the curves

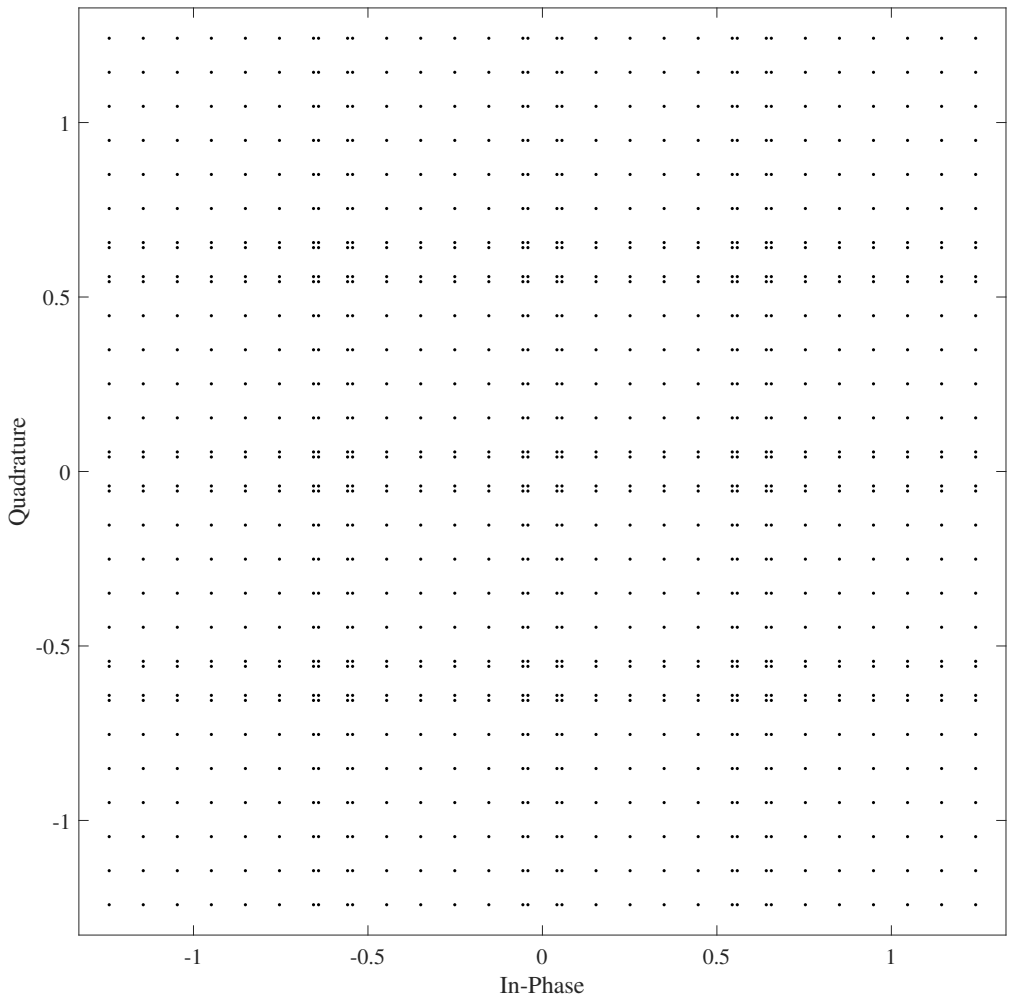
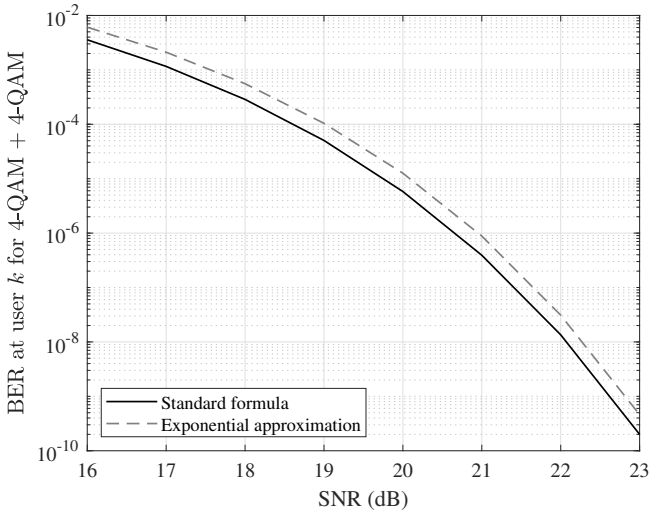
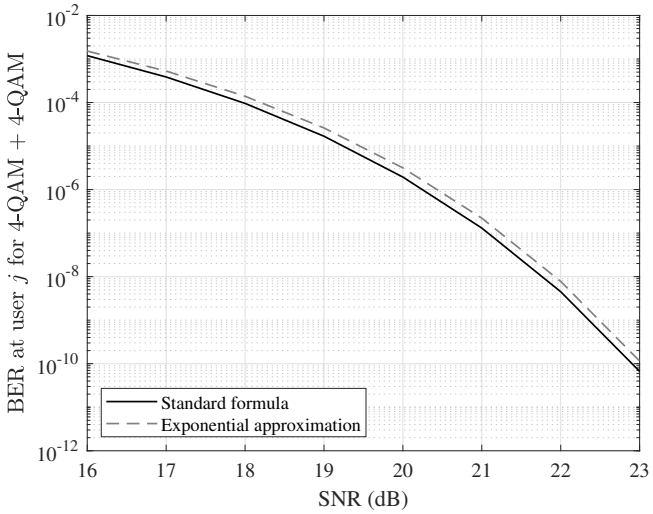


Figure 3.16: 64-QAM + 16-QAM superconstellation for $\alpha_k = 0.1$.



(a) BER at user k



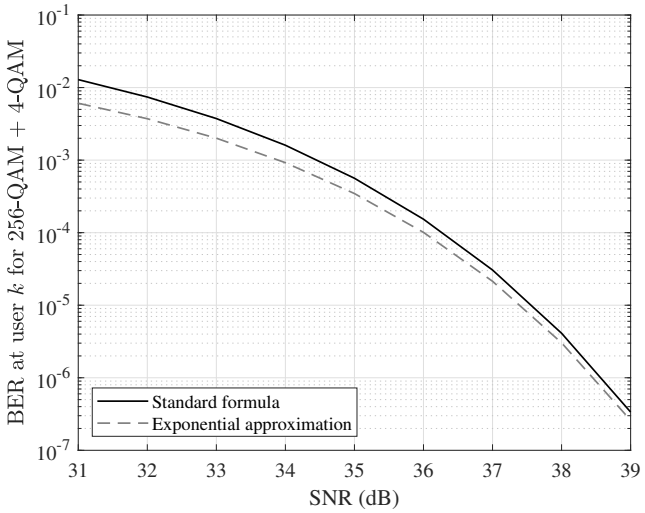
(b) BER at user j

Figure 3.17: BER exponential approximations for 4-QAM + 4-QAM.

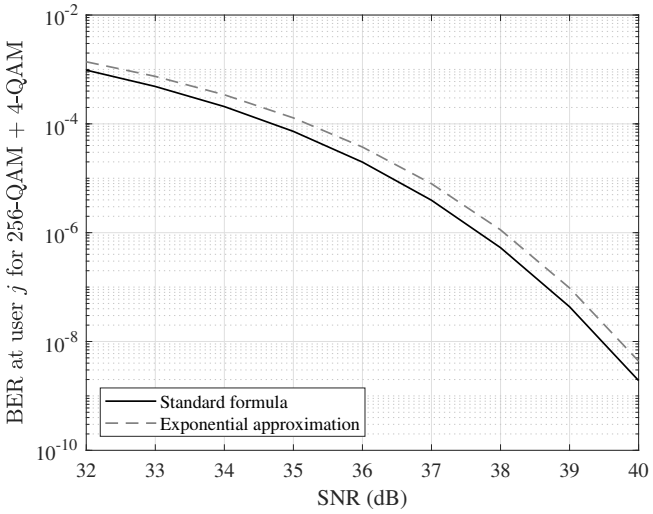
represent the BER in the AWGN channel; these curves are equivalent to those of the instantaneous BER under fading when the channel gains of users j and k are given by $|h_j|= 1$ and $|h_k|= 1$, respectively.

Figure 3.17 represents the validity of the exponential approximations for 4-QAM + 4-QAM, for both NOMA users. For user k, the exponential approximation becomes more accurate for higher transmit power and lower target BER.

However, at user j, the accuracy of the exponential approximation remains approximately constant within the transmit power interval of interest.



(a) BER at user k

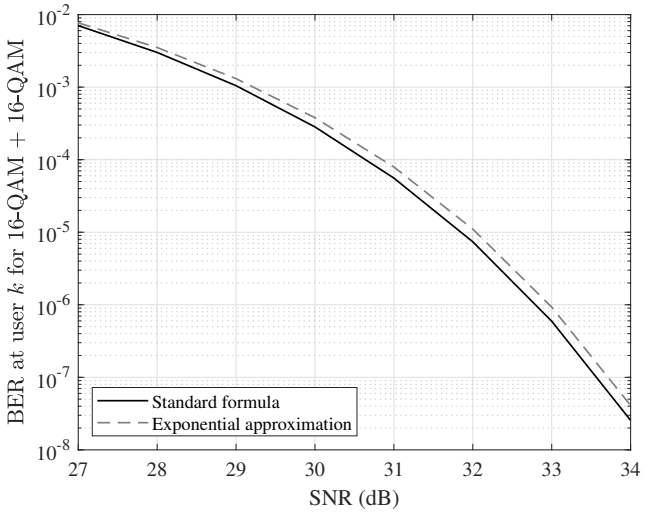


(b) BER at user j

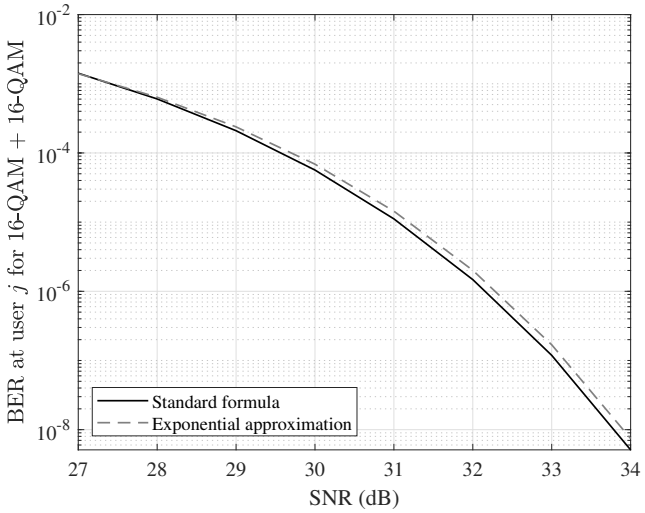
Figure 3.18: BER exponential approximations for 256-QAM + 4-QAM.

Figure 3.18 represents the accuracy of the exponential approximations for 256-QAM + 4-QAM. At user k, the approximated BER is below the theoretical value. Moreover, the curvature of the approximated BER differs from that of the theoretical curve. However, the accuracy still remains good and it is within

0.6dB. At user j, the accuracy of the approximation worsens with respect to the result obtained in Figure 3.17. The exponential approximation (3.52) is accurate to within 0.51 dB with respect to the standard formula (3.30).



(a) BER at user k

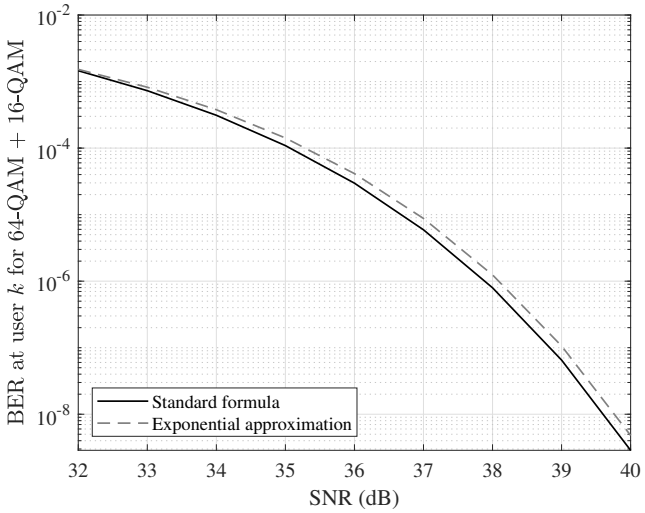


(b) BER at user j

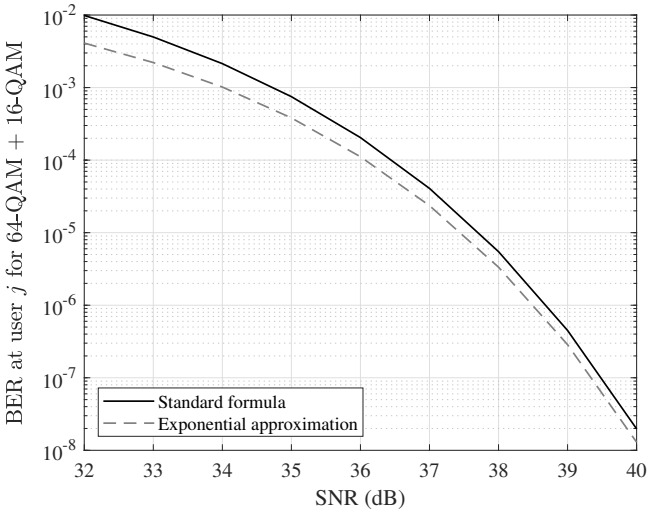
Figure 3.19: BER exponential approximations for 16-QAM + 16-QAM.

Next, Figure 3.19 represents the accuracy of the exponential approximations for 16-QAM + 16-QAM. In this case, the accuracy of the BER approximations improves substantially for both user j and user k, although it is observed that the

curvature of (3.52) differs with respect to that of (3.30). However, the accuracy remains within 0.25dB for all cases of interest at both user j and user k.



(a) BER at user k



(b) BER at user j

Figure 3.20: BER exponential approximations for 64-QAM + 16-QAM.

Last, Figure 3.20 represents the accuracy of the exponential approximations for 64-QAM + 16-QAM. At user k, the accuracy of (3.53) differs with respect to that of (3.36) is within 0.25dB for all cases of interest. At user j, (3.52) becomes a lower bound to (3.30), and the worst accuracy is obtained for low transmit

power and high BER values. For a BER of 10^{-3}, the accuracy is within 0.67dB, and it improves for lower BER values.

In conclusion, at user j, the approximation error introduced by (3.52) decreases for increasing $M_{j,s}$ and decreasing $M_{j,s}M_{k,s}$. In the worst case, for 256-QAM + 4-QAM, the exponential approximation (3.52) is accurate to within 0.51 dB with respect to the standard formula (3.30).

At user k, the exponential approximation (3.53) is more accurate with respect to the standard formula (3.36) for increasing values of $M_{j,s}$ and $M_{k,s}$. In the worst case, for 4-QAM + 4-QAM and 16-QAM + 4QAM, at user k, (3.53) is accurate to within 0.67dB with respect to (3.36).

In most cases, for $\beta_0 = 10^{-3}$, the exponential approximations constitute an upper-bound to the standard BER expressions in terms of the Q-function. The exceptions are the approximations to the BER at user k for 256-QAM + 4-QAM and 64-QAM + 16-QAM, as given in Figures 3.18 and 3.20, respectively.

3.9 Conclusions

In this Chapter, theoretical BER expressions were presented for two-user NOMA, for BPSK + BPSK, multilevel M_k-QAM + BPSK and multilevel M_k-QAM + M_j-QAM. To the best of the author's knowledge, this work provides the first attempt to derive BER expressions for multilevel M_k-QAM + M_j-QAM in NOMA.

It was analytically proved that, for the user who comes second in the SIC decoding order, there exists a certain value of the power allocation factor for which the BER is minimized. From the higher-layer user's point of view, the value of the power allocation factor must be high enough as to increase the power level of its own received signal, but it must also be low enough as to ensure successful decoding and removal of the lower-layer data stream. Further, the optimal power

allocation factor value is also the value that minimizes the transmit power in a BER-constrained NOMA system. The optimal level of the power allocation factor is dependent on the modulation level assigned to both NOMA users and, as the modulation level assigned to the lower-layer user increases, the value of the optimal power allocation rises sharply.

The theoretical BER expressions were used for calculating the optimal channel gain ratio for a pair of NOMA users, i.e. the ratio of channel gains that maximize the achievable sum rate for a given BER constraint. Unlike previous research in NOMA [51], it was demonstrated that, in NOMA systems with practical QAM, albeit different, the channel gains of two NOMA users must be of approximately the same order of magnitude in order to guarantee that the effect of inter-user interference can be overcome at the receivers, especially when the modulation levels assigned to users are similar.

Further, accurate exponential approximations were proposed for the BER in NOMA under multilevel M_k-QAM + M_j-QAM. Based on the BER exponential approximations, numerical boundaries were found for the values of the channel gain ratios of NOMA users that fulfill individual BER constraints for different combinations of modulation levels. It was proved that a pair of NOMA users whose channel gain ratio falls outside the numerical boundaries for a given modulation level cannot meet the individual BER constraints. In addition, the exponential BER approximations are easily invertible and differentiable, unlike their Q-function-based equivalent expressions. These qualities make them suitable for their application to solving the problem of resource allocation in NOMA.

Chapter 4

Resource Allocation in Multicarrier NOMA Based on Channel Gain Ratios

Contents

4.1 Introduction

4.1.1 Motivation

Beyond-5G paradigms such as the IoE require of novel technologies that are capable of providing massive scale connectivity in an efficient manner. It is expected that, in the connected society of 2030, massive machine-type communication networks will be occupied by the traffic of billions of devices, reaching densities of the order of 10^7 devices/km^2, with data rates of up to 1 Tbps and vast increases in energy and spectrum efficiency [12].

It has been proved that NOMA with SC can approach the channel capacity boundary in both the uplink and the downlink [46, 56, 63, 94]. SC takes advantage of the near-far effect in wireless communications [62] to simultaneously convey data of multiple users over the same communication channel. In contrast, the near-far effect may significantly degrade the achievable performance of OMA-based systems. In addition, NOMA can be easily combined with other

technologies [52], further increasing the scalability, spectral efficiency and energy efficiency of future wireless networks.

For the reasons above, NOMA has received considerable attention as a potential multiple access technique over the last few years. Although it was decided not to continue with NOMA as a study item for 5G in 3GPP [95], the proliferation of new use cases with an ultra-dense number of devices in beyond-5G networks is a motivating paradigm for the use of NOMA.

Up to this date, the relative performance gain of NOMA over OMA is still smaller than its high implementation cost. In particular, optimal user pairing in NOMA is specially challenging for systems with a large number of users, since it requires a large channel state information acquisition and feedback overhead, as well as very computationally complex optimization algorithms [95]. Further, the spectral efficiency of NOMA needs to be improved to provide larger gains than its OMA counterpart.

By combining NOMA with subcarrier-based schemes, the capabilities of both can be enhanced. Multicarrier NOMA systems take advantage of the orthogonality of radio resources in subcarrier-based schemes, in terms of a reduced implementation complexity, whilst increasing the performance by multiplexing users in a non-orthogonal manner in the same resource. The key to achieve the full potential of NOMA is resource allocation, which carries out the assignment of radio resources with the objective of optimizing a certain performance metric. Appropriate resource allocation in NOMA comprises the joint optimization of subcarrier allocation, data rate allocation and power allocation [52–54], which quite often yields an intractable problem, for which finding the optimal solution requires of prohibitively complex exhaustive search algorithms [55]. Therefore, in order to reduce the implementation cost of NOMA, it is critical to investigate how to reduce the computational complexity of resource allocation in multicarrier

NOMA systems, while achieving a performance close to optimal.

4.1.2 Contributions

In this Chapter, several contributions are made in terms of user pairing and resource allocation in multicarrier NOMA systems. Due to the intractability of the formulated problem, this is divided into the sub-problems of subcarrier allocation, and power and data rate allocation. These two sub-problems are separately optimized.

First, for the problem of subcarrier allocation, knowledge about the optimal channel gain ratios and channel gain gaps in NOMA that maximize the achievable sum-rate in BER-constrained scenarios is used to propose a user pairing algorithm that achieves quasi-linear complexity with respect to the number of users. This is achieved by implementing a search procedure with prior knowledge about the optimal value of the users' channel gains ratio and gap. In contrast, the complexity of exhaustive search procedures is of the order of the squared number of users. Other sub-optimal user pairing schemes in the literature [55, 56] also achieve a much larger computational complexity than the algorithm presented in the thesis.

Second, BER exponential approximations in NOMA are used for solving the problem of power and data rate allocation by applying a Lagrangian optimization method [58], with the objective of maximizing the sum-rate. The optimization result is applied to proposing a novel IRA-DRS power and data rate allocation algorithm for BER-constrained multicarrier NOMA systems. Unlike existing works, continuous power levels and discrete modulation levels are considered. Through numerical simulations, it is demonstrated that the proposed resource allocation scheme for multicarrier NOMA systems yields a performance close to optimal, and it outperforms other suboptimal schemes such as FTPC [56]. The minimum

performance loss introduced by IRA-DRS with respect to the optimal solution is 1.5%. Therefore, IRA-DRS yields an excellent trade-off between sum-rate performance and computational complexity. Further, the proposed scheme greatly benefits from multi-user diversity in terms of achievable sum-rate, number of iterations required for convergence, and degrees of freedom in choosing different combinations of modulation levels at each subcarrier that yield the same sum-rate.

Moreover, unlike previous works, it is demonstrated that the benefit of pairing NOMA users with very distinct channel conditions is lost in BER-constrained multicarrier NOMA with QAM. In fact, the average channel gain ratio obtained from simulations has a value close to two.

4.1.3 Structure

The remainder of this Chapter is organized as follows. A thorough literature review on user pairing and resource allocation in multicarrier NOMA systems is presented in Section 4.2. The system model and the problem formulation are presented in Section 4.3 and Section 4.4, respectively. The subcarrier allocation problem is solved in Section 4.5, where the user pairing algorithm is presented. Derivations for the power allocation and modulation level selection are presented in Section 4.6. Then, the IRA-DRS algorithms are presented in Section 4.7 and in Section 4.8, respectively. Numerical results on the effect of multi-user diversity, sum-rate performance, system optimization parameters and computational complexity of IRA-DRS are provided in Section 4.9. Finally, Section 4.10 concludes the Chapter.

4.2 Literature Review

Multicarrier NOMA systems [47] combine the benefits of OMA and NOMA. One of the major advantages of NOMA is that it is compatible with multicarrier based schemes, and therefore it can be easily integrated into existing schemes such as OFDMA. In multicarrier NOMA systems, different resource blocks are orthogonal in the time and frequency domains, but several NOMA users can be multiplexed into the same resource block in a non-orthogonal manner.

Resource allocation, i.e. the joint optimization of subcarrier and power allocation, is required in order to obtain the full benefits of multicarrier NOMA. However, in NOMA systems, this joint optimization problem normally leads to a mixed-integer, NP-hard problem [55]. NP-hard problems are intractable, and therefore their optimal solution can only be found through exhaustive search, which is a combinatorial optimization problem. A more practical approach to solving the joint problem of subcarrier and power allocation in multicarrier NOMA is to separate the problems of subcarrier allocation and power allocation, fix one of them and optimize the other [55], [57]. This leads to suboptimal but practical and efficient solutions that provide a better trade-off between implementation complexity and system performance [79, 96–98].

The main focus of this literature review is the study of existing works related to resource allocation in multicarrier NOMA, in terms of subcarrier allocation and power allocation.

4.2.1 Subcarrier Allocation

The problem of subcarrier allocation in NOMA consists in carrying out user pairing or user grouping at each subcarrier in an optimal manner. One approach to select the best NOMA user set on each subcarrier is to search over all pos-

sible combinations of users, then select the pair that optimizes a given system optimization parameter, such as the maximum sum-rate [56]. However, exhaustive search procedures require a computational complexity that is too large for implementation in practical systems.

The landmark work in [51] provided the first thorough study on the impact of user pairing on the performance of NOMA systems. It was proven that, in NOMA with fixed power allocation, both the probability that NOMA achieves a higher sum-rate than OMA and the sum-rate gain gap of NOMA over OMA can be increased by pairing NOMA users whose channel gains are significantly distinctive. It was also shown that, in terms of maximizing individual user data rates in NOMA systems, it is also preferable to pair two users with significantly distinctive channel gains.

The findings from [51] have since been applied extensively to the design of user pairing and user grouping schemes, which offer a more efficient solution to the subcarrier allocation problem than exhaustive search-based schemes. A common approach in the existing literature was the development of suboptimal schemes based on user grouping, i.e. roughly classifying users as strong or weak, depending on their channel conditions, and multiplexing them according to this classification. The procedure in [99] was to randomly allocate users in groups, then pair the users with best and worst channel conditions within each group. A similar approach was followed in [100], where dynamic user clustering and power allocation was proposed. Users were clustered and then paired within each cluster depending on their channel conditions.

Other more practical approaches considered aspects such as fairness or data rate constraints. For example, [101] proposed a sort-based pairing algorithm for downlink NOMA, where the users were sorted according to their channel condition and then paired. The algorithm guaranteed the minimum achievable data rate

at both users to be at least equal to the achievable with OMA schemes. In [102], a low-complexity algorithm for user pairing was proposed, in which a priority coefficient was derived in terms of data rate constraints, and the user with the largest priority coefficient was paired with the user with the best channel condition. Furthermore, a *fair* NOMA power allocation scheme was proposed in [103], where users were opportunistically paired irrespective of their channel conditions. The authors proved again that a larger difference in channel conditions between paired users leads to a larger increase in sum-rate.

Other investigations considered user grouping-based strategies, but with an increased degree of complexity in order to improve the balance between system performance and computational complexity. For example, in [104], a generalized user grouping approach was proposed, in which NOMA users are allowed to participate in several groups simultaneously, according to their channel gains, and under individual power constraints. By introducing this increased flexibility with respect to conventional user grouping schemes where users are placed into disjoint groups, a significant performance gain in terms of the sum-rate can be achieved, striking a good balance between computational complexity and performance. In [105], user pairing was carried out while considering the optimal power allocation between user pairs. A classification of user pairs was established based on the relationship between the users' channel conditions and the signal to noise and interference constraints. Then, a Tabu search algorithm was used to find a suboptimal user pair in a computationally efficient manner.

However, many challenges still remain open in terms of the subcarrier allocation problem. For example, the impact of user pairing on the users' individual BER has not been studied so far. It is also needed to further study how a certain ratio of users' channel conditions affects system parameters such as the system sum-rate and outage probability in multicarrier NOMA systems. This is necessary

in order to gain more insight into how to develop user grouping and user pairing strategies that can provide a performance close to that of exhaustive search, but with a reduced computational complexity. This is of paramount importance in the machine-type communications paradigm, due to the low processing capabilities of IoT devices.

4.2.2 Power Allocation

It is well known that the joint optimization of subcarrier and power allocation in multicarrier NOMA may lead to an NP-hard problem. A thorough analysis of the tractability of resource allocation in multicarrier NOMA was carried out in [55]. It was proven that the joint power and channel allocation problems in multicarrier NOMA under the utilities of sum-rate and weighted sum-rate are both NP-hard.

In the literature, the problem of optimal power allocation in multicarrier NOMA is commonly studied in a joint manner with the problem of subcarrier allocation. For intractable joint subcarrier and power allocation problems in multicarrier NOMA, the optimal solution requires of exhaustive search [107], which is not applicable in practical systems due to its prohibitively high computational complexity. It is usually preferable to design suboptimal schemes, that can provide a good trade-off between system performance and implementation complexity [108]. A performance bound on the global optimum was developed in [55] for both sum-rate and weighted sum-rate formulations, which can be used as a framework for performance evaluation of suboptimal subcarrier and power allocation schemes.

Power allocation in multicarrier NOMA systems has been studied under different performance metrics. A comprehensive investigation of joint subcarrier and power allocation in downlink NOMA systems was provided in [108], for several

problem formulations based on different performance criteria. Closed-form and semi-closed-form solutions were derived for the optimal power allocation, and these were used to jointly optimize subcarrier and power allocation in a suboptimal but practical and efficient manner.

The problem of sum-rate maximization in multicarrier NOMA is one of the most commonly adopted in the literature. This is an NP-hard problem [55] and, due to its complexity, most researches divided it into two equivalent subproblems, namely subcarrier and power allocation, and only proposed heuristic methods for resource allocation based on user grouping for pairing users sharing the same subcarrier. In [109], the authors formulated a resource allocation problem for downlink OFDM-based NOMA systems, with the objective of heuristically maximizing the system sum-rate, subject to power constraints. For subcarrier allocation, users with similar channel gains were allocated into a group. NOMA was applied to users from different groups in a greedy manner, by imposing a constraint to ensure a large enough difference in mean channel gain ratio between any two groups, based on the findings on optimal user pairing in [51]. [107] investigated resource allocation for the maximization of the weighted sum throughput of full-duplex multicarrier NOMA systems. A low-complexity suboptimal algorithm based on successive convex approximation was proposed, and its performance was shown to be close to optimal. A near-optimal solution to the sum-rate maximization problem was proposed in [55], through the discretization of the user power budget. However, in practical systems with a large number of power levels, the achievable computational complexity remained too high. In [110], optimal and approximate algorithms were proposed for joint subcarrier and power allocation. However, these relied on using a precomputation procedure in order to reduce the complexity of the power control algorithm. Therefore, this solution is not viable for a practical system where the channel conditions of all users are time variant,

and adaptive resource allocation is needed.

In summary, these works assumed continuous data rate allocation, based on theoretical data rate expressions, and did not consider practical systems using discrete modulation levels. Therefore, more knowledge is needed on how to achieve a good trade-off between complexity and performance in such a system, for example when QAM is employed. Moreover, existing works mainly considered system constraints such as minimum individual data rates or maximum transmit power, but did not consider constraints such as BER, which plays a critical role in ensuring that users receive a certain quality of service.

4.3 System Model

Consider a single-cell, multicarrier downlink NOMA system as represented in Figure 4.1, where it is assumed that the downlink channel gains of all subcarriers are perfectly estimated by each user through pilot channels and channel state information. Studying the impact of channel estimation error is not within the scope of this thesis, but insight into its impact in NOMA systems is provided in [88–91].

The scenario under consideration is assumed to have a set of active users $\mathcal{K} = \{1, \cdots, K\}$. The entire bandwidth of W Hertz is partitioned into S orthogonal subcarriers contained within the set $\mathcal{S} = \{1, \cdots, S\}$. It is assumed that there is no interference among adjacent subcarriers due to the orthogonal frequency partitioning. For all users, all subcarriers are assumed to be Rayleigh fading channels with AWGN, with double-sided spectral density $N_0/2$. The channel response of user j on subcarrier s is given by $h_{j,s}$, with $\mathbb{E}\{|h_{j,s}|^2\} = 1$, where $\mathbb{E}\{\cdot\}$ denotes statistical expectation. It is assumed that $|h_{j,s}|$ is Rayleigh distributed, $\forall j, s$. Further, $h_{j,s}$ and $h_{k,s}$ are independent for $j \neq k$.

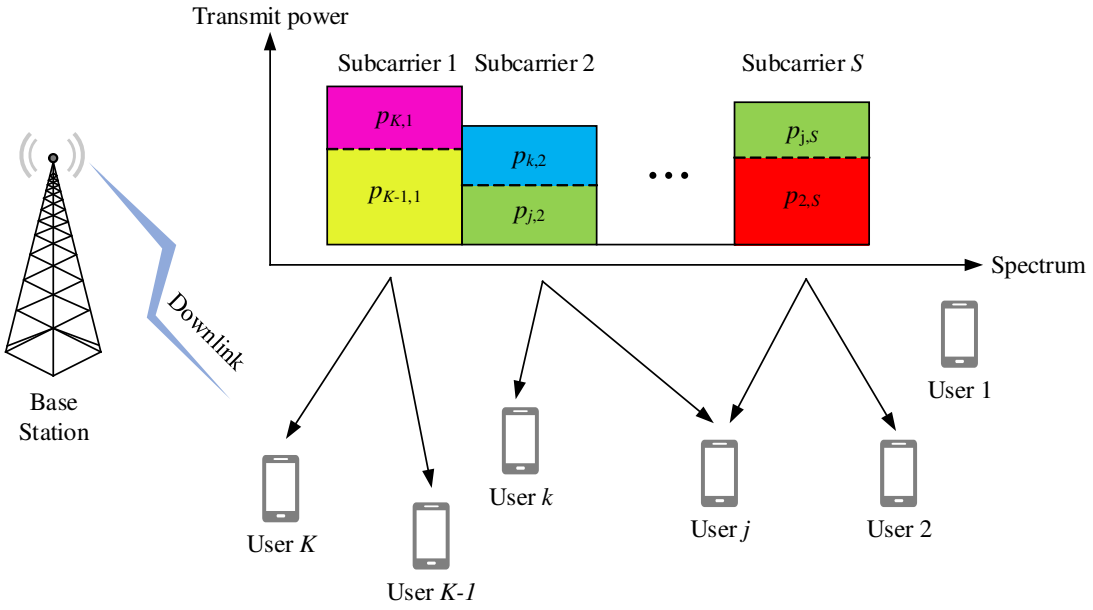


Figure 4.1: Multicarrier NOMA system model.

It is assumed that one user may be simultaneously assigned to several subcarriers. A particular subcarrier may be unused during a certain time resource if the channel conditions of users in that subcarrier are unfavorable. Without loss of generality, it is assumed that all users' channel gains on subcarrier s have been ordered from lowest to highest, i.e. $|h_{1,s}|^2 \leq \cdots \leq |h_{K,s}|^2$. Assume that users $\{j,k\} \in \mathcal{K}$ are jointly selected to perform NOMA on subcarrier $s \in \mathcal{S}$. The base station transmits a signal of the form

$$x_{\{j,k\},s} = \sqrt{p_{j,s}}x_{j,s} + \sqrt{p_{k,s}}x_{k,s}, \tag{4.1}$$

where $x_{j,s} \in \mathbb{C}$ denotes user j's transmit symbol, and $p_{j,s}$ is the transmit power to user j on subcarrier s. Further, $\mathbb{E}\{|x_{j,s}|^2\} = \mathbb{E}\{|x_{k,s}|^2\} = 1$ and $p_{j,s} = \alpha_{j,s}p_s$, where p_s is the total transmit power on subcarrier s and $\alpha_{j,s}$ is the power allocation factor of user j on that subcarrier. For any two multiplexed users j and k, the conditions $p_{j,s} + p_{k,s} = p_s$ and $\alpha_{j,s} + \alpha_{k,s} = 1$ hold for subcarrier s. Hence, the

received signals at user j and user k on subcarrier s can be expressed as

$$y_{j,s} = \sqrt{p_{j,s}} h_{j,s} x_{j,s} + \sqrt{p_{k,s}} h_{j,s} x_{k,s} + z_{j,s}, \tag{4.2}$$

$$y_{k,s} = \sqrt{p_{k,s}} h_{k,s} x_{k,s} + \sqrt{p_{j,s}} h_{k,s} x_{j,s} + z_{k,s}, \tag{4.3}$$

respectively, where $z_{j,s} \sim \mathcal{CN}(0, \sigma^2_{z_{j,s}})$, $z_{k,s} \sim \mathcal{CN}(0, \sigma^2_{z_{k,s}})$ denote the AWGN. Note that $\mathcal{CN}(a, \sigma^2)$ represents a circularly symmetric complex Gaussian distribution with mean a and variance σ^2.

It is assumed that the lower-layer user on each subcarrier, i.e. user j, uses maximum-likelihood detection with hard decision. At user k, the maximum-likelihood method is applied to decode user j's data stream. Then, user k uses SIC to remove interference from user j, where imperfect SIC is assumed, i.e. errors in detecting user j's signals are carried over to the second stage of detection, thus affecting the probability of correct detection of user k's own symbols. After removing user j's symbols, user k uses maximum-likelihood detection with hard decision to decode its own data stream [92].

The high implementation complexity of SIC remains an open research challenge [50], [111]. Therefore, in order to achieve a reasonable SIC implementation complexity, it is assumed that the base station multiplexes at most two NOMA users during each resource allocation instant. Resource allocation is performed at the base station under the constraints of total downlink transmit power, p_{total}, and identical maximum instantaneous BER per user, β_0. The base station multiplexes users into each subcarrier according to their channel gain ratios, such that the NOMA principle can be successfully applied. In the event that the application of NOMA is not feasible, a subcarrier may be unused.

The base station feeds forward information about subcarrier and modulation level assignment to each active user through downlink control signaling. In

addition, where two users are multiplexed according to the NOMA principle, information about user j's assigned power and modulation level is also fed forward to user k, thus enabling SIC decoding at the receiver.

The instantaneous BER at user j on subcarrier s is represented as $\beta_{j,s}$, and that at user k is represented as $\beta_{k,s}$. β_0 denotes the individual BER constraint, which is assumed to be equal at both user j and user k, for all subcarriers.

4.4 Problem Formulation

The optimization objective under consideration in this thesis is the maximization of the system sum-rate under individual, instantaneous BER constraints at each user and total transmit power constraint at the base station. This is given by the following indication function:

$$\max_{p_{j,s},p_{k,s},\delta_{j,s},\delta_{k,s}} \quad \sum_{s=1}^{S}\sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{k=1}^{K} \delta_{j,s}R_{j,s} + \delta_{k,s}R_{k,s} \tag{4.4a}$$

$$\text{s.t.} \quad \text{C1: } 0 \leq p_{j,s}, \forall j, s, \tag{4.4b}$$

$$\text{C2: } \delta_{j,s} \in \{0, 1\}, \forall j, s, \tag{4.4c}$$

$$\text{C3: } \sum_{j=1}^{K} \delta_{j,s} \leq 2, \forall s, \tag{4.4d}$$

$$\text{C4: } \sum_{s=1}^{S} (p_{j,s} + p_{k,s}) \leq p_{total}, \tag{4.4e}$$

$$\text{C5: } \beta_{j,s} \leq \beta_0, \forall j, s, \tag{4.4f}$$

where $R_{j,s}$ and $R_{k,s}$ are given by (3.58) and (3.59) in terms of the channel gain gap. $\delta_{j,s}$ is a binary variable defined as

$$\delta_{j,s} = \begin{cases} 1, & \text{if user } j \text{ is assigned to subcarrier } s, \\ 0, & \text{otherwise.} \end{cases} \tag{4.5}$$

Constraint C1 ensures non-negative transmit power. Constraints C2 and C3 guarantee that a maximum of two users may be allocated to one subcarrier. Constraint C4 is the maximum transmit power allowance at the base station, p_{total}. Further, constraint C5 guarantees that the individual instantaneous BERs at users on subcarrier s, i.e. $\beta_{j,s}$, $\forall j$, are lower than the BER threshold, β_0.

The problem in (4.4) was found to be NP-hard in [55]. Therefore, its optimal solution can only be found through exhaustive search, which is not practical due to its large computational complexity [56]. In order to make the maximization problem (4.4) more tractable, the proposed solution is to divide it into the sub-problems of subcarrier and power allocation, and then optimize each of them separately. This yields a sub-optimal but computationally efficient solution.

First, in order to solve the problem of subcarrier allocation in a simplified manner, the channel gain gap expression derived in Section 3.7 is applied to the development of a user pairing algorithm in Section 4.5. Second, by using the channel gain gap expression, a procedure for power and rate allocation is proposed in Sections 4.6–4.8.

4.5 Subcarrier Allocation

The main idea of subcarrier allocation is to allocate the best pair of NOMA users to a subcarrier, with no prior information on the transmit power, such that the sum-rate is maximized. In this work, it is assumed that one user can be assigned

to more than one subcarrier at the same time.

A relevant conclusion that can be extracted from (3.60)–(3.63) is that, in a scenario with a finite number of users, there might not be a suitable NOMA partner for the user with the highest channel gain. Therefore, in order to optimize $H_{\{j,k\},s}$, it might be feasible to select a user with a lower channel gain, but for whom a suitable partner that yields an adequate value of $g_{\{j,k\},s}$ can be found. Based on the optimal channel gain gap values from Table 3.1 and on the channel gain ratio condition (3.63), an efficient user pairing algorithm is proposed to carry out optimal user pairing at each subcarrier.

The aim of the user pairing algorithm is to select the pair of users k and j that yield the largest channel gain gap for a given subcarrier, $H_{\{j,k\},s}$. This is achieved by pairing a user with a very strong channel condition, and a user with the lowest possible channel condition that fulfills the BER constraint. Therefore, the user pairing algorithm is an iterative search procedure where the user with the strongest channel condition is considered, and a suitable partner with the lowest possible channel condition is searched for; when there are no suitable partners for the user with the strongest channel condition, the procedure is repeated for the second user with the best channel condition, and so on, until a suitable pair of NOMA users is found.

In the user pairing algorithm, the channel gains of users are ordered first from smallest to largest on subcarrier s, i.e. $|h_{1,s}|^2 \leq \cdots \leq |h_{K-1,s}|^2 \leq |h_{K,s}|^2$. The user with the largest channel gain is selected first, since this is the user that can potentially yield the largest $H_{\{j,k\},s}$ value, and consequently a higher sum-rate.

Next, the set $\mathbf{J}$ of possible partners that fulfill condition (3.63) when paired with user K is found. The user j that provides the highest $H_{\{j,k\},s}$ value, $\forall j$, is selected as the partner to user K, i.e. $j = \min(\mathbf{J})$, in order to maximize the value of $H_{\{j,k\},s}$ while fulfilling the BER constraint.

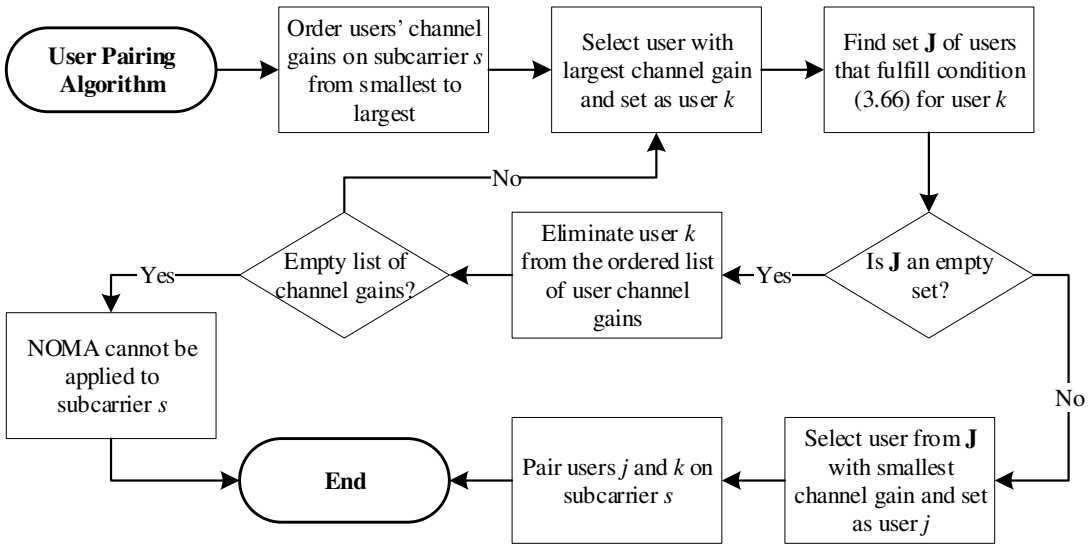


Figure 4.2: User pairing algorithm flowchart.

If no users fulfill (3.63) when paired with user K, the procedure is repeated for user $K-1$, and so on, until a suitable pair of users is found. Iterations are carried out for users in decreasing order of channel gains, in order to maximize the value of the channel gain gap. The user pairing algorithm is summarized in Algorithm 4.1, and its flowchart is given in Figure 4.2.

In scenarios with low multi-user diversity, it is more likely that users with a small channel gain ratio are paired on subcarrier s after applying user pairing. This yields a smaller $H_{\{j,k\},s}$ and therefore, according to (3.59) and Table 3.1, a smaller achievable sum-rate at subcarrier s. In the power allocation procedure in Section 4.6, this is penalized by allocating less power to subcarriers with a smaller $H_{\{j,k\},s}$. However, since both (3.37) and (3.63) are fulfilled, it is guaranteed that the effect of inter-user interference is manageable at user j, even for small channel gain ratios.

The presented user pairing procedure is applied in Section 4.7 to iteratively select the optimal pair of NOMA users allocated on every subcarrier for any given transmit power and BER constraints.

Numerical results in Section 4.9 prove that the complexity of the user pairing

Algorithm 4.1: User Pairing Algorithm

initialization;

order users from lowest to largest channel gain, as

$|h_{1,s}|^2 \leq \cdots \leq |h_{K-1,s}|^2 \leq |h_{K,s}|^2$;

set $i = K$ and $j = \emptyset$;

while $j = \emptyset$ *and* $i > 1$ **do**

  set $k = i$;

  find set $\mathbf{J}$ of possible partners for user k that fulfill (3.63);

  if $\mathbf{J} = \emptyset$ **then**

    i = i-1;

  end

  else

    set j as element from $\mathbf{J}$ that yields the maximum $H_{\{j,k\},s}$, i.e.

    $j = \min(\mathbf{J})$;

  end

end

Result: pair of users $\{j, k\}$.

algorithm is quasi-linear in practice, of the order of $\mathcal{O}(1.54K)$ for a system with a number of users between five and fifty-five. In comparison, the complexity of the user pairing prodecure in FTPC [56] and that of exhaustive search are of the order of $\mathcal{O}(K^2)$ per subcarrier, whereas the complexity of the sub-optimal SCUS scheme in [55] is of the order of $\mathcal{O}(2CK)$ after an initial stage of precomputation, where C is the number of discrete power allocation factor levels and it is of the order of 100-1000.

4.6 Power Allocation

After all subcarriers are allocated, the original objective in (4.4) leads to the development of power allocation in all subcarriers under the given subcarrier allocation result. Assume that the pairs of NOMA users allocated to each subcarrier are collected in the subcarrier allocation vector $\mathbf{U}$. Further, through the application of (3.58) and (3.59), it is imposed that the BER constraint be marginally met at all users. Therefore, the optimization problem given in (4.4) can be simplified as follows:

$$\underset{p_{j,s},p_{k,s}}{max} \quad \sum_{s=1}^{S} R_{j,s} + R_{k,s} \tag{4.6a}$$

$$\text{s.t.} \quad \text{C1: } 0 \leq p_s, \forall j, s, \tag{4.6b}$$

$$\text{C2: } \sum_{s=1}^{S} p_s \leq p_{total} \tag{4.6c}$$

In order to further simplify the analysis, the data rate variables $\{R_{j,s}, R_{k,s}\} \in \mathbb{I}^+$ in (4.6), which are positive integers, are respectively transformed into their equivalent real versions, $\{\hat{R}_{j,s}, \hat{R}_{k,s}\} \in \mathbb{R}$. According to the expression given in (3.59), the continuous sum-rate on subcarrier s can be written in terms of the

transmit power as

$$\hat{R}_{j,s} + \hat{R}_{k,s} = \log_2 \left(1 + \frac{c p_s H_{\{j,k\},s}}{\log 0.18} \right). \tag{4.7}$$

Since the second derivative of $\hat{R}_{j,s} + \hat{R}_{k,s}$ is always positive with respect to p_s for the interval $p_s > 0$, then the function (4.7) is concave, and the optimization problem given by (4.6) can be solved by using the Lagrangian method [58]. Consider the Lagrangian function

$$\mathcal{L} = \sum_{s=1}^{S} \left(\hat{R}_{j,s} + \hat{R}_{k,s} \right) - \lambda \left(\sum_{s=1}^{S} p_s - p_{total} \right), \tag{4.8}$$

where λ is the Lagrange multiplier for constraint (4.4e) and p_s is the total power allocated to subcarrier s. The solution to the optimal power allocation can be found by differentiating $\mathcal{L}$ in (4.8) with respect to p_s and equaling each derivation to zero, i.e.

$$\frac{\partial \mathcal{L}}{\partial p_s} = \frac{\mathrm{d}\hat{R}_{j,s}}{\mathrm{d}p_s} + \frac{\mathrm{d}\hat{R}_{k,s}}{\mathrm{d}p_s} - \lambda = 0. \tag{4.9}$$

In order to obtain $\mathrm{d}\hat{R}_{j,s}/\mathrm{d}p_s$ and $\mathrm{d}\hat{R}_{k,s}/\mathrm{d}p_s$, both sides of (3.58) are differentiated with respect to p_s, yielding

$$\begin{aligned} &\exp\left(\frac{c H_{\{j,k\},s} p_s}{2^{\hat{R}_{j,s}+\hat{R}_{k,s}} - 1} \right) \\ &\cdot \left(c H_{\{j,k\},s} \frac{\left(2^{\hat{R}_{j,s}+\hat{R}_{k,s}} - 1 \right) - \ln 2 \cdot 2^{\hat{R}_{j,s}+\hat{R}_{k,s}} p_s \left(\frac{\mathrm{d}\hat{R}_{j,s}}{\mathrm{d}p_s} + \frac{\mathrm{d}\hat{R}^{s}_{u_{k,s}}}{\mathrm{d}p_s} \right)}{2^{\hat{R}_{j,s}+\hat{R}_{k,s}} - 1} \right) = 0. \end{aligned} \tag{4.10}$$

In order for (4.10) to be zero, the following condition must be fulfilled:

$$\left(2^{\hat{R}_{j,s}+\hat{R}_{k,s}} - 1 \right) - \ln 2 \cdot 2^{\hat{R}_{j,s}+\hat{R}_{k,s}} p_s \left(\frac{\mathrm{d}\hat{R}_{j,s}}{\mathrm{d}p_s} + \frac{\mathrm{d}\hat{R}_{k,s}}{\mathrm{d}p_s} \right) = 0. \tag{4.11}$$

Hence, the derivative of the data rate at user k with respect to the transmit power on subcarrier s can be expressed as

$$\frac{\mathrm{d}\hat{R}_{j,s}}{\mathrm{d}p_s} + \frac{\mathrm{d}\hat{R}_{k,s}}{\mathrm{d}p_s} = \frac{2^{\hat{R}_{j,s}+\hat{R}_{k,s}} - 1}{2^{\hat{R}_{j,s}+\hat{R}_{k,s}} p_s \ln 2}. \tag{4.12}$$

Substituting (4.12) in (4.9) yields

$$\lambda = \frac{2^{\hat{R}_{j,s}+\hat{R}_{k,s}} - 1}{2^{\hat{R}_{j,s}+\hat{R}_{k,s}} p_s \ln 2}, \ s = 1, \cdots, S. \tag{4.13}$$

Thus, the transmit power of subcarrier s is derived as

$$p_s = \frac{2^{\hat{R}_{j,s}+\hat{R}_{k,s}} - 1}{2^{\hat{R}_{j,s}+\hat{R}_{k,s}}} \cdot \frac{1}{\lambda \cdot \ln 2} \tag{4.14}$$

Consider now that the constraint given in (4.4e) is marginally met, i.e.

$$\sum_{s=1}^{S} p_s = p_{total}. \tag{4.15}$$

After substituting (4.14) into (4.15), λ is derived as

$$\lambda = \frac{1}{\ln 2 \cdot p_{total}} \sum_{s=1}^{S} \frac{2^{\hat{R}_{j,s}+\hat{R}_{k,s}} - 1}{2^{\hat{R}_{j,s}+\hat{R}_{k,s}}}. \tag{4.16}$$

By equaling the value of λ in (4.14) and (4.16), the transmit power allocated to subcarrier s is derived as

$$p_s = \frac{p_{total}\left(2^{\hat{R}_{j,s}+\hat{R}_{k,s}} - 1\right)}{2^{\hat{R}_{j,s}+\hat{R}_{k,s}} \sum_{t=1}^{S}\left(1 - \frac{1}{2^{\hat{R}_{j,t}+\hat{R}_{k,t}}}\right)} = \frac{p_{total}\left(2^{\hat{R}_{j,s}+\hat{R}_{k,s}} - 1\right)}{2^{\hat{R}_{j,s}+\hat{R}_{k,s}} S - \sum_{t=1}^{S}\left(\frac{2^{\hat{R}_{j,s}+\hat{R}_{k,s}}}{2^{\hat{R}_{j,t}+\hat{R}_{k,t}}}\right)}. \tag{4.17}$$

From (3.59) and (4.14), the variable $H_{\{j,k\},s}$ can be written in terms of λ as

$$H_{\{j,k\},s} = \lambda \frac{\log 0.18 \log 2}{c} 2^{\hat{R}_{j,s}+\hat{R}_{k,s}}. \tag{4.18}$$

Then, by replacing $H_{\{j,k\},s}$ into (4.17),

$$p_s = \frac{p_{total}\left(2^{\hat{R}_{j,s}+\hat{R}_{k,s}} - 1\right)}{2^{\hat{R}_{j,s}+\hat{R}^s_{u_{k,s}}} S - \sum_{t=1}^{S} \frac{H_{\{j,k\},s}}{H_{\{j,k\},t}}}. \tag{4.19}$$

Further, by expressing $2^{\hat{R}_{j,s}+\hat{R}_{k,s}}$ in terms of $H_{\{j,k\},s}$ and p_s according to (3.59),

$$2^{\hat{R}_{j,s}+\hat{R}_{k,s}} = 1 + \frac{c p_s H_{\{j,k\},s}}{\log 0.18}, \tag{4.20}$$

and the transmit power allocation at subcarrier s is finally given by

$$p_s = \overline{p} + \frac{\ln 0.18}{cS} \sum_{t=1}^{S} \frac{1}{H_{\{j,k\},t}} - \frac{\ln 0.18}{c H_{\{j,k\},s}} \geq 0, \tag{4.21}$$

where $\overline{p} = p_{total}/S$ is the average transmit power per subcarrier, and the term $(1/S) \cdot \sum_{t=1}^{S} 1/H_{\{j,k\},t}$ is the average of $1/H_{\{j,k\},t}$ over all subcarriers $t = 1, \cdots S$.

According to (3.63) and (4.21), the power allocated to subcarrier s increases when the channel gain gap $H_{\{j,k\},s}$ between user j and user k on subcarrier s increases. This fact follows the water-filling principle in multiuser environments [62].

Note that, due to constraint (4.4b), if the result of (4.21) for subcarrier s is negative, then p_s is set to zero. In the event that one or more subcarriers are allocated with zero power, power allocation is carried out again among subcarriers allocated with positive power, in order to allocate all unused transmit power.

After power allocation, the theoretical sum-rate on each subcarrier with power

larger than zero can be derived from (3.59), by expressing

$$2^{\hat{R}_{j,s}+\hat{R}_{k,s}} = \frac{cp_s}{\ln 0.18} H_{\{j,k\},s} + 1. \tag{4.22}$$

From (4.21), it is found that (4.22) is equivalent to

$$2^{\hat{R}_{j,s}+\hat{R}_{k,s}} = \frac{cH_{\{j,k\},s}}{\ln 0.18} \cdot \overline{p} + \frac{1}{S} \sum_{t=1}^{S} \frac{H_{\{j,k\},s}}{H_{\{j,k\},t}}. \tag{4.23}$$

Therefore,

$$\hat{R}_{j,s} + \hat{R}_{k,s} = \log_2 \left(\frac{cH_{\{j,k\},s}}{\ln 0.18} \cdot \overline{p} + \frac{1}{S} \sum_{t=1}^{S} \frac{H_{\{j,k\},s}}{H_{\{j,k\},t}} \right). \tag{4.24}$$

From (4.24), it is clear that the highest sum data rate corresponds to the pair of NOMA users with the highest $H_{\{j,k\},s}$ value, which is consistent with the results obtained in (3.58) and Table 3.1.

Power Correction Factor

The approximation error introduced by (3.52) and (3.53) with respect to (3.30) and (3.36) results in excess power allocated to subcarriers. By introducing a power correction factor, the power allocated to each subcarrier can be reduced. Any unused power can then be applied to certain subcarriers in order to further increase the sum-rate.

Let $\varepsilon_{j,s}$ denote the excess power introduced on subcarrier s by (3.52) with respect to (3.30), and let $\varepsilon_{k,s}$ denote the excess power introduced on subcarrier s by (3.53) with respect to (3.36). By reducing the powers allocated to user j and user k by factors of $\varepsilon_{j,s}$ and $\varepsilon_{k,s}$ respectively, the same sum rate can be achieved while fulfilling the BER constraint β_0. The excess powers allocated to user j and user k on subcarrier s become equal by setting the power allocation factor at user j to $\breve{\alpha}_{j,s} = \alpha_{j,s}\sqrt{\varepsilon_{j,s}/\varepsilon_{k,s}}$. At user k, the power allocation factor is set to

Ratio $\varepsilon_{k,s}/\varepsilon_{j,s}s$				
	$\beta_0 = 10^{-3}$	$\beta_0 = 10^{-4}$	$\beta_0 = 10^{-5}$	$\beta_0 = 10^{-6}$
4-QAM + 4-QAM	1.101	1.088	1.074	1.066
16-QAM + 4-QAM	1.124	1.226	1.084	1.073
64-QAM + 4-QAM	1.110	1.096	1.083	1.068
256-QAM + 4-QAM	0.926	0.928	0.936	0.943
16-QAM + 16-QAM	1.023	1.040	1.038	1.036
64-QAM + 16-QAM	0.989	0.970	0.965	0.968

Table 4.1: Numerical evaluation of the ratio $\varepsilon_{k,s}/\varepsilon_{j,s}$ for M_k-QAM + M_j-QAM.

$\check{\alpha}_{k,s} = \alpha_{k,s}\sqrt{\varepsilon_{k,s}/\varepsilon_{j,s}}$. Therefore, a power correction factor $\mathcal{F}_s$ is defined as

$$\begin{aligned} \mathcal{F}_s &= \sqrt{\varepsilon_{k,s}/\varepsilon_{j,s}}, \text{ if } \varepsilon_{k,s}/\varepsilon_{j,s} \geq 1 \\ \mathcal{F}_s &= \sqrt{\varepsilon_{j,s}/\varepsilon_{k,s}}, \text{ otherwise.} \end{aligned} \tag{4.25}$$

$\mathcal{F}_s$ is calculated from (4.25) by applying the numerical values for $\varepsilon_{k,s}/\varepsilon_{j,s}$, as given on Table 4.1. The transmit power per subcarrier is readjusted by applying $\mathcal{F}_s$ to (4.21), i.e.

$$\check{p}_s = \left(\overline{p} + \frac{\ln 0.18}{cS}\sum_{t=1}^{S}\frac{1}{H_{\{j,k\},t}} - \frac{\ln 0.18}{cH_{\{j,k\},s}}\right)\cdot\frac{1}{\mathcal{F}_s} \geq 0. \tag{4.26}$$

$\mathcal{F}_s$ is applied after data rate allocation, given that prior knowledge about $M_{j,s}$ and $M_{k,s}$ is needed.

4.7 Iterative Resource Allocation Algorithm

Equation (4.24) is the theoretical sum-rate achieved for a subcarrier for a given power allocation. However, different individual user data rates may yield the same sum-rate for a given subcarrier. Consider, for example, a sum-rate of 8 bits/symbol for a given subcarrier; this sum-rate can be achieved for combinations of individual data rates such as 4+4, 5+3 or 6+2. In this Section, a heuristic

IRA algorithm is proposed to select suitable individual user data rates based on the optimal power allocation and theoretical sum-rate result, such that the BER constraint can be met by all users. In IRA, user pairing, power allocation and modulation selection are performed in an iterative manner, by applying (3.59), the numerical results from Table 3.1 and the optimal power allocation result given by (4.21).

In IRA, an initial pair of NOMA users is selected first for each subcarrier, based on the average transmit power per subcarrier $\overline{p}$. Expression (3.60) can be written in terms of $\overline{p}$ as follows,

$$\exp\left(\frac{c\overline{p}H_{\{j,k\},s}}{2^{\hat{R}^s_{u_{j,s}}+\hat{R}^s_{u_{k,s}}}-1}\right)=0.18. \tag{4.27}$$

According to (3.60), the expectation of the channel gain gap $H_{\{j,k\},s}$ across all subcarriers can be expressed as

$$\mathbb{E}\{H_{\{j,k\},s}\}=\left(1-\frac{1}{g_{\{j,k\},s}}\right)\mathbb{E}\{|h_{k,s}|^2\}, \tag{4.28}$$

where $\mathbb{E}\{|h_{k,s}|^2\}=1$. Therefore, from (4.27), the average data rate on subcarrier s is given by

$$\mathbb{E}\{\hat{R}_{j,s}+\hat{R}_{k,s}\}=\log_2\left(\frac{c\overline{p}\mathbb{E}\{H_{\{j,k\},s}\}}{\ln 0.18}+1\right)=\log_2\left(\frac{c\overline{p}\left(1-1/g_{\{j,k\},s}\right)}{\ln 0.18}+1\right). \tag{4.29}$$

For simplicity of notation, assume that the subset $\mathbf{V}\subset\mathbf{U}$ is the collection of users j allocated to each subcarrier, whereas $\mathbf{W}\subset\mathbf{U}$ is the subset of users k allocated to each subcarrier, and $\mathbf{V}\cup\mathbf{W}=\mathbf{U}$. The first user in $\mathbf{V}$ is paired with the first user in $\mathbf{W}$ on the first subcarrier, the second user in $\mathbf{V}$ is paired with the second user in $\mathbf{W}$ on the second subcarrier, and so on. The average data rates of the sets of users $\mathbf{V}$ and $\mathbf{W}$ across all subcarriers are given by

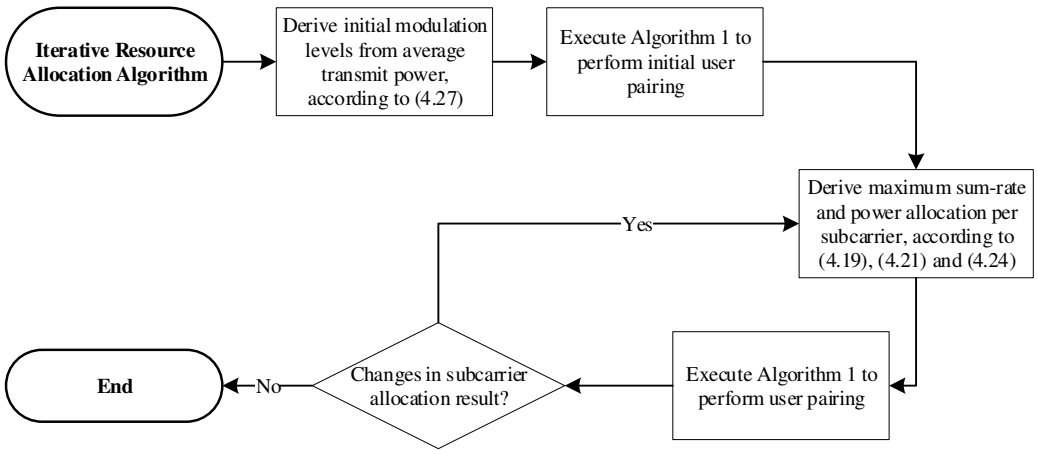


Figure 4.3: IRA flowchart.

Algorithm 4.2: Iterative Resource Allocation (IRA) Algorithm

initialization;

derive $\hat{\mathbf{R}}_{\mathbf{V}_{(0)}}$, $\hat{\mathbf{R}}_{\mathbf{W}_{(0)}}$, $\hat{\mathbf{M}}_{\mathbf{V}_{(0)}}$ and $\hat{\mathbf{M}}_{\mathbf{W}_{(0)}}$ from $\hat{\mathbf{P}}_{(0)}$ and β_0, according to (4.29);

execute Algorithm 4.1 to derive $\mathbf{V}_{(0)}$ and $\mathbf{W}_{(0)}$ from $\hat{\mathbf{M}}_{\mathbf{V}_{(0)}}$, $\hat{\mathbf{M}}_{\mathbf{W}_{(0)}}$ and β_0;

derive $\hat{\mathbf{P}}_{(1)}$ from $\mathbf{V}_{(0)}$ and $\mathbf{W}_{(0)}$, according to (4.21);

derive $\hat{\mathbf{R}}_{\mathbf{V}_{(1)}}$, $\hat{\mathbf{R}}_{\mathbf{W}_{(1)}}$, $\hat{\mathbf{M}}_{\mathbf{V}_{(1)}}$ and $\hat{\mathbf{M}}_{\mathbf{W}_{(1)}}$ from $\hat{\mathbf{P}}_{(1)}$, according to (4.24);

update $\hat{\mathbf{P}}_{(1)}$, according to (4.26);

execute Algorithm 4.1 to derive $\mathbf{V}_{(1)}$ and $\mathbf{W}_{(1)}$ from $\hat{\mathbf{M}}_{\mathbf{V}_{(1)}}$ and $\hat{\mathbf{M}}_{\mathbf{W}_{(1)}}$;

set $i = 1$;

while $\mathbf{V}_{(i)} \neq \mathbf{V}_{(i-1)}$ *or* $\mathbf{W}_{(i)} \neq \mathbf{W}_{(i-1)}$ **do**

  $i = i + 1$;

  derive $\hat{\mathbf{P}}_{(i)}$ from $\mathbf{V}_{(i-1)}$ and $\mathbf{W}_{(i-1)}$, according to (4.21);

  derive $\hat{\mathbf{R}}_{\mathbf{V}_{(i)}}$, $\hat{\mathbf{R}}_{\mathbf{W}_{(i)}}$, $\hat{\mathbf{M}}_{\mathbf{V}_{(i)}}$ and $\hat{\mathbf{M}}_{\mathbf{W}_{(i)}}$ from $\hat{\mathbf{P}}_{(i)}$, according to (4.24);

  update $\hat{\mathbf{P}}_{(i)}$ according to (4.26);

  execute Algorithm 4.1 to derive $\mathbf{V}_{(i)}$ and $\mathbf{W}_{(i)}$ from $\hat{\mathbf{M}}_{\mathbf{V}_{(i)}}$ and $\hat{\mathbf{M}}_{\mathbf{W}_{(i)}}$;

end

make $\hat{\mathbf{P}} = \hat{\mathbf{P}}_{(i)}$, $\hat{\mathbf{R}}_{\mathbf{V}} = \hat{\mathbf{R}}_{\mathbf{V}_{(i)}}$, $\hat{\mathbf{R}}_{\mathbf{W}} = \hat{\mathbf{R}}_{\mathbf{W}_{(i)}}$, $\hat{\mathbf{M}}_{\mathbf{V}} = \hat{\mathbf{M}}_{\mathbf{V}_{(i)}}$, $\hat{\mathbf{M}}_{\mathbf{W}} = \hat{\mathbf{M}}_{\mathbf{W}_{(i)}}$, $\mathbf{V} = \mathbf{V}_{(i)}$ and $\mathbf{W} = \mathbf{W}_{(i)}$;

Result: $\hat{\mathbf{P}}$, $\hat{\mathbf{R}}_{\mathbf{V}}$, $\hat{\mathbf{R}}_{\mathbf{W}}$, $\hat{\mathbf{M}}_{\mathbf{V}}$, $\hat{\mathbf{M}}_{\mathbf{W}}$, $\mathbf{V}$ and $\mathbf{W}$.

$\overline{R}_{\mathbf{V}} = \mathbb{E}\{\hat{R}_{k,s}\}$ and $\overline{R}_{\mathbf{W}} = \mathbb{E}\{\hat{R}_{j,s}\}$, respectively. The average modulation levels are given by $\overline{M}_{\mathbf{V}} = 2^{\overline{R}_{\mathbf{V}}}$ and $\overline{M}_{\mathbf{W}} = 2^{\overline{R}_{\mathbf{W}}}$, respectively. Note that there is an implicit dependency between $\overline{R}_{\mathbf{V}}$, $\overline{R}_{\mathbf{W}}$ and the BER constraint β_0, which is given by the values of the channel gain ratios on each subcarrier, $g_{\{j,k\},s}$.

In IRA, the initial modulation levels per user are derived first based on the average transmit power per subcarrier. Then, the user pairing algorithm is applied, based on the initial modulation levels. According to (4.21), the sum-rate per subcarrier increases with $H_{\{j,k\},s}$; therefore, the initial pair of NOMA users allocated to subcarrier s corresponds to the pair $\{j,k\}$ such that $H_{\{j,k\},s} = \max\{\mathbf{H}_s\}$, where $\mathbf{H}_s$ is the set of all possible channel gain gaps on subcarrier s that fulfill the channel gain ratio requirement for $\mathcal{G}(\lfloor\overline{M}_{\mathbf{W}}\rfloor, \lfloor\overline{M}_{\mathbf{V}}\rfloor, \beta_0)$. Note that the use of the function $\lfloor\cdot\rfloor$ is necessary given that $\{\overline{M}_{\mathbf{W}}, \overline{M}_{\mathbf{V}}\} \in \mathbb{R}$ and the fact that the channel gain ratio values on Table 3.1 are given for positive integer modulation values.

Let $\mathbf{V}_{(0)}$ and $\mathbf{W}_{(0)}$ denote the initial sets of users on each subcarrier. The initial power allocation set is denoted as $\hat{\mathbf{P}}_{(0)} = \{p_{total}/S, \cdots, p_{total}/S\}$. The set of achievable data rates for the set of users $\mathbf{V}_{(0)}$ is expressed as

$$\hat{\mathbf{R}}_{\mathbf{V}_{(0)}} = \{\overline{R}_{\mathbf{V}}, \cdots, \overline{R}_{\mathbf{V}}\}, \tag{4.30}$$

and that for the set of users $\mathbf{W}_{(0)}$ is given by

$$\hat{\mathbf{R}}_{\mathbf{W}_{(0)}} = \{\overline{R}_{\mathbf{W}}, \cdots, \overline{R}_{\mathbf{W}}\}. \tag{4.31}$$

The initial modulation levels of users $\mathbf{V}_{(0)}$ is given by the set $\{\overline{M}_{\mathbf{V}}, \cdots, \overline{M}_{\mathbf{V}}\}$, and that of the set of users $\mathbf{W}_{(0)}$ is given by $\{\overline{M}_{\mathbf{W}}, \cdots, \overline{M}_{\mathbf{W}}\}$. User pairing is applied initially to derive $\mathbf{V}_{(0)}$ and $\mathbf{W}_{(0)}$ in terms of $\overline{M}_{\mathbf{W}}$, $\overline{M}_{\mathbf{V}}$ and β_0.

After the initial sets of users are paired to each subcarrier, the achievable

modulation level on each subcarrier is calculated in an iterative manner, according to (4.24). A new pair of users $\{l, m\}$ is allocated on subcarrier s during each iteration, such that the BER constraints are met for any given power allocation. In this manner, more power is iteratively allocated to subcarriers where a higher sum-rate can be achieved, i.e. subcarriers where there exists a user with a large channel condition and for whom a suitable partner can be found.

The IRA algorithm finishes when an equilibrium state is found, i.e. when the sets of users $\mathbf{V}_{(i)}$ and $\mathbf{W}_{(i)}$ remain unchanged after the i-th iteration. IRA is summarized in Figure 4.3 and Algorithm 4.2.

4.8 Digitization of Data Rates

Recall that this work assumes practical, discrete-level modulation. However, after optimal power allocation, the theoretical sum-rate $\hat{R}_{j,s} + \hat{R}_{k,s}$ achieved per subcarrier, given by (4.24), is a real-valued number. Further, the resulting data rates from IRA are real, i.e. $\{\hat{\mathbf{R}}_{\mathbf{V}}, \hat{\mathbf{R}}_{\mathbf{W}}\} \subset \mathbb{R}$.

In order to obtain discrete modulation levels, it is necessary to carry out a digitization (or data rate conversion) process to transform the sets of data rates obtained from IRA into equivalent, practical discrete-valued data rates. When the digitization process is carried out, the discrete-valued dates are lower than their real-valued equivalents. Since lower power is required to achieve these discrete-valued data rates, the digitization process results in unused power at the transmitter.

In this Section, a DRS algorithm is proposed in order to carry out the digitization of data rates in all subcarriers after optimal, theoretical power allocation. In addition, any unused power after the conversion process is re-allocated among certain subcarriers to further increase the system sum-rate.

In order to obtain discrete modulation levels, it is first necessary to transform the real-valued sets of data rates resulting from IRA, i.e. $\hat{\mathbf{R}}_{\mathbf{V}}$ and $\hat{\mathbf{R}}_{\mathbf{W}}$, into equivalent sets of positive integers, i.e. $\tilde{\mathbf{R}}_{\mathbf{V}} = \{\tilde{R}_{k,1}, \cdots, \tilde{R}_{k,s}\} \subset \mathbb{I}^+$ and $\tilde{\mathbf{R}}_{\mathbf{W}} = \{\tilde{R}_{j,1}, \cdots, \tilde{R}_{j,s}\} \subset \mathbb{I}^+$. $\tilde{R}_{j,s}$ and $\tilde{R}_{k,s}$ are, initially, the largest integer values less than or equal to $\hat{R}_{j,s}$ and $\hat{R}_{k,s}$, respectively.

In order to carry out data rate conversion after IRA, the discrete data rates $\tilde{R}_{j,s}$ and $\tilde{R}_{k,s}$ are allocated to user j and user k, respectively, on subcarrier s. Let $\tilde{p}_s$ express the necessary power at subcarrier s, and denote $\tilde{\mathbf{P}} = \{\tilde{p}_1, \cdots, \tilde{p}_s, \cdots, \tilde{p}_S\}$. After initial modulation selection, given that $\tilde{R}_{j,s} \leq \hat{R}_{j,s}$ and $\tilde{R}_{k,s} \leq \hat{R}_{k,s}$, the required transmit power is given by $\sum_{s=1}^{S} \tilde{p}_s \leq \sum_{s=1}^{S} \hat{p}_s = p_{total}$, i.e. not all available power is used. The unused power after data rate conversion can be evaluated from (4.26) as

$$p_{unused} = \sum_{s=1}^{S} (\hat{p}_s - \tilde{p}_s) = \frac{\ln 0.18}{c} \sum_{s=1}^{S} \left(\frac{2^{\hat{R}_{j,s}+\hat{R}_{k,s}}}{\hat{\mathcal{F}}_s H_{\{j,k\},s}} - \frac{2^{\tilde{R}_{j,s}+\tilde{R}_{k,s}}}{\tilde{\mathcal{F}}_s H_{\{j,k\},s}} \right). \tag{4.32}$$

The unused power can be assigned to certain subcarriers in order to further increase the data rate at user k, without affecting user j's allocated data rate on that subcarrier. Let $R^*_{k,s} \in \mathbb{I}^+$ denote $\lceil \hat{R}_{k,s} \rceil$, i.e. $R^*_{k,s} = \tilde{R}_{k,s} + 1$. In order to achieve $M^*_{k,s} = 2^{R^*_{k,s}}$, if $g_{\{j,k\},s} \notin [1, \mathcal{G}(M^*_{k,s}, \beta_0)]$ it is necessary to apply user pairing in order to select a different pair of NOMA users $\{l, m\}$ on subcarrier s such that $g_{\{l,m\},s} \in [1, \mathcal{G}(M^*_{k,s}, \beta_0)]$. The pair $\{l, m\}$ is allocated to subcarrier s if and only if $H_{\{l,m\},s} \geq H_{\{j,k\},s}$. Otherwise, the pair $\{j, k\}$ remains allocated to subcarrier s and the data rate is set to $\tilde{R}_{k,s}$.

In the case where the pair $\{l, m\}$ is allocated to subcarrier s, the required power increase for a data rate of $R^*_{k,s}$ with respect to $\hat{R}_{k,s}$ is given by

$$\Delta p_s = \frac{\ln 0.18 \cdot 2^{\tilde{R}_{j,s}}}{c} \left(\frac{2^{R^*_{m,s}}}{\mathcal{F}^*_s H_{\{l,m\},s}} - \frac{2^{\hat{R}_{k,s}}}{\hat{\mathcal{F}}_s H_{\{j,k\},s}} \right). \tag{4.33}$$

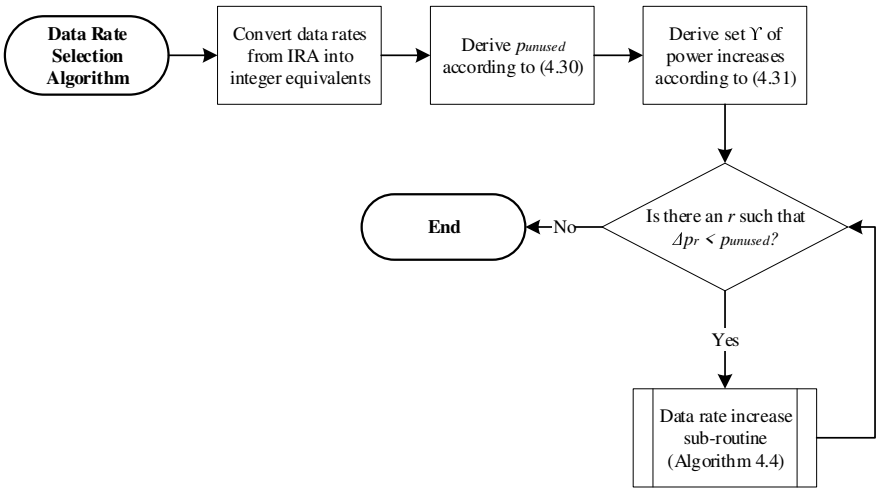


Figure 4.4: DRS flowchart.

Algorithm 4.3: Data Rate Selection (DRS) Algorithm

Data: $\tilde{\mathbf{P}}$, $\hat{\mathbf{R}}_{\mathbf{V}}$, $\hat{\mathbf{M}}_{\mathbf{V}}$, $\mathbf{V}$ and $\mathbf{W}$.

initialization;

convert $\hat{\mathbf{R}}_{\mathbf{V}} \subset \mathbb{R}$ to $\tilde{\mathbf{R}}_{\mathbf{V}} \subset \mathbb{I}^{+}$ and derive p_{unused} according to (4.32);

derive $\Upsilon = \{\Delta p_1, \cdots, \Delta p_s, \cdots, \Delta p_S\}$ according to (4.33);

while $\exists r : \Delta p^r < p_{unused}$ **do**

  sub-routine `Algorithm 4.4;`

end

Result: $\tilde{\mathbf{R}}_{\mathbf{V}}$, $\tilde{\mathbf{P}}$, $\mathbf{V}$, $\mathbf{W}$ and p_{unused}.

For lower values of Δp_s, lower additional extra power is required to increase the data rate at user k on subcarrier s. Thus, in order to maximize the system sum-rate, extra power is assigned to subcarriers in a strictly increasing order of Δp_s while $p_{unused} > 0$. A DRS algorithm is proposed in Algorithm 4.3 to assign the optimal integer data rate to each subcarrier, and to allocate p_{unused} to selected subcarriers to maximize the sum-rate. The overall computational complexity of IRA-DRS is presented in Section 4.9 for a practical system simulation.

Figure 4.4 represents the flowchart of the DRS algorithm. For simplicity, parts of the pseudo-code in Algorithm 4.3 are presented in Algorithm 4.4 and Algorithm 4.5, and flowcharts are given in Figure 4.5 and Figure 4.6.

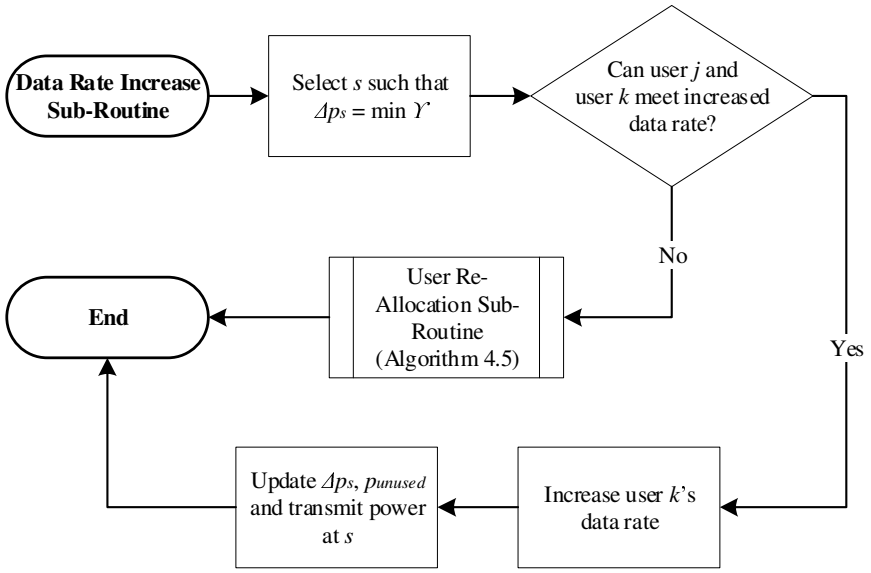


Figure 4.5: Data rate increase sub-routine flowchart.

Algorithm 4.4: Data Rate Increase Sub-Routine

Data: $\tilde{\mathbf{P}}$, $\hat{\mathbf{R}}_{\mathbf{V}}$, $\hat{\mathbf{M}}_{\mathbf{V}}$, $\mathbf{V}$, $\mathbf{W}$ and Υ.

initialization;

select s such that $\Delta p_s = \min \Upsilon$;

if $g_{\{j,k\},s} \notin \left[1, \mathcal{G}(\tilde{M}_{j,s}, M^*_{k,s}, \beta_0)\right]$ **then**

  sub-routine `Algorithm 4.5;`

end

else

  make $\tilde{R}_{k,s} \in \tilde{\mathbf{R}}_{\mathbf{V}}$ equal to $R^*_{m,s}$;

  update Δp_s, according to (4.33);

  make $p_{unused} = p_{unused} - \Delta p_s$, $\tilde{p}^s = \tilde{p}^s + \Delta p_s$;

end

Result: $\tilde{\mathbf{R}}_{\mathbf{V}}$, $\tilde{\mathbf{P}}$, $\mathbf{V}$, $\mathbf{W}$ and p_{unused}.

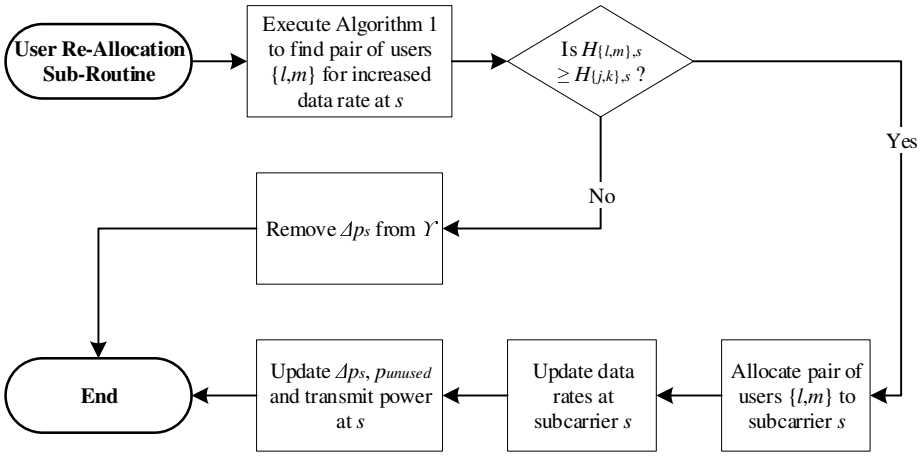


Figure 4.6: User re-allocation sub-routine flowchart.

Algorithm 4.5: User Re-Allocation Sub-Routine

Data: $\tilde{\mathbf{P}}$, $\hat{\mathbf{R}}_{\mathbf{V}}$, $\mathbf{V}$, $\mathbf{W}$ and $H_{\{j,k\},s}$.

initialization;

execute Algorithm 4.1 to find pair of users $\{l, m\}$ such that $g_{\{l,m\},s} \in [1, \mathcal{G}(\tilde{M}_{j,s}, M^*_{k,s}, \beta_0)]$;

if $H_{\{l,m\},s} \geq H_{\{j,k\},s}$ **then**

  allocate users $\{l, m\}$ to subcarrier s;

  make $\tilde{R}_{k,s} \in \tilde{\mathbf{R}}_{\mathbf{V}}$ equal to $R^*_{m,s}$;

  re-calculate Δp_s for $H_{\{l,m\},s}$, according to (4.33);

  make $p_{unused} = p_{unused} - \Delta p_s$ and $\tilde{p}^s = \tilde{p}^s + \Delta p_s$;

end

else

  remove Δp_s from Υ;

end

Result: $\tilde{\mathbf{R}}_{\mathbf{V}}$, $\tilde{\mathbf{P}}$, $\mathbf{V}$, $\mathbf{W}$, Υ and p_{unused}.

4.9 Numerical Results

The performance of the user pairing algorithm and that of IRA-DRS are evaluated in this Section. Results on the effect of multi-user diversity in the system are presented first, in order to provide the reader with some prior insight about the impact of multi-user diversity in the overall system performance. Then, sum-rate performance results are presented and compared to other schemes previously proposed in the literature. Further, some additional results are given on the impact of system optimization parameters, specifically the channel gain ratio and the power allocation factor. Last, the computational complexity and convergence of the user pairing and IRA-DRS algorithms are presented.

The simulated system model is formed by a single cell with one base station and a varying number of users. The number of subcarriers is set to $S = 64$, with a downlink system bandwidth of $W = 10$MHz, and a channel bandwidth of $W_s = W/S$. It is assumed that the broadband channel is frequency selective and independent for all users in all subcarriers. The normalized channel fading factor of all users on each subcarrier follows a Rayleigh distribution with mean square of one. Further, the system BER constraint is set to $\beta_0 = 10^{-3}$ for all users and the average transmit SNR per subcarrier is set to 23dB.

4.9.1 Effect of Multi-User Diversity

Before presenting sum-rate performance results, it is convenient to give some prior insight about the effect of multi-user diversity in the performance of IRA-DRS.

Recall that, in the user pairing procedure, two NOMA users can only be paired into subcarrier s if their individual channel condition gains fulfill (3.63), according to the values in Table 3.1.

Figure 4.7 represents the percentage of subcarriers where NOMA cannot be

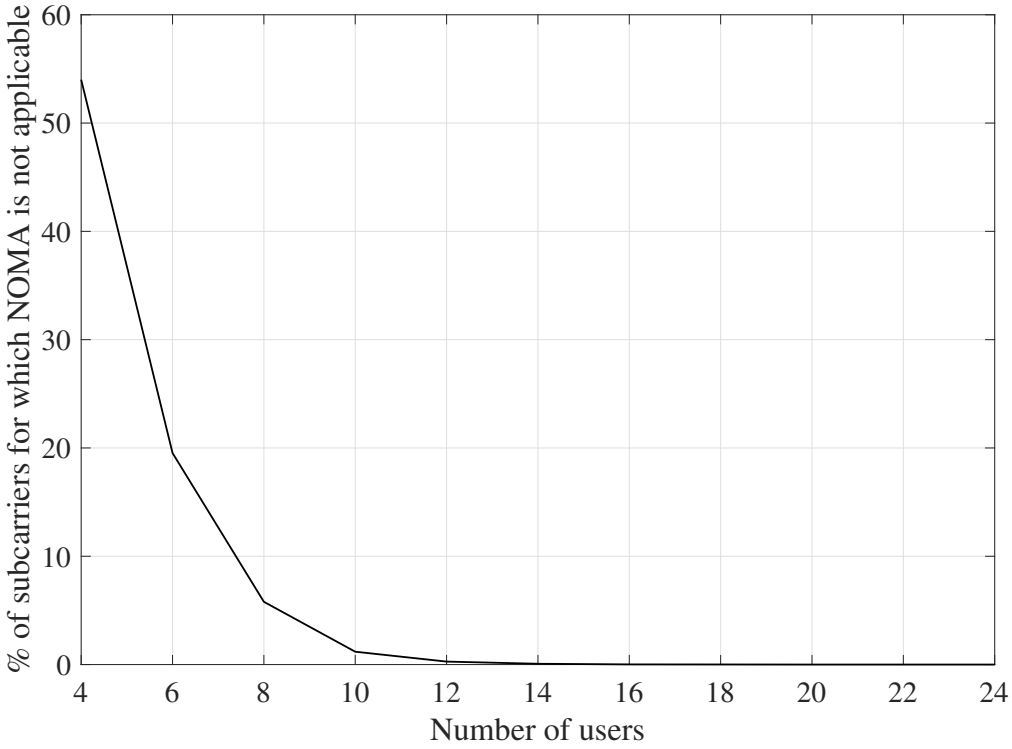


Figure 4.7: Percentage of subcarriers for which NOMA is not applicable.

applied, due to the impossibility of finding a pair of users with a channel condition gain ratio such that the BER constraint can be jointly meet at both users. Specifically, for a system with $K = 4$, NOMA can be applied in under half of the subcarriers. However, the probability of finding a suitable pair of NOMA users increases rapidly with K. For a system with $K = 8$, the percentage of subcarriers where NOMA cannot be applied decreases to under 6%, and for $K \geq 20$ the percentage becomes approximately zero. A conclusion that can be gathered from this result is that, for systems with a small number of users, the performance gain introduced by NOMA with respect to OMA is relatively small, whereas the implementation complexity is much larger. However, the performance of NOMA increases rapidly with the number of users. This result corroborates that NOMA yields a good performance-versus-complexity trade-off in scenarios with a high density of users, enabling massive connectivity.

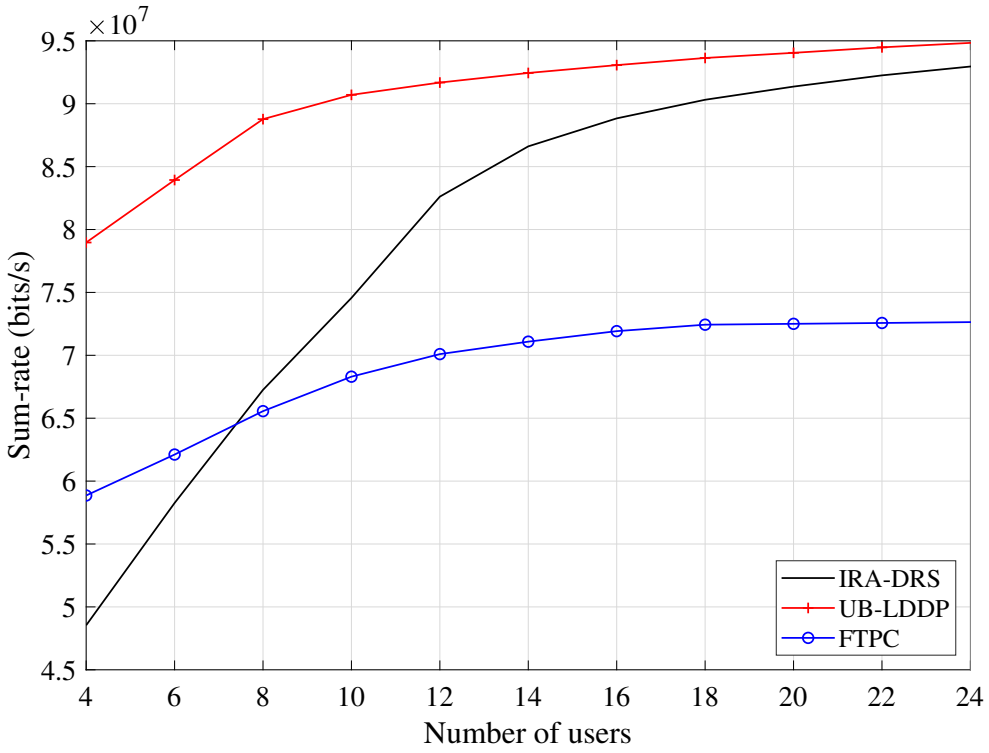


Figure 4.8: System sum-rate versus number of users.

4.9.2 Sum-Rate Performance

In this section, results on the system sum-rate performance of IRA-DRS are presented in terms of the number of users and the average transmit SNR per subcarrier. The performance of IRA-DRS is compared to those of upper-bound Lagrangian duality and dynamic programming (UB-LDDP) [55] and FTPC [56]. UB-LDDP is a Lagrangian duality and dynamic programming scheme which serves as a very tight upper-bound for the optimum achievable value of exhaustive search, and therefore it is used as a theoretical framework for performance evaluation. In FTPC, a greedy user pairing scheme is applied and power is dynamically allocated according to the channel gains of the multiplexed users. FTPC is a commonly used framework in the literature for sub-optimal resource allocation schemes [55].

Figure 4.8 represents the sum-rate achieved by IRA-DRS, for different num-

bers of users. For a system with twenty-four users, IRA-DRS has a performance loss within 1.5% of the maximum rate achieved with UB-LDDP, and a performance gain of over 28% with respect to FTPC. However, for a system with a number of users smaller than eight, IRA-DRS is outperformed by FTPC. The explanation for this effect is as follows. When the number of users in the system is small, there is a low probability of finding a pair of users with a suitable channel condition ratio in IRA-DRS, such that their channel gain ratio fulfills (3.63), according to the values in Table 3.1. Therefore, it might not be possible to apply the NOMA principle to some of the subcarriers. Further, in IRA-DRS, a minimum modulation level of 4-QAM + 4-QAM is assumed. In FTPC, a continuous modulation level is assumed, so it might be feasible to apply NOMA to a subcarrier yielding a total modulation level smaller than $4 \cdot 4$ in a situation where NOMA cannot be applied through IRA-DRS. However, it is observed that IRA-DRS greatly benefits from multi-user diversity. For a larger number of users, there is a higher probability that two users with a channel gain ratio close to the optimal value can be found. Therefore, the system sum-rate increases rapidly for an increasing number of users. For $K \geq 14$, the slope of the sum-rate curve decreases, as the system approaches the situation where multi-user gain saturation is reached. Overall, the IRA-DRS scheme provides an excellent tradeoff between achievable performance and system complexity for $K \geq 8$.

Figure 4.9 compares the performance of IRA-DRS with that of UB-LDDP and FTPC, for an increasing average transmit SNR per subcarrier, in a system with $K = 20$ users. For an average SNR per subcarrier greater than 23dB, the performance gap between IRA-DRS and UB-LDDP becomes approximately constant. The explanation for this effect is that, for higher SNR values, there are more combinations of modulation levels that fulfill the system BER constraint. Recall the BER results given in Figure 3.13a and Figure 3.13b, where it is clear

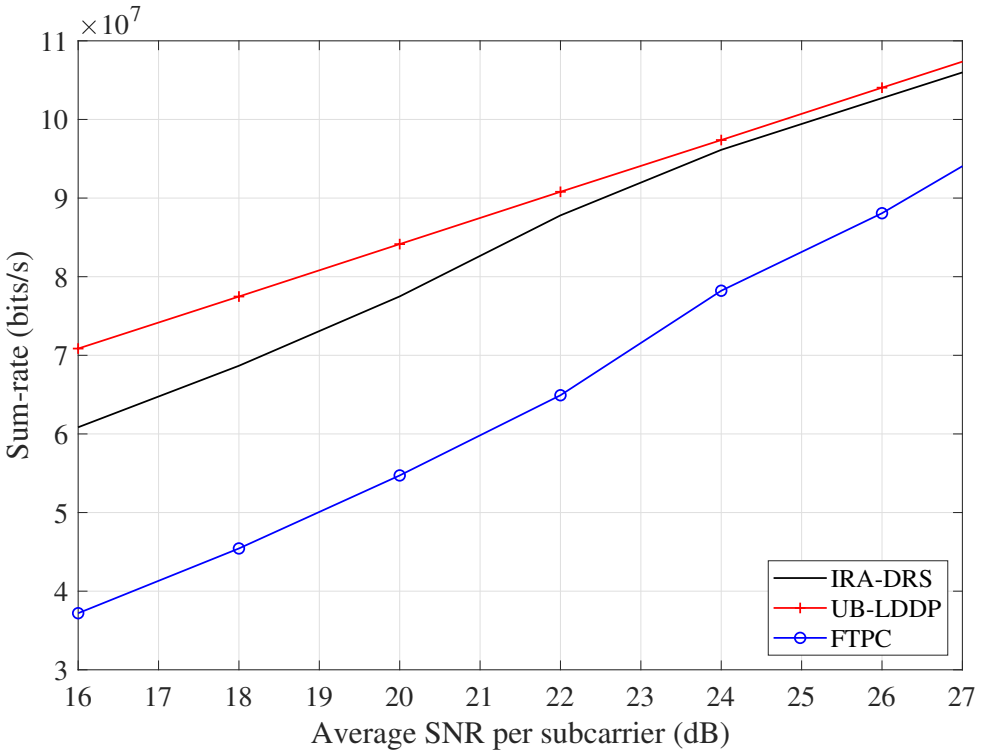


Figure 4.9: System sum-rate in terms of average SNR per subcarrier (dB).

that a much better receive SNR is required at user j as the level of $M_{j,s}$ increases. Therefore, according to Figure 4.9, there is a greater probability of finding a pair of NOMA users with a channel gain ratio closer to the optimal value for SNR$\geq$23dB. In contrast, for lower SNR values, there are less combinations of modulation levels for which a pair of users can simultaneously meet the BER constraints. Therefore, there is a lower probability of finding a pair of NOMA users with the channel gain ratio required for achieving higher modulation levels.

4.9.3 Results on System Optimization Parameters: Channel Gain Ratio and Power Allocation Factor

In this section, the channel gain ratio and power allocation factor are studied in terms of the number of users and the average transmit SNR. Further, their effect on the overall system performance is studied.

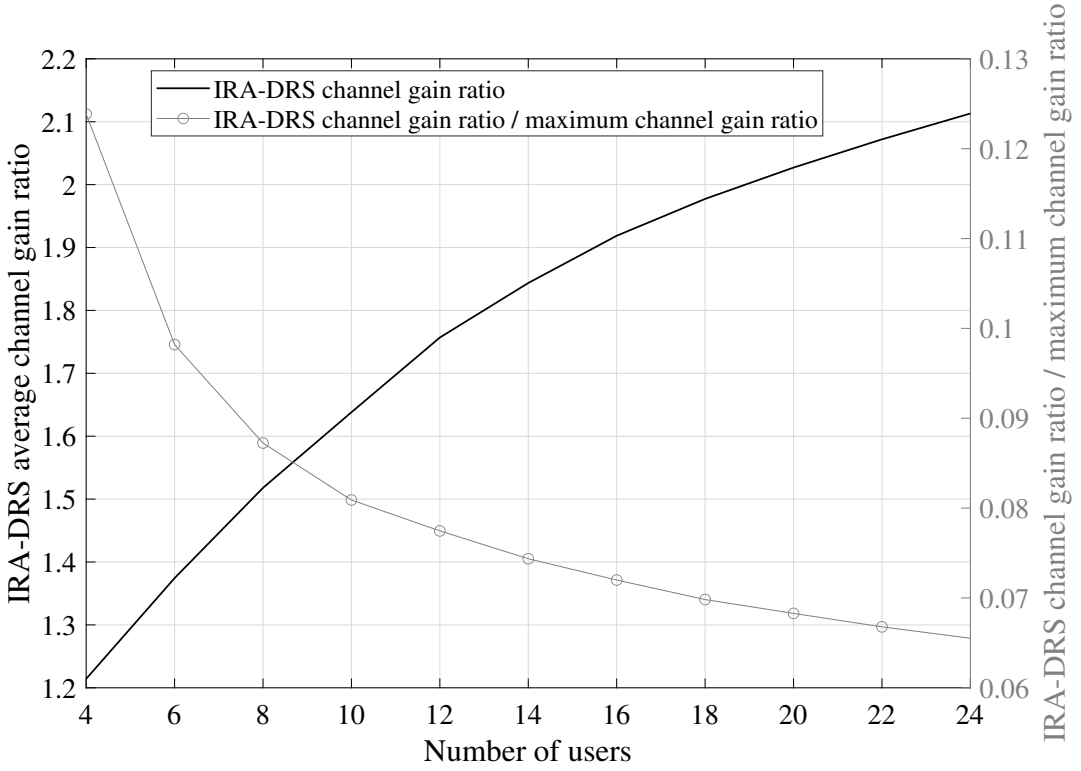


Figure 4.10: IRA-DRS average channel gain ratio.

Figure 4.10 shows the average channel gain ratio for IRA-DRS versus the number of users. The quotient of this curve over the maximum channel gain ratio is also presented, where the maximum channel gain ratio is calculated as follows. A set of randomized channel gains are obtained for each subcarrier over 10^7 realizations. Only the channel gain values between the 5th and 95th percentiles of the resulting distribution are considered, as a means of regularizing the data and avoiding distortions introduced by either very large or very small values of the channel gain. For each subcarrier, the ratio between the largest and smallest channel gains is calculated, and this value is averaged over all the system realizations. When the number of users increases, the IRA-DRS average channel gain ratio increases, but the coefficient between the IRA-DRS average channel gain ratio and the maximum channel gain ratio decreases. Therefore, the maximum channel gain ratio increases faster than the IRA-DRS channel gain ratio. This is due to the fact that, for a fixed transmit SNR, there exists a limit on

the sum-rate that IRA-DRS can achieve. Therefore, there also exists a limit to the IRA-DRS maximum average channel gain ratio. Moreover, Figure 4.10 illustrates the fact that, when practical QAM schemes are applied to NOMA, the benefit of pairing users with the most distinct channel condition [51] is lost, due to the inability of users with poor channel conditions to meet BER constraints. In the IRA-DRS setting simulated in Figure 4.10, the average maximum ratio between NOMA pairs is of the order of 2.1 as the optimal sum-rate is approached. This result is meaningful since it implies that NOMA user pairing schemes can be greatly simplified, by only searching among a handful of users that fulfill certain channel gain ratio conditions.

The average channel gain ratio for IRA-DRS versus increasing average transmit SNR per subcarrier is represented in Figure 4.11. This is compared to the average maximum channel gain ratio calculated over 10^7 realizations for a system with $K = 20$ users. In IRA-DRS, the average channel gain ratio is between 11 and 21 times smaller than the average maximum channel gain ratio per subcarrier, which equals 32.07. In a scenario where two users are paired with such a large channel gain ratio, a very large share of the power has to be allocated to the user with poorer channel condition in order to make the application of NOMA feasible. However, in systems with practical QAM schemes, this implies that the BER at the user with the stronger channel condition is far from the optimal value. The results in Figure 4.11 show again that, even though the average channel gain ratio increases with the transmit SNR, the optimal performance under discrete QAM levels is achieved for users with channel gains of the same order of magnitude. This condition facilitates the BER constraints being met at both NOMA users.

Figure 4.12 shows the sum-rate performance loss introduced by the IRA-DRS scheme with respect to UB-LDDP, versus the coefficient between the average channel gain ratio over the maximum channel gain ratio. The minimum per-

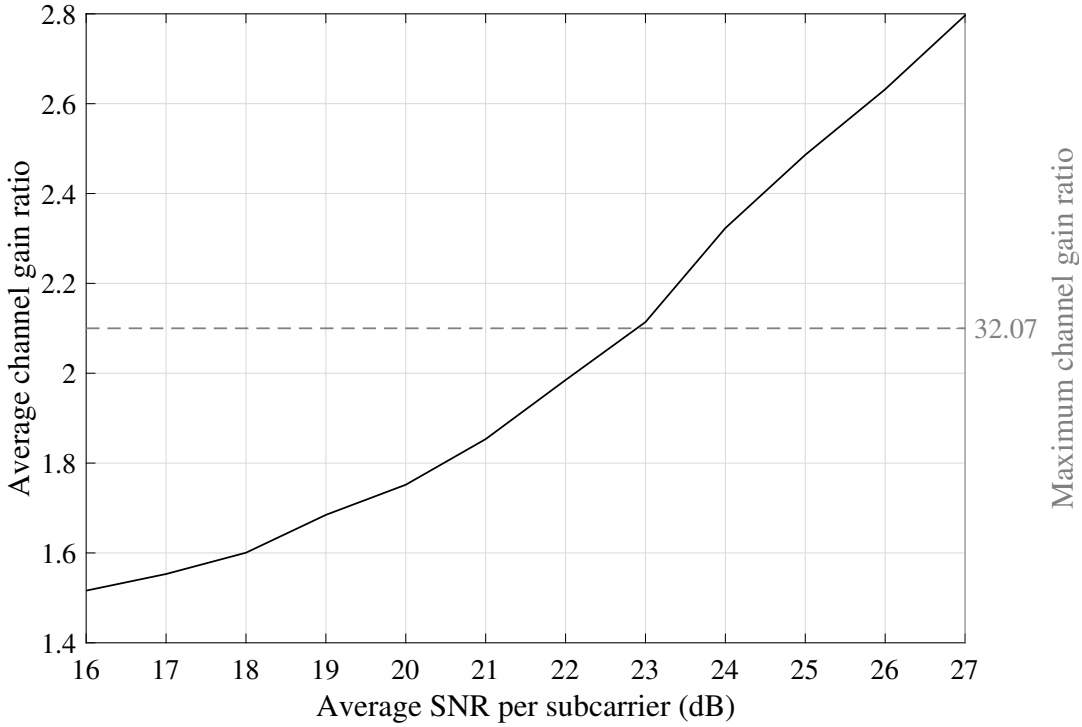


Figure 4.11: Average channel gain ratio in terms of average SNR per subcarrier (dB).

formance gap between IRA-DRS and UB-LDDP is just below 1.5%, and this is achieved when the average channel gain ratio is, approximately, 0.072 times the value of the maximum channel gain ratio. Increasing the value of the coefficient betweent the average channel gain ratio and the maximum channel gain ratio has a negligible effect on the sum-rate performance loss of IRA-DRS with respect to UB-LDDP, which becomes approximately constant for values over 0.072. However, when the coefficient is reduced below this value, the negative impact on the sum-rate performance loss is exponential.

Figure 4.13 presents the standard deviation of the IRA-DRS average channel gain ratio with respect to the optimal value, $\mathcal{G}(M_j, M_k, \beta_0)$. The curve decreases with the number of users due to multi-user diversity. This result is consistent with the obtained in Figure 4.7, since a larger standard deviation with respect to the optimal value of $\mathcal{G}(M_j, M_k, \beta_0)$ results in a larger probability of not finding a

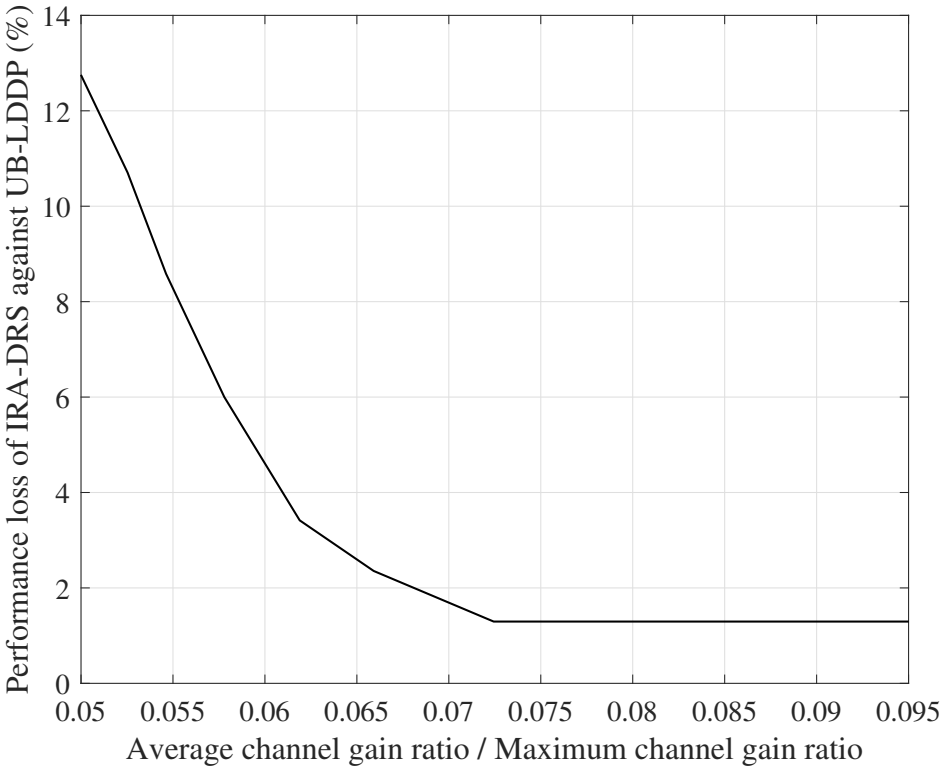


Figure 4.12: Performance loss of IRA-DRS against UB-LDDP (%) versus the coefficient of average channel gain ratio over maximum channel gain ratio.

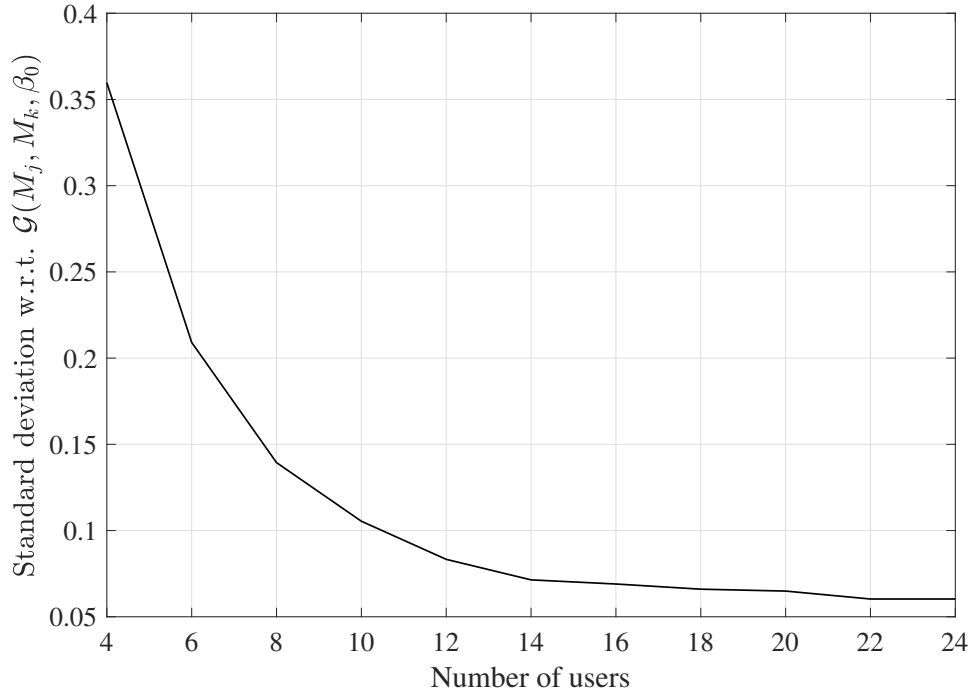


Figure 4.13: Standard deviation w.r.t. $\mathcal{G}(M_j, M_k, \beta_0)$.

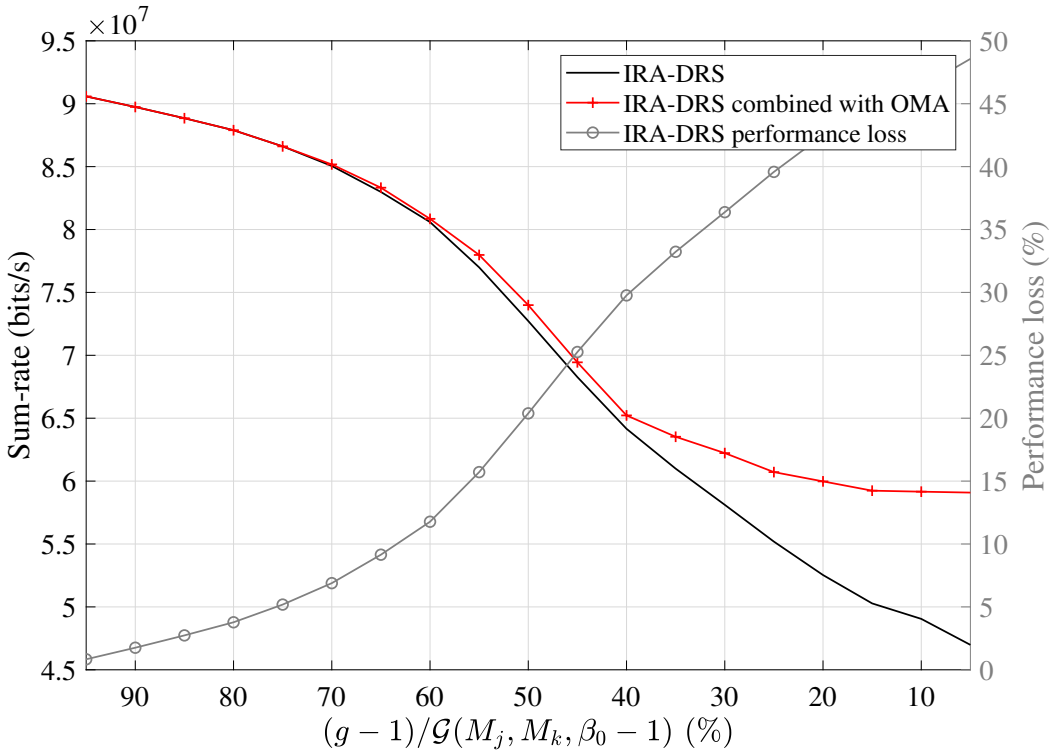


Figure 4.14: System sum-rate versus channel gain ratio.

suitable pair of NOMA users on each subcarrier. By comparing Figures 4.8 and 4.13, it is observed that IRA-DRS outperforms FTPC when the standard deviation with respect to $\mathcal{G}(M_j, M_k, \beta_0)$ is smaller than 0.15. At this point, according to Figure 4.7, the percentage of subcarriers where NOMA is not applicable drops to a 5%.

Figure 4.14 shows the system performance for $K = 20$ users, in terms of the sum-rate versus a ratio defined as

$$\frac{g-1}{\mathcal{G}(M_j, M_k, \beta_0) - 1}, \tag{4.34}$$

where

$$\frac{g-1}{\mathcal{G}(M_j, M_k, \beta_0) - 1} = 100\% \implies g = \mathcal{G}(M_j, M_k, \beta_0), \tag{4.35}$$

and

$$\frac{g-1}{\mathcal{G}(M_j, M_k, \beta_0) - 1} = 0\% \implies g = 1. \tag{4.36}$$

Recall that $g = \mathcal{G}(M_j, M_k, \beta_0)$, and $g = 1$ are the boundaries for the value of the channel gap ratio, according to (3.63). The performance loss of IRA-DRS becomes more critical for values of $(g-1)/(\mathcal{G}(M_j, M_k, \beta_0) - 1)$ under 65%, when the negative slope of the sum-rate curve becomes steeper, and the performance loss reaches 10%. For values of $(g-1)/(\mathcal{G}(M_j, M_k, \beta_0) - 1)$ under 40%, a large benefit can be obtained by applying OMA to unused subcarriers. Further, when the ratio $(g-1)/(\mathcal{G}(M_j, M_k, \beta_0) - 1)$ falls under 50%, the performance achieved by IRA-DRS and IRA-DRS combined with NOMA falls under the achieved by FTPC, as given in Figure 4.8. Therefore, by increasing the lower boundary of the channel gain ratio in (3.63), a better trade-off between performance and simplicity can be achieved, for example by applying OMA or FTPC to subcarriers where the value of the NOMA channel gain ratio falls under a given threshold.

Figure 4.15 shows the average power allocation factor per subcarrier at user k, for different number of users and an average transmit SNR of 23dB per subcarrier. The shape of the curve is similar to that of the sum-rate given in Figure 4.8, and that of the average channel gain ratio, as given in Figure 4.10. As the number of users increases and the sum-rate approaches its maximum achievable value, so does the power allocation factor, which converges to a maximum value of approximately 0.249 according to Figure 4.15. The reason is that, due to multi-user diversity, the average modulation level of the system converges to its maximum level, and therefore the power allocation factor converges to the optimal value for that modulation level. This limitation in the maximum power allocation factor ensures that the level of inter-user interference remains manageable and that the BER constraints can be simultaneously met at both NOMA users.

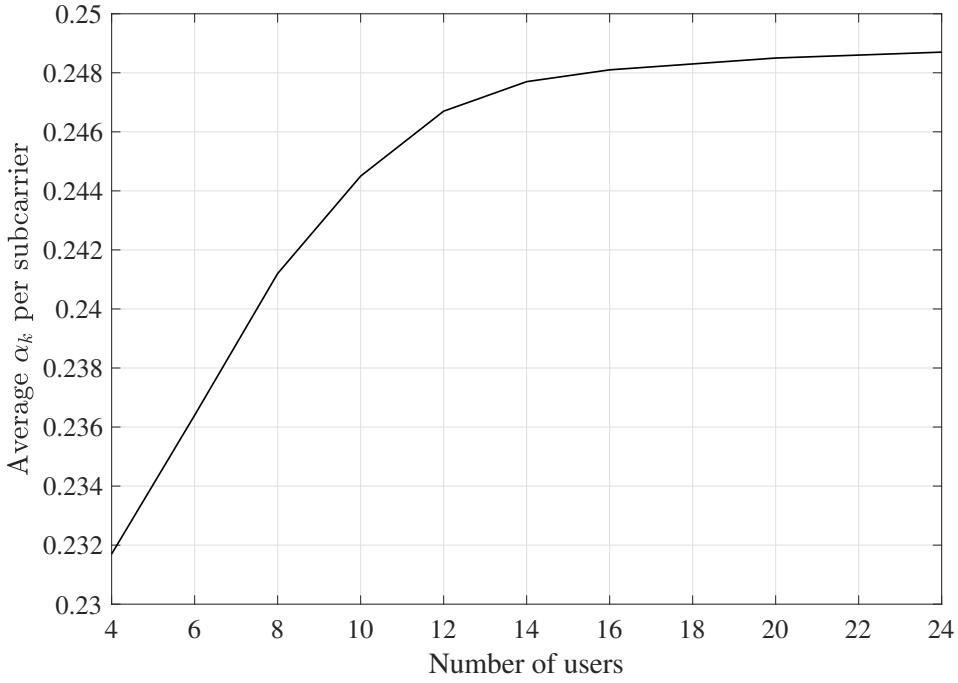


Figure 4.15: Average α_k per subcarrier versus number of users.

4.9.4 Computational Complexity and Convergence

The computational complexity of the user pairing algorithm and that of IRA-DRS are presented in this section. The computational complexity has been computed numerically rather than theoretically in all cases, for two reasons. First, it is overly complicated to obtain an analytical complexity expression. And second, the complexity obtained in a practical scenario may differ from the analytical value, due to the effect of multi-user diversity. While in a worst case scenario the user pairing algorithm would search over all combinations of users, in practice the search is reduced to only a few iterations, for a sufficiently large number of normally distributed user channel conditions. In all cases, the complexity of the algorithms has been numerically computed over $6.4 \cdot 10^4$ repetitions, for a varying number of users and subcarriers.

Figure 4.16 demonstrates that the complexity of the user pairing algorithm is quasi-linear in practice, of the order of $\mathcal{O}(1.54K)$ for $K = 55$ users. In com-

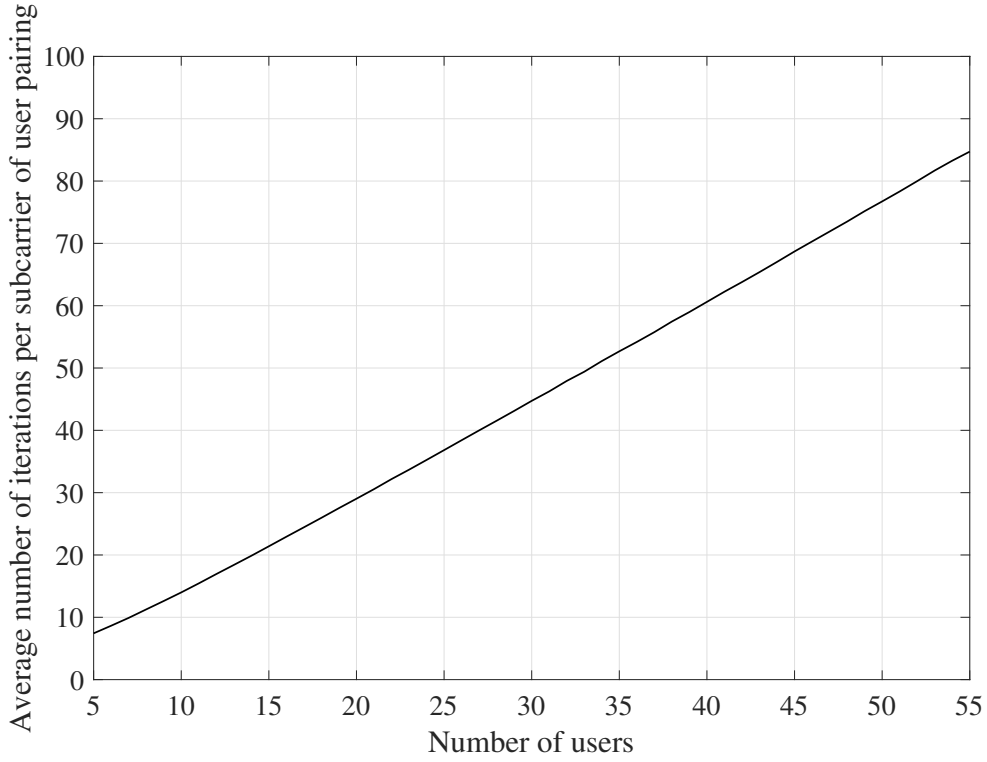


Figure 4.16: Complexity of the user pairing algorithm versus number of users.

parison, the complexity of the user pairing prodecure in FTPC [56] and that of exhaustive search are of the order of $\mathcal{O}(K^2)$ per subcarrier, whereas the complexity of the sub-optimal single-carrier user selection (SCUS) scheme in [110] is of the order of $\mathcal{O}(2CK)$ after an initial stage of precomputation, where C is the number of discrete power allocation factor levels, of the order of $10^2 - 10^3$. The user pairing procedure proposed in this work uses prior knowledge about the optimal channel gain ratio from Table 3.1 and expression (3.63). Therefore, it is possible to tune the user pairing algorithm to search only among a limited number of users, and the achieved complexity is greatly reduced compared to that of FTPC and SCUS.

Figure 4.17 shows the complexity of IRA-DRS versus the number of users, for a fixed average transmit SNR per subcarrier. When multi-user diversity gain saturation is achieved, the number of iterations of IRA-DRS is constant with the

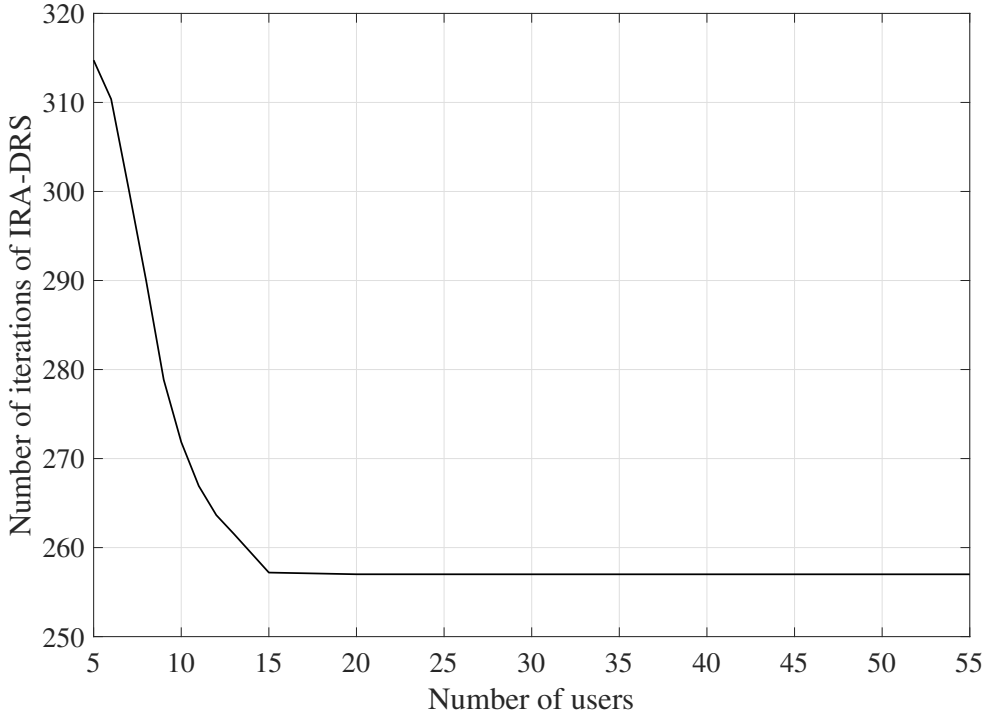


Figure 4.17: Complexity of IRA-DRS versus number of users.

number of users. However, for scenarios with a small number of users, additional iterations are required. The reason is that, when an initial modulation level is assigned to a subcarrier, it might not be possible to find a pair of NOMA users that fulfill the BER constraints. In this case a lower modulation level must be assigned to the subcarrier, unused power shared among the remaining subcarriers, and so on. In contrast, for systems with a large multi-user diversity gain, it is highly probable to find a pair of users that fulfill the modulation level initially assigned to a given subcarrier, and this reduces the necessary number of iterations for IRA-DRS to converge.

Last, Figure 4.18 shows the complexity of IRA-DRS versus the number of subcarriers in the system. It is demonstrated that the complexity of IRA-DRS is linear with the number of subcarriers, as it is also the case with the FTPC, UB-LDDP and SCUS schemes.

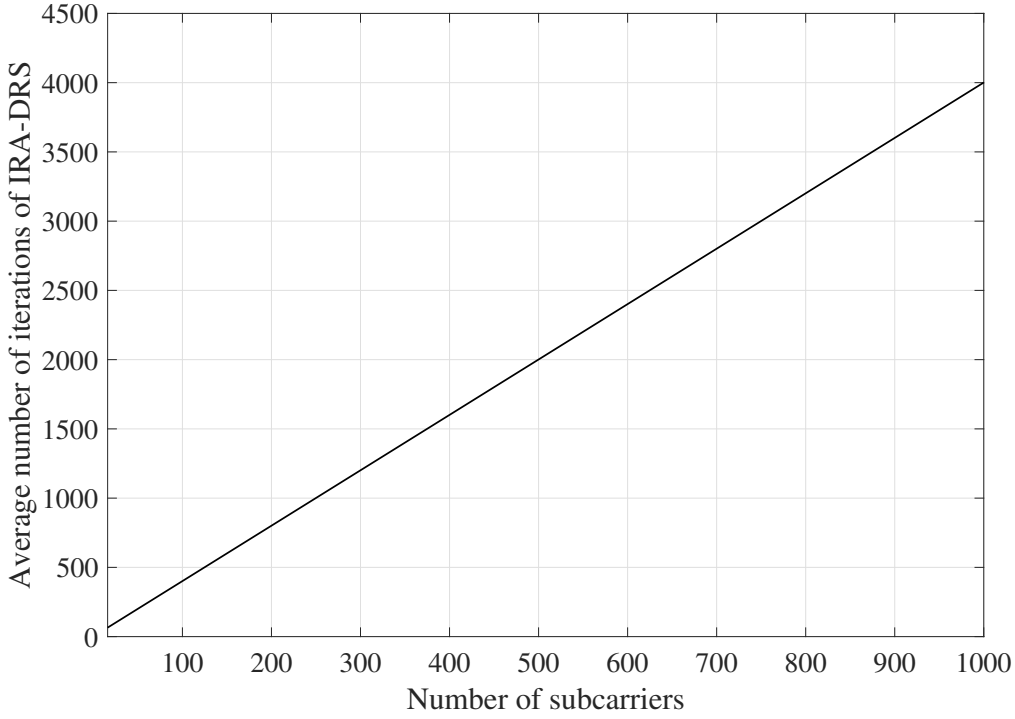


Figure 4.18: Complexity of IRA-DRS versus number of subcarriers.

4.10 Conclusions

In this Chapter, the problem of resource allocation in multicarrier NOMA systems was addressed. The optimization objective was the maximization of the system sum-rate under maximum transmit power and individual BER constraints. Due to the intractability of this problem, it was divided into the sub-problems of subcarrier allocation, and power and data rate allocation. These two sub-problems were separately optimized.

In order to solve the problem of subcarrier allocation, knowledge about optimal channel gain ratios and channel gain gaps in NOMA that maximize the achievable sum-rate in BER-constrained scenarios was applied. This knowledge was used to propose a search procedure, where the complexity is greatly reduced due to the prior knowledge about the optimal value of the users' channel gains ratio and gap. A user pairing algorithm was proposed that achieves quasi-

linear complexity with the number of users, in contrast with other sub-optimal schemes [55, 56] that yield a much larger computational complexity.

In terms of the power and data rate allocation problem, BER exponential approximations in NOMA were used for proposing a resource allocation scheme for the same BER-constrained multicarrier NOMA system, with the objective of maximizing the sum-rate. In order to solve this problem, a Lagrangian optimization method [58] was applied, yielding an analytical result that was applied to proposing an IRA-DRS scheme for resource allocation. Unlike existing works, continuous power levels and discrete modulation levels were considered. Through numerical simulations, it was demonstrated that the proposed IRA-DRS scheme yields a performance close to optimal, and it outperforms other suboptimal schemes such as FTPC [56]. The minimum performance loss introduced by IRA-DRS with respect to the optimal solution was measured as 1.5%. Therefore, it was demonstrated that IRA-DRS yields an excellent trade-off between sum-rate performance and computational complexity. Further, it was demonstrated that the proposed scheme greatly benefits from multi-user diversity in terms of achievable sum-rate, number of iterations required for convergence, and degrees of freedom in choosing different combinations of modulation levels at each subcarrier that yield the same sum-rate.

Unlike previous works, it was demonstrated that the benefit of pairing NOMA users with very distinct channel conditions is lost in BER-constrained multicarrier NOMA with QAM. In fact, the average channel gain ratio obtained from simulations has a value close to two.

Chapter 5

Conclusions and Future Research

Contents

5.1 Summary and Conclusions

The application of NOMA technologies to future wireless communications systems can improve the spectrum efficiency and enable massive connectivity in beyond-5G networks. NOMA can contribute to addressing the increasing need for massive connectivity by multiplexing several users into the same time and frequency resources. Further, the capabilities of NOMA can be extended through its combination with subcarrier-based schemes, such as OFDM, which is widely used in current wireless communications systems. However, some research challenges remain unaddressed. For example, many previous researches assumed perfect SIC and mainly considered the pairing of users with very distinct channel conditions [3, 51]. This can yield inefficient power allocations in terms of sum-rate. Further, the assumption of perfect SIC is not realistic in practical sys-

tems, where the SIC error propagation greatly impacts the achievable BER at the receivers. The implementation complexity of NOMA remains another major research challenge. Specifically, efficient resource allocation algorithms that provide a good trade-off between system performance and implementation practicality are needed.

The contributions in this thesis were two-fold. First, several contributions were made in terms of new performance bounds on the BER of NOMA systems, that prove the complex interplay between critical parameters in NOMA and the achievable BER. And second, several contributions were made in terms of user pairing and resource allocation in multicarrier NOMA systems.

First, theoretical BER expressions were presented for NOMA, assuming multi-layer, multi-level QAM. The optimal value of the power allocation factor in terms of BER transmit power was analytically derived. Further, the theoretical BER expressions were used for calculating the ratios of user channel gains that maximize the sum-rate for a given BER constraint. Unlike previous research in NOMA, it was demonstrated that, in NOMA systems with practical QAM, the channel gains of two NOMA users must be of approximately the same order of magnitude in order to guarantee that inter-user interference can be overcome at the receivers. By pairing two users such that their channel gain ratio equals the maximum allowable value (i.e. the optimal channel gain ratio), it is guaranteed that both users can marginally meet the BER constraint, while maximizing the sum-rate.

Additionally, accurate BER approximations were presented in the form of exponential functions, that were then applied for finding numerical boundaries for the values of the channel gain ratios of NOMA users that fulfill BER constraints for different combinations of modulation levels.

In terms of the resource allocation problem, the sub-problems of user pairing (i.e. subcarrier allocation), and power and data rate allocation were divided and

separately optimized. Knowledge about optimal channel gain ratios and channel gain gaps in NOMA was used to develop a user pairing algorithm that achieves quasi-linear complexity with respect to the number of users. In contrast, the complexity of exhaustive search procedures and FTPC is of the order of the squared number of users.

In terms of the power and data rate allocation problems, a Lagrangian optimization method based on BER exponential approximations was applied, with the objective of maximizing the system sum-rate. The optimization result was applied to proposing a novel IRA-DRS algorithm. Unlike existing works, continuous power levels and discrete modulation schemes were considered. Through numerical simulations, it was demonstrated that IRA-DRS yields a performance close to optimal, and it outperforms other suboptimal schemes such as FTPC. IRA-DRS yields an excellent trade-off between sum-rate performance and computational complexity.

The proposed scheme greatly benefits from multi-user diversity in terms of achievable sum-rate, number of iterations required for convergence, and degrees of freedom in choosing different combinations of modulation levels at each subcarrier that yield an equal sum-rate.

5.2 Future Research Directions

Despite the many potential benefits introduced by NOMA, its relative performance gain with respect to OMA is still relatively small compared to its large implementation cost in many use-case scenarios of interest. To make NOMA more suitable for implementation in future wireless networks, it is imperative to improve its spectral efficiency and practicality of implementation. Moreover, NOMA schemes can improve fairness among users, and this advantage should

be exploited in order to improve the performance gap with respect to OMA. In addition, it is important to better understand the performance limits of NOMA in more realistic scenarios that have been considered in the literature so far, such as under imperfect channel state information. Some interesting future research directions related to spectral efficiency, implementation complexity and studying the effect of imperfect channel state information are listed below.

User Fairness

NOMA schemes offer the potential to improve user fairness with respect to OMA systems.

1. An important research direction for this work is to extend the resource allocation problem formulation by considering the optimization of the system weighted sum-rate under BER constraints, as a means to improve user fairness.

2. In fairness-constrained systems, this work can be extended to consider the situation where the higher-level modulation is assigned to the NOMA user that enjoys a poor channel condition but has a high data rate requirement, whereas the lower-level modulation is assigned to the NOMA user with a better channel condition and lower data-rate requirements.

Spectral Efficiency

A way of improving the current performance gap between NOMA schemes and their OMA counterpart is to further increase the spectral efficiency of NOMA.

1. One method for improving the spectral efficiency in NOMA is the multiplexing of more than two users per subcarrier. Therefore, a logical extension to this work is to extend the derivations to an arbitrary number of users, and

study the performance boundaries of the system under BER constraints. This analysis can be performed for systems where SIC is applied, or for less complex systems where users decode their data directly.

2. Another possible research direction is to derive the optimal number of NOMA users per subcarrier, such that the spectral efficiency is maximized, but subject to minimum data rate requirements and maximum BER constraints. The number of users per subcarrier can then be adaptively selected according to the system performance metric, constraints and data rate requirements of users.

3. Another interesting research direction is to extend this work to consider scenarios where users can be assigned to any level of the transmit superconstellation, independently of their channel condition (eg. the case where the user with the highest channel gain is assigned to the lower layer of the transmit superconstellation), and study the performance bounds of such scenarios. Although such allocations may not be optimal in terms of system sum-rate, they can contribute to enhancing the spectral efficiency, for example by multiplexing users with a good channel condition and low data rate requirements together with users that enjoy a poorer channel condition but require a higher data rate.

Implementation Complexity

Receiver complexity remains one of the major drawbacks of NOMA systems, specially in use cases where end devices have low processing capabilities. Another source of complexity is the tremendous signaling overhead and computational resource consumption of multicarrier allocation schemes. Some future directions of work are proposed below to tackle these issues.

1. In machine-type communications applications where the capabilities of IoT devices are constrained in terms of power and processing capabilities, the power consumption and complexity of NOMA could be reduced by not applying SIC at the receivers, and instead directly decoding data by disregarding other users' information as noise. This would improve the number of users in the system at the expense of reduced sum-rates, which might be acceptable for machine-type communications devices given their normally low data rate requirements. The work in this thesis could be extended to derive the modulation levels and power allocation factor boundaries for which such a simplified system is feasible, such that decoding can be guaranteed at the receivers with a certain quality of service.

2. One of the major types of NOMA receiver is minimum mean squared error (MMSE)-hard interference cancellation (IC) [49]. A big source of complexity in MMSE-hard IC receivers is demodulation weight calculation, which involves matrix inversion operations. In scenarios with high correlation between channel coefficients over adjacent symbols, the weight vector of the MMSE-hard IC receiver might not need to be calculated and updated for every modulation symbol, thus reducing the computational complexity drastically.

   Therefore, a promising future direction of study is to apply the findings of this thesis to practical system models based on MMSE-hard IC receivers, and evaluating the performance loss when the updating of the system optimization parameters is spaced out in time, based on the channel correlation.

   The computational complexity of the MMSE-hard IC receiver could also be reduced through the application of chunk-based resource allocation [16, 112] in multicarrier NOMA. Effectively, chunk-based resource allocation is equi-

valent to spacing out the optimization of system parameters in frequency, based on the correlation of adjacent subcarriers.

3. Since wireless communication environments can change rapidly and optimal resource allocation requires real-time computation, the application of machine learning techniques can achieve a fast adaptation of the system parameters to the time-varying environment, and thus greatly reduce the computational complexity of resource allocation. Therefore, a possible future direction of study is the application of machine learning techniques to design low-complexity resource allocation schemes in multicarrier NOMA. Further, machine learning techniques can be applied to adaptively calculate how frequently the parameters of the the weight vector of the MMSE-hard IC receiver must be updated, according to the variability of the channel conditions in time, frequency and space.

Imperfect Channel State Information

In [113], it was proven that imperfect channel state information significantly degrades the performance of NOMA. Therefore, it is critical to understand the performance limits and benefits of NOMA in terms of channel state information in realistic scenarios.

1. In this sense, an interesting research direction is to study the effect of both statistical and imperfect channel state information on BER performance of NOMA. Specifically, the BER expressions can be extended to the case of statistical and imperfect channel state information, and use this knowledge to understand the performance boundaries of NOMA under BER constraints in such cases.

2. Another exciting application of machine learning in wireless communication

systems is the prediction of the fading channel [114]. This could be applied to NOMA systems under imperfect channel state information, in order to study its impact on the BER and develop techniques to improve the system performance.

3. Another important line of research is the study of NOMA under more challenging channel conditions, such as in millimeter-wave communication scenarios. By understanding the BER performance limits of NOMA in such scenarios, its performance can be enhanced.

References

[1] 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 Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28–41, 2019.

[2] L. Bariah, L. Mohjazi, S. Muhaidat, P. C. Sofotasios, G. K. Kurt, H. Yanikomeroglu, and O. A. Dobre, "A prospective look: Key enabling technologies, applications and open research topics in 6G networks," *IEEE Access*, pp. 1–1, 2020.

[3] S. M. R. Islam *et al.*, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commununications Surveys Tutorials*, vol. PP, no. 99, pp. 1–1, 2016.

[4] N. Alliance, "5G white paper," Tech. Rep., Feb. 2015.

[5] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, 2014.

[6] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, 2019.

[7] M. N. Patwary, S. Junaid Nawaz, M. A. Rahman, S. K. Sharma, M. M. Rashid, and S. J. Barnes, "The potential short- and long-term disruptions and transformative impacts of 5G and beyond wireless networks: Lessons learnt from the development of a 5G testbed environment," *IEEE Access*, vol. 8, pp. 11 352–11 379, 2020.

[8] K. David and H. Berndt, "6G vision and requirements: Is there any need for beyond 5G?" *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 72–80, 2018.

[9] Z. Nezami and K. Zamanifar, "Internet of things/internet of everything: Structure and ingredients," *IEEE Potentials*, vol. 38, no. 2, pp. 12–17, 2019.

[10] A. Shilpa, V. Muneeswaran, D. D. K. Rathinam, G. A. Santhiya, and J. Sherin, "Exploring the benefits of sensors in internet of everything (IoE)," in *2019 5th International Conference on Advanced Computing Communication Systems (ICACCS)*, 2019, pp. 510–514.

[11] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, pp. 1–9, 2019.

[12] M. Elbayoumi, M. Kamel, W. Hamouda, and A. Youssef, "NOMA-assisted machine-type communications in UDN: State-of-the-art and challenges," *IEEE Communications Surveys Tutorials*, vol. 22, no. 2, pp. 1276–1304, 2020.

[13] J. J. Sánchez-Sánchez, M. C. Aguayo-Torres, and U. Fernández-Plazaola, "Spectral efficiency of interleaved SC-FDMA with adaptive modulation and coding over Nakagami-m fading channels," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 8, pp. 3663–3670, 2013.

[14] C. Santivanez and I. Stavrakakis, "Study of various TDMA schemes for wireless networks in the presence of deadlines and overhead," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 7, pp. 1284–1304, 1999.

[15] L. L. Hanzo, L. L. Yang, E. L. Kuan, and K. Yen, "CDMA overview," in *Single and Multi-Carrier DS-CDMA: Multi-User Detection, Space-Time Spreading, Synchronisation, Networking and Standards*, 2004, pp. 35–80.

[16] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems - part I: chunk allocation," *IEEE Transactions on Communications*, vol. 57, no. 9, pp. 2734–2744, Sep. 2009.

[17] M. Vaezi, R. Schober, Z. Ding, and H. V. Poor, "Non-orthogonal multiple access: Common myths and critical questions," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 174–180, 2019.

[18] 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 Communications Surveys Tutorials*, pp. 1–1, 2020.

[19] Y. Lin, Z. Yang, and H. Guo, "Proportional fairness-based energy-efficient power allocation in downlink MIMO-NOMA systems with statistical CSI," *China Communications*, vol. 16, no. 12, pp. 47–55, 2019.

[20] K. Yang, N. Yang, N. Ye, M. Jia, Z. Gao, and R. Fan, "Non-orthogonal multiple access: Achieving sustainable future radio access," *IEEE Communications Magazine*, vol. 57, no. 2, pp. 116–121, 2019.

[21] Z. Song, Q. Ni, and X. Sun, "Spectrum and energy efficient resource allocation with QoS requirements for hybrid MC-NOMA 5G systems," *IEEE Access*, vol. 6, pp. 37 055–37 069, 2018.

[22] G. Liu, Z. Wang, J. Hu, Z. Ding, and P. Fan, "Cooperative NOMA broadcasting/multicasting for low-latency and high-reliability 5G cellular V2X communications," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7828–7838, 2019.

[23] Q. Guo, C. W. Sung, Y. Chen, and C. S. Chen, "Power control for coordinated NOMA downlink with cell-edge users," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, 2018, pp. 1–6.

[24] T. N. Do, D. B. da Costa, T. Q. Duong, and B. An, "Improving the performance of cell-edge users in NOMA systems using cooperative relaying," *IEEE Transactions on Communications*, vol. 66, no. 5, pp. 1883–1901, 2018.

[25] 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 Communications Surveys Tutorials*, pp. 1–1, 2020.

[26] 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 Transactions on Cognitive Communications and Networking*, vol. 5, no. 4, pp. 900–919, 2019.

[27] L. Zhu, Z. Xiao, X. Xia, and D. Oliver Wu, "Millimeter-wave communications with non-orthogonal multiple access for B5G/6G," *IEEE Access*, vol. 7, pp. 116 123–116 132, 2019.

[28] J. Zuo, Y. Liu, E. Basar, and O. A. Dobre, "Intelligent reflecting surface enhanced millimeter-wave NOMA systems," *IEEE Communications Letters*, pp. 1–1, 2020.

[29] Z. Xiao, L. Zhu, Z. Gao, D. O. Wu, and X. Xia, "User fairness non-orthogonal multiple access (NOMA) for millimeter-wave communications with analog beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3411–3423, 2019.

[30] Z. Xiao, L. Zhu, J. Choi, P. Xia, and X. Xia, "Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter wave communications," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 2961–2974, 2018.

[31] J. Li, X. Li, N. Ye, and A. Wang, "Performance evaluation of MIMO-NOMA in millimeter wave communication for broadcast services," in *2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2019, pp. 1–4.

[32] Y. Liu, "Exploiting NOMA for cooperative edge computing," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 99–103, 2019.

[33] L. Lv, J. Chen, Q. Ni, Z. Ding, and H. Jiang, "Cognitive non-orthogonal multiple access with cooperative relaying: A new wireless frontier for 5G spectrum sharing," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 188–195, 2018.

[34] C. Zhang, F. Jia, Z. Zhang, J. Ge, and F. Gong, "Physical layer security designs for 5G NOMA systems with a stronger near-end internal eavesdropper," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2020.

[35] D. Wang and S. Men, "Secure energy efficiency for NOMA based cognitive radio networks with nonlinear energy harvesting," *IEEE Access*, vol. 6, pp. 62 707–62 716, 2018.

[36] B. Lin, X. Tang, and Z. Ghassemlooy, "Optical power domain NOMA for visible light communications," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1260–1263, 2019.

[37] A. Kiani and N. Ansari, "Edge computing aware NOMA for 5G networks," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1299–1306, 2018.

[38] Y. Yuan, Z. Yuan, G. Yu, C. Hwang, P. Liao, A. Li, and K. Takeda, "Non-orthogonal transmission technology in LTE evolution," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 68–74, 2016.

[39] L. Yu, P. Fan, and Z. Han, "Maximizing spectral efficiency for SCMA systems with codebooks based on star-QAM signaling constellations," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1163–1166, 2019.

[40] Y. Han, W. Zhou, M. Zhao, and S. Zhou, "Enabling high order SCMA systems in downlink scenarios with a serial coding scheme," *IEEE Access*, vol. 6, pp. 33 796–33 809, 2018.

[41] X. Wu and Z. Wu, "Performance analysis of SCMA system based on polar codes," in *2018 IEEE Globecom Workshops*, 2018, pp. 1–5.

[42] Hujun Yin and Hui Liu, "Performance of space-division multiple-access (SDMA) with scheduling," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 611–618, 2002.

[43] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern division multiple access—a novel nonorthogonal multiple access for fifth-generation radio networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3185–3196, 2017.

[44] M. Mhedhbi and F. E. Boukour, "Analysis and evaluation of pattern division multiple access scheme jointed with 5G waveforms," *IEEE Access*, vol. 7, pp. 21 826–21 833, 2019.

[45] X. Bian, J. Tang, H. Wang, M. Li, and R. Song, "An uplink transmission scheme for pattern division multiple access based on DFT spread generalized multi-carrier modulation," *IEEE Access*, vol. 6, pp. 34 135–34 148, 2018.

[46] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Nakamura, "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access," in *2013 International Symposium on Intelligent Signal Processing and Communication Systems*, 2013, pp. 770–774.

[47] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, 2017.

[48] H. Haci, H. Zhu, and J. Wang, "Performance of non-orthogonal multiple access with a novel asynchronous interference cancellation technique," *IEEE Transactions on Communications*, vol. 65, no. 3, pp. 1319–1335, Mar. 2017.

[49] Y. Yuan, Z. Yuan, and L. Tian, "5G non-orthogonal multiple access study in 3GPP," *IEEE Communications Magazine*, vol. 58, no. 7, pp. 90–96, 2020.

[50] B. Makki, K. Chitti, A. Behravan, and M. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020.

[51] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6010–6023, 2016.

[52] J. Zeng, T. Lv, R. P. Liu, X. Su, M. Peng, C. Wang, and J. Mei, "Investigation on evolving single-carrier NOMA into multi-carrier NOMA in 5G," *IEEE Access*, vol. 6, pp. 48 268–48 288, 2018.

[53] F. Liu and M. Petrova, "Performance of dynamic power and channel allocation for downlink MC-NOMA systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1650–1662, 2020.

[54] Z. Wei, D. W. K. Ng, J. Yuan, and H. Wang, "Optimal resource allocation for power-efficient MC-NOMA with imperfect channel state information," *IEEE Transactions on Communications*, vol. 65, no. 9, pp. 3944–3961, 2017.

[55] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Power and channel allocation for non-orthogonal multiple access in 5G systems: Tractability and computation," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8580–8594, 2016.

[56] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2013, pp. 611–615.

[57] P. Parida and S. S. Das, "Power allocation in OFDM based NOMA systems: A DC programming approach," in *2014 IEEE Globecom Workshops*, Dec. 2014, pp. 1026–1031.

[58] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[59] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.

[60] R. M. Buehrer, *Code Division Multiple Access (CDMA)*, 2006.

[61] Cheong Yui Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1758, 1999.

[62] A. Goldsmith, *Wireless communications*. Cambridge University Press, 2005.

[63] R. Zhang and L. Hanzo, "A unified treatment of superposition coding aided communications: Theory and practice," *IEEE Communications Surveys Tutorials*, vol. 13, no. 3, pp. 503–520, Third Quarter 2011.

[64] T. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, 1972.

[65] S. Tomida and K. Higuchi, "Non-orthogonal access with SIC in cellular downlink for user fairness enhancement," in *2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*, 2011, pp. 1–6.

[66] S. Vanka, S. Srinivasa, Z. Gong, P. Vizi, K. Stamatiou, and M. Haenggi, "Superposition coding strategies: Design and experimental evaluation," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2628–2639, 2012.

[67] L. Ping, J. Tong, X. Yuan, and Q. Guo, "Superposition coded modulation and iterative linear mmse detection," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 6, pp. 995–1004, 2009.

[68] P. K. Vitthaladevuni and M. S. Alouini, "BER computation of 4/M-QAM hierarchical constellations," *IEEE Transactions on Broadcasting*, vol. 47, no. 3, pp. 228–239, Sep. 2001.

[69] ——, "A closed-form expression for the exact ber of generalized PAM and QAM constellations," *IEEE Transactions on Communications*, vol. 52, no. 5, pp. 698–700, May 2004.

[70] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of gaussian multiple-access and broadcast channels," *IEEE Transactions on Information Theory*, vol. 50, no. 5, pp. 768–783, 2004.

[71] S. Vanka, S. Srinivasa, Z. Gong, P. Vizi, K. Stamatiou, and M. Haenggi, "Superposition coding strategies: Design and experimental evaluation," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2628–2639, 2012.

[72] A. Samuel and C. Sipes, "Making internet of things real," *IEEE Internet of Things Magazine*, vol. 2, no. 1, pp. 10–12, 2019.

[73] "Study on downlink multiuser superposition transmission for LTE," in *document, 3rd Generation Partnership Project (3GPP)*, Mar. 2011.

[74] B. Makki *et al.*, "A survey of noma: Current status and open research challenges," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020.

[75] Seong Taek Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1561–1571, Sep. 2001.

[76] F. Wei, T. Zhou, T. Xu, and H. Hu, "BER analysis for uplink NOMA in asymmetric channels," *IEEE Communications Letters*, pp. 1–1, 2020.

[77] X. Wang, F. Labeau, and L. Mei, "Closed-form BER expressions of QPSK constellation for uplink non-orthogonal multiple access," *IEEE Communications Letters*, vol. 21, no. 10, pp. 2242–2245, 2017.

[78] E. Carmona Cejudo, H. Zhu, and O. Alluhaibi, "On the power allocation and constellation selection in downlink NOMA," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, 2017, pp. 1–5.

[79] E. Carmona Cejudo, H. Zhu, and J. Wang, "Resource allocation in multicarrier noma systems based on optimal channel gain ratios," *IEEE Transactions on Wireless Communications, under review*, 2020.

[80] P. A. Hoeher and T. Wo, "Superposition modulation: myths and facts," *IEEE Communications Magazine*, vol. 49, no. 12, pp. 110–116, Dec. 2011.

[81] L. F. Wei, "Coded modulation with unequal error protection," *IEEE Transactions on Communications*, vol. 41, no. 10, pp. 1439–1449, Oct 1993.

[82] H. Jiang and P. A. Wilford, "A hierarchical modulation for upgrading digital broadcast systems," *IEEE Transactions on Broadcasting*, vol. 51, no. 2, pp. 223–229, Jun. 2005.

[83] F. Kara and H. Kaya, "BER performances of downlink and uplink NOMA in the presence of SIC errors over fading channels," *IET Communications*, vol. 12, no. 15, pp. 1834–1844, 2018.

[84] Y. Yin, Y. Peng, M. Liu, J. Yang, and G. Gui, "Dynamic user grouping-based NOMA over Rayleigh fading channels," *IEEE Access*, vol. 7, pp. 110 964–110 971, 2019.

[85] 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. 134 539–134 555, 2019.

[86] 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 Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10 807–10 821, 2019.

[87] A. K. Dutta, "MBER criterion assisted power NOMA design and performance analysis with estimated channel," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 11 816–11 826, 2019.

[88] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, "On the performance of non-orthogonal multiple access systems with partial channel information," *IEEE Transactions on Communications*, vol. 64, no. 2, pp. 654–667, 2016.

[89] S. Guo and X. Zhou, "Robust resource allocation with imperfect channel estimation in NOMA-based heterogeneous vehicular networks," *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 2321–2332, 2019.

[90] Z. Wei, D. W. K. Ng, and J. Yuan, "Power-efficient resource allocation for MC-NOMA with statistical channel state information," in *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–7.

[91] F. Fang, H. Zhang, J. Cheng, S. Roy, and V. C. M. Leung, "Joint user scheduling and power allocation optimization for energy-efficient NOMA systems with imperfect CSI," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 12, pp. 2874–2885, 2017.

[92] F. Kara and H. Kaya, "Improved user fairness in decode-forward relaying non-orthogonal multiple access schemes with imperfect SIC and CSI," *IEEE Access*, vol. 8, pp. 97 540–97 556, 2020.

[93] J. Proakis, *Digital communications*, 5th ed. Mc-Graw Hill International Editions, 2008.

[94] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, 2013, pp. 1–5.

[95] B. Makki, K. Chitti, A. Behravan, and M. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Open J. Commun. Society*, vol. 1, pp. 179–189, 2020.

[96] E. Carmona Cejudo, H. Zhu, J. Wang, and O. Alluhaibi, "A fast algorithm for resource allocation in downlink multicarrier noma," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, 2019, pp. 1–5.

[97] E. Carmona Cejudo, H. Zhu, and J. Wang, "Resource allocation in downlink multicarrier NOMA under a fairness constraint," in *2020 IEEE 92nd Vehicular Technology Conference (VTC-Fall)*, 2020.

[98] E. Carmona Cejudo, H. Zhu, and J. Wang, "Resource allocation in BER-constrained multicarrier NOMA systems," in *2021 IEEE International Conference on Communications (ICC), submitted.*

[99] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 537–552, Jan 2016.

[100] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.

[101] J. Guo, X. Wang, J. Yang, J. Zheng, and B. Zhao, "User pairing and power allocation for downlink non-orthogonal multiple access," in *2016 IEEE Globecom Workshops*, Dec. 2016, pp. 1–6.

[102] J. He and Z. Tang, "Low-complexity user pairing and power allocation algorithm for 5G cellular network non-orthogonal multiple access," *IET Electronics Letters*, vol. 53, no. 9, pp. 626–627, May 2017.

[103] J. A. Oviedo and H. R. Sadjadpour, "A fair power allocation approach to NOMA in multi-user SISO systems," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, Mar. 2017.

[104] W. Chen, S. Zhao, R. Zhang, and L. Yang, "Generalized user grouping in NOMA based on overlapping coalition formation game," *IEEE Journal on Selected Areas in Communications*, pp. 1–1, 2020.

[105] L. Chen, L. Ma, and Y. Xu, "Proportional fairness-based user pairing and power allocation algorithm for non-orthogonal multiple access system," *IEEE Access*, vol. 7, pp. 19 602–19 615, 2019.

[106] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Joint optimization of power and channel allocation with non-orthogonal multiple access for 5G cellular systems," in *2015 IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–6.

[107] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems," *IEEE Transactions on Communications*, vol. 65, no. 3, pp. 1077–1091, Mar. 2017.

[108] J. Zhu, J. Wang, Y. Huang, S. He, X. You, and L. Yang, "On optimal power allocation for downlink non-orthogonal multiple access systems," *IEEE Journal on Selected Areas in Communications*, vol. PP, no. 99, pp. 1–1, 2017.

[109] W. Cai, C. Chen, L. Bai, Y. Jin, and J. Choi, "Subcarrier and power allocation scheme for downlink OFDM-NOMA systems," *IET Signal Processing*, vol. 11, no. 1, pp. 51–58, 2017.

[110] L. Salaün, M. Coupechoux, and C. S. Chen, "Joint subcarrier and power allocation in NOMA: Optimal and approximate algorithms," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2215–2230, 2020.

[111] W. Liang, Z. Ding, Y. Li, and L. Song, "User pairing for downlink non-orthogonal multiple access networks using matching algorithm," *IEEE Transactions on Communications*, vol. 65, no. 12, pp. 5319–5332, 2017.

[112] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems - part II: joint chunk, power and bit allocation," *IEEE Transactions on Communications*, vol. 60, no. 2, pp. 499–509, Feb. 2012.

[113] X. Chen, Z. Zhang, C. Zhong, R. Jia, and D. W. K. Ng, "Fully non-orthogonal communication for massive access," *IEEE Transactions on Communications*, vol. 66, no. 4, pp. 1717–1731, 2018.

[114] Y. Sui, W. Yu, and Q. Luo, "Jointly optimized extreme learning machine for short-term prediction of fading channel," *IEEE Access*, vol. 6, pp. 49 029–49 039, 2018.