
PhD-FSTM-2023-145
The Faculty of Sciences, Technology and Medicine

DISSERTATION

Defence held on 15/12/2023 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG
EN INFORMATIQUE

by

Duc-Dung TRAN
Born on 27 July 1990 in Thua Thien Hue (Vietnam)

# DESIGN AND OPTIMIZATION OF ULTRA-RELIABLE LOW-LATENCY COMMUNICATIONS IN BEYOND 5G NETWORKS

## Dissertation defence committee

Dr. Symeon Chatzinotas, dissertation supervisor
*Professor, Université du Luxembourg*

Dr. Miguel A. Olivares-Mendez, Chairman
*Professor, Université du Luxembourg*

Dr. Eva Lagunas, Vice Chair
*Research Scientist, Université du Luxembourg*

Dr. Petar Popovski
*Professor, Aalborg University*

Dr. Le-Nam Tran
*Professor, University College Dublin*

# Abstract

The fifth generation (5G) and beyond wireless networks mark a pivotal shift in the realm of telecommunications. These advanced networks aim to provide an array of different services, fulfilling the diverse needs of modern-day connectivity. They are expected to provide services with high data rates, large connection density, ultra-low latency, and extraordinary reliability. To achieve these goals, there are three primary service categories in 5G and beyond networks: enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable low-latency communications (URLLC). With eMBB, users can communicate with a substantial increase in data rates, enabling swift and high-bandwidth content consumption. On the other hand, mMTC sets the stage for the seamless integration of billions of devices into the network. However, it's URLLC that stands out as the linchpin of these networks, providing unprecedented levels of reliability and ultra-low latency, considering mission-critical applications and real-time responsiveness as the norm. This service is expected to open groundbreaking changes in fields such as healthcare, autonomous vehicles, industrial automation, and beyond. Given the above context, this dissertation focuses on designing effective communication protocols for different URLLC-related systems. In particular, the study delves into three key aspects: (1) Average block error rate (BLER) and minimum blocklength analysis for short-packet communications, a promising transmission method for URLLC; (2) Deep reinforcement learning (DRL)-based resource management strategy for uplink URLLC within the context of grant-free access, an advanced access technology for latency-sensitive dense networks; and (3) Joint optimization and DRL-based resource allocation for harmonious coexistence of diverse services such as eMBB, mMTC, and URLLC.

Firstly, we study a promising transmission method for URLLC, namely short packet communications (SPC), to fulfill its stringent requirements. Specifically, we investigate SPC in downlink non-orthogonal multiple access (NOMA) systems using multiple-input

multiple-output (MIMO) schemes. The main focus of this work is a comprehensive evaluation of system performance by analyzing the average block error rate (BLER) and minimizing the blocklength to reduce transmission latency. Our findings reveal that MIMO NOMA exhibits the capability to efficiently serve multiple users in a concurrent fashion while employing a lower blocklength in comparison with MIMO Orthogonal Multiple Access (OMA). These results effectively highlight the advantages of MIMO NOMA-based SPC, primarily in its ability to significantly reduce transmission latency.

Secondly, we investigate the application of DRL techniques for designing highly efficient resource management solutions in grant-free NOMA (GF-NOMA) systems tailored to meet the stringent demands of URLLC. Our focus centers on maximizing network energy efficiency (EE) and ensuring the fulfillment of URLLC users' specific requirements. The outcomes of our simulations demonstrate that the methods we propose achieve better convergence properties, smaller signaling overhead, and larger network EE than other benchmark methods.

Finally, our focus turns to the seamless combination of diverse services including eMBB, mMTC, and URLLC in NOMA-based systems. In this context, we develop an innovative resource management solution applying a joint optimization and cooperative multi-agent DRL approach. The primary goal of this strategy is to maximize network EE for the considered system while adhering to users' diverse demands. Our extensive simulations indicate that our proposed method provides superior performance regarding convergence property and system EE over other considered benchmark methods.

# Acknowledgements

First and foremost, I would like to express my heartfelt gratitude, to my supervisors Prof. Symeon Chatzinotas and Dr. Shree Krishna Sharma, for their invaluable instruction, motivation, and patience. Throughout my doctoral journey, they dedicatedly helped me determine research direction and improve my research skills. Their expertise, encouragement, and constructive feedback have not only enriched my knowledge but also instilled in me the confidence to overcome challenges in research and life. Also, I would like to express my sincere gratitude to the external member of my CET committee, Prof. Petar Popovski, for giving me valuable suggestions which have helped me significantly improve the quality of the papers. Furthermore, I would like to thank the members of the dissertation defence committee, Prof. Symeon Chatzinotas, Prof. Petar Popovski, Prof. Le-Nam Tran, Prof. Miguel A. Olivares-Mendez, and Dr. Eva Lagunas, for dedicating their time to review my thesis and providing me with valuable comments to enhance the quality of the dissertation. In addition, I am thankful to Dr. Vu Nguyen Ha for his constructive comments and suggestions to refine ideas and revise the papers.

Moreover, I would like to sincerely thank my research collaborators, Dr. Shree Krishna Sharma, Dr. Vu Nguyen Ha, Prof. Symeon Chatzinotas, Prof. Björn Ottersten, and Prof. Isaac Woungang, and Dr. Ti Ti Nguyen, for giving me valuable advice and suggestions to enhance my technical and writing skills, the overall quality of my research papers, and have a right direction for my doctoral journey. Furthermore, I am thankful to my kind colleagues at SnT for their enthusiastic support and for fostering a conducive work environment. These have not only enriched my professional journey but have also contributed to the enjoyable moments I have experienced at SnT.

I am also deeply grateful to my devoted parents and family for their nurturing, boundless love, and unwavering support. Thanks to their unconditional sacrifices and tireless efforts, I have been given the opportunity to pursue my studies. Additionally, I would like to express my special thanks to my beloved wife, Han Dan Anh Nguyen, for her understanding, sacrifices and unwavering assistance. She has always been my rock that helped me overcome the difficulties and challenges I faced during my doctoral journey. I am also thankful to my beloved son, Tony, for filling my daily life with love and joyous moments.

Moreover, I would like to extend my appreciation to my dear friends who have stood by my side during challenging moments.

Last but not least, I would like to sincerely thank the University of Luxembourg community for providing me with an inspiring environment and numerous opportunities for personal and professional growth. The generous financial help from the Fonds National de la Recherche (FNR-Luxembourg National Research Fund) via the University of Luxembourg is gratefully acknowledged.

# Preface

This Ph.D. Thesis has been carried out from January, 2020 to December, 2023 at the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg, under the supervision of Prof. Symeon Chatzinotas and Dr. Shree Krishna Sharma at SnT, University of Luxembourg, Luxembourg. The time-to-time evaluation of the Ph.D. Thesis was duly performed by the CET members constituting the supervisors at SnT, University of Luxembourg, Luxembourg and Prof. Petar Popovski from Aalborg University, Denmark.

## Contents

This Ph.D. Thesis entitled *"Design and Optimization of Ultra-Reliable and Low-Latency Communications in Beyond 5G Networks"* is divided into six chapters. In Chapter 1, the literature review, limitations of existing works, and contributions of this thesis are described. Chapter 2 provides common knowledge about different techniques applied to URLLC-related scenarios in this thesis such as short-packet communications (SPC), grant-free access, machine learning, and non-orthogonal multiple access (NOMA). Chapter 3 presents the analysis of block error rate and minimum blocklength of the SPC in low-latency multiple-input multiple-output NOMA systems under Nakagami-$m$ fading conditions. Chapter 4 focuses on resource management strategy based on multi-agent deep reinforcement learning (MADRL) for URLLC in grant-free NOMA networks to optimize the network performance regarding energy efficiency. In Chapter 5, a joint optimization and cooperative MADRL framework is proposed for the coexistence of different services based on heterogeneous NOMA (H-NOMA) with the purpose of optimizing network energy efficiency while guaranteeing the diverse requirements from various services. Finally, Chapter 6 offers the conclusions and potential avenues for future research.

## Support of the Thesis

This Ph.D. Thesis has been supported by the Luxembourg National Research Fund under Project FNR CORE 5G-Sky, grant C19/IS/13713801 and ERC-funded project AGNOSTIC, grant 742648. Additionally, the time-to-time support from SIGCOM is also gratefully acknowledged.

# List of Tables

2.1 Reliability and latency requirements for different applications . . . . . 28

4.1 Solution Summary of Related Works . . . . . 85
4.2 Simulation Parameters . . . . . 86

5.1 Experimental Parameters . . . . . 118

# List of Figures


2.1 Potential URLLC applications . . . . . 27
2.2 Key enablers for URLLC scenarios. . . . . 28
2.3 Handshake procedures of GB and GF access methods . . . . . 30
2.4 Machine learning approaches. . . . . 32
2.5 Supervised learning algorithms classification. . . . . 33
2.6 Unsupervised learning algorithms classification. . . . . 33
2.7 Reinforcement learning algorithms classification. . . . . 34
2.8 Illustration of two-user power-domain NOMA . . . . . 35

3.1 Model of a MIMO NOMA system under SPC over Nakagami-$m$ fading. . . 44
3.2 Rate comparison between SPC and LPC, where $\gamma_0 = 20$ (dB), $m_H = m_L = 2$, and $K_S = K_H = K_L = I = J = 2$. . . . . 61
3.3 Average BLER at user $H_{\hat{i}}$ vs. $\gamma_0$ with different methods, where $m_H = m_L = 2$ and $(K_S, K_H, I) = (2, 2, 1)$. . . . . 61
3.4 Average BLER at user $L_{\hat{j}}$ vs. $\gamma_0$ with different methods, where $\psi = 0$, $m_H = m_L = 2$ and $(K_S, K_L, J) = (2, 2, 1)$. . . . . 62
3.5 Average BLER at user $H_{\hat{i}}$ vs. $\gamma_0$ with different values of $(K_S, K_H, I)$, where $m_H = m_L = 2$. . . . . 62
3.6 Average BLER at user $L_{\hat{j}}$ vs. $\gamma_0$ with different values of $(K_S, K_L, J)$, where $\psi = 0$ and $m_H = m_L = 2$. . . . . 63
3.7 Average BLER at user $L_{\hat{j}}$ vs. $\gamma_0$ with different values of $\psi$ (residual interference caused due to the ISIC process), where $m_H = m_L = 2$ and $(K_S, K_L, J) = (2, 2, 1)$. . . . . 63
3.8 Average BLER at users $H_{\hat{i}}$ and $L_{\hat{j}}$ vs. $m$ with different methods, where $\gamma_0 = 20$ (dB) and $(K_S, K_H, K_L, I, J) = (2, 2, 2, 1, 1)$. . . . . 64

3.9 Minimum blocklength for users $H_{\hat{i}}$ and $L_{\hat{j}}$ vs. $\alpha_{L_{\hat{j}}}$ with different methods, where $\psi = 0$, $m_H = m_L = 2$, $K_S = K_H = K_L = I = J = 2$, and $\gamma_0 = 20$ (dB). 64
3.10 Blocklength comparison between NOMA and OMA. . . . . 65

4.1 Illustration of an uplink URLLC-GF-NOMA system. . . . . 72
4.2 Illustration of DQN/2DQN model. . . . . 79
4.3 Illustration of 3DQN model. . . . . 82
4.4 Convergence analysis with different approaches, where $M = 4$, $K = 2$, $L = 7$. 87
4.5 Convergence analysis with different network states and MA3DQN method, where $M = 4$, $K = 2$, $L = 7$. . . . . 88
4.6 Effect of state-action spaces on the achieved reward with different approaches. 89
4.7 Effect of URLLC requirements $(\varepsilon_m, \tau)$ on the achieved reward, where $M = 4$, $K = 2$, and $L = 10$. . . . . 90
4.8 Performance comparison between the methods using GF-OMA and GF-NOMA, where $M = 4$, $L = 10$. . . . . 90
4.9 Effect of number of users on the EE performance with different approaches, where $K = 2$, $L = 10$. . . . . 91
4.10 EE performance comparison between different methods, where $M = 4$, $K = 2$. 91
4.11 EE performance of MADQN method with centralized and decentralized rewards. . . . . 92
4.12 Achievable sum rate and power consumption of different problems, where $M = 4$, $K = 2$, and $L = 7$. . . . . 93
4.13 EE performance of different MADRL solutions for GF-NOMA systems, where $M = 4$ and $K = 2$. . . . . 94
4.14 EE performance of different SIC methods, where $M = 4$, $K = 2$, and $L = 7$. 94

5.1 Illustration of a H-NOMA-based uplink system. . . . . 101
5.2 Convergence performance with different learning rate values. . . . . 119
5.3 Convergence performance of learning approaches with different values of $K$. 119
5.4 Convergence performance of learning approaches with different values of $L$. 120
5.5 Energy efficiency of different approaches vesrus number of mMTC users. . . 120
5.6 Energy efficiency of different approaches versus maximum transmission power. 121
5.7 Energy efficiency of different approaches versus URLLC/eMBB requirement set, where $K = 2$. . . . . 121

# Contents

**Abstract** **2**

**Acknowledgements** **4**

**Preface** **7**
Contents . . . 7
Support of the thesis . . . 7

**List of Tables** **9**

**List of Figures** **11**

**1 Introduction** **16**
1.1 Related Works . . . 17
1.1.1 Short Packet Communications for URLLC-Related Systems . . . 17
1.1.2 Resource Allocation Optimization for URLLC Systems . . . 18
1.1.3 Coexistence of eMBB, mMTC, and URLLC Heterogeneous Services 19
1.2 Limitations of Existing Works . . . 20
1.3 Thesis Contributions . . . 20
1.4 Other contributions beyond the scope of the thesis . . . 23
Journal papers . . . 23
Conference Papers . . . 23

**2 Background** **26**
2.1 Short Packet Communications . . . 27
2.2 Grant-Free Access . . . 29
2.3 Machine Learning for Wireless Communications . . . 31

2.4 Non-Orthogonal Multiple Access . . . . . 35

**3 Short-Packet Communications in URLLC-Enabled Systems: BLER and Minimum Blocklength Analysis 38**
3.1 Introduction . . . . . 39
3.2 System Model . . . . . 43
3.2.1 Antenna and User Selection . . . . . 44
3.2.2 Information Transmission Process and Channel Statistics . . . . . 46
3.3 Proposed approach for BLER performance analysis with SPC . . . . . 48
3.3.1 Preliminaries . . . . . 48
3.3.2 Derivation for Cumulative Distribution Function (CDF) of Channel Power Gains . . . . . 50
3.3.3 Average BLER Analysis of HCS Method . . . . . 52
3.3.4 Average BLER Analysis of LCS Method . . . . . 53
3.4 Proposed analytical framework for optimal power allocation and minimum blocklength . . . . . 54
3.4.1 Asymptotic Average BLER Analysis . . . . . 55
3.4.2 Power and Blocklength Optimization at High SNR . . . . . 57
3.4.3 Comparison of MIMO NOMA and MIMO OMA Schemes . . . . . 59
3.5 Numerical Results . . . . . 60
3.6 Summary . . . . . 65

**4 Deep Reinforcement Learning for Resource Allocation Optimization in URLLC Systems 67**
4.1 Introduction . . . . . 68
4.2 System Model . . . . . 71
4.2.1 Uplink GF-NOMA Transmission Process . . . . . 72
4.2.2 URLLC Communication Model . . . . . 73
4.2.3 Energy Efficiency Maximization . . . . . 74
4.3 MADRL-Based Energy Efficiency Resource Allocation Solution For URLLC-GF-NOMA Systems . . . . . 75
4.3.1 MADRL Framework . . . . . 76
4.3.2 Proposed MADRL Algorithms For URLLC-GF-NOMA Systems . . 79
4.3.3 Analysis of The Proposed Methods . . . . . 83
4.4 Simulation Results . . . . . 85
4.5 Summary . . . . . 95

**5 Coexistence of eMBB, mMTC, and URLLC Heterogeneous Services 97**
5.1 Introduction . . . . . 98
5.2 System Model and Problem Formulation . . . . . 101

5.2.1 Uplink Transmission Strategy for H-NOMA . . . 102
5.2.2 Achievable Rate of Users . . . 104
5.2.3 Energy Efficiency Maximization Problem . . . 106
5.3 Proposed Joint Optimization and Cooperative MADDQN Method . . . 106
5.3.1 Power Allocation for Given SC Assignment . . . 107
5.3.2 CMADDQN-based SC Assignment Strategy . . . 111
5.4 Proposed Distributed Reinforcement Learning Method . . . 114
5.4.1 FDDQN Method . . . 114
5.4.2 Complexity Analysis . . . 116
5.4.3 Convergence Discussion . . . 117
5.5 Simulation Results . . . 118
5.6 Summary . . . 122

**6 Conclusions and Future Research** **124**
6.1 Conclusions . . . 124
6.2 Potential Avenues for Future Research . . . 125

**Appendix A Appendices for Chapter 3** **130**
A.1 Proof for Proposition 1 in Chapter 3 . . . 130
A.2 Proof of Theorem 1 in Chapter 3 . . . 131
A.3 Proof of Theorem 2 in Chapter 3 . . . 132

**Bibliography** **134**

Chapter 1

# Introduction

The fifth generation (5G) and beyond wireless networks mark a pivotal shift in the realm of telecommunications. These advanced networks aim to provide an array of different services, fulfilling the diverse needs of modern-day connectivity. They are expected to provide services with high data rates, large connection density, ultra-low latency, and extraordinary reliability [1]. To achieve these goals, there are three primary service categories in 5G and beyond: enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable low-latency communications (URLLC). With eMBB, users can communicate with a substantial increase in data rates, enabling swift and high-bandwidth content consumption. On the other hand, mMTC sets the stage for the seamless integration of billions of devices into the network. However, it's URLLC that stands out as the linchpin of these networks, providing unprecedented levels of reliability and ultra-low latency, considering mission-critical applications and real-time responsiveness as the norm [2].

URLLC services are designed to meet the most stringent demands of applications where reliability and ultra-low latency are mission-critical. These services are expected to open groundbreaking changes in fields such as healthcare, autonomous vehicles, industrial automation, and beyond [3]. In a URLLC environment, data is transmitted with an almost imperceptible delay, ensuring real-time interactions and precise control in scenarios where split-second decisions are imperative. In particular, a general URLLC requirements has been defined by the Third Generation Partnership Project (3GPP) standard [4, 5], specifying the need for a reliability level of $1 - 10^{-5}$ within 1 ms user plane latency for a data payload of 32 bytes. Despite its immense benefits, URLLC still face several challenges that must be addressed for their successful implementation. Within this chapter, we offer a comprehensive examination of recent advancements in URLLC and its related technologies in recent years. Subsequently, we highlight the constraints encountered in these prior studies and clarify our contributions to this field.

## 1.1 Related Works

In this section, we show an insight into the closest studies relevant to the primary contributions made in this dissertation. Specifically, we present the related works on short packet communications (SPC) for URLLC-related systems, optimization for resource management in URLLC-related systems, and the coexistence of diverse services such as eMBB, mMTC, and URLLC.

### 1.1.1 Short Packet Communications for URLLC-Related Systems

URLLC has emerged as a highly potential candidate for 5G and future networks, catering to the demands of novel applications with unparalleled requirements in terms of both reliability and latency [6, 7]. In this regard, a novel communication method, known as SPC, is gaining significant traction to meet these strict criteria. This approach arises due to the demand for advanced transmission techniques tailored to latency-sensitive networks, a context where conventional analytical methods relying on longer packet communications prove inadequate [8]. The assessment of SPC performance has prompted the introduction of new evaluation metrics in the research domain [9, 10], such as the ratio of pilots to the information payload, namely overhead ratio, and block error rate (BLER).

Recently, there have been many studies on SPC under different scenarios such as performance analysis regarding average achievable rate and BLER [11–18], SPC optimization [19–21], and application of machine learning to SPC [22]. Specifically, considering performance analysis, the work in [11] considered SPC based on a single-input single-output (SISO) NOMA scheme over Rayleigh fading, where base station (BS) communicates with two users. The network performance is characterized by deriving the average BLER. In [12], a NOMA scenario in stochastic geometry was investigated and average BLER was analyzed to investigate system performance. In [13], Lai *et al.* evaluated the performance of a cooperative NOMA SPC system, where average BLER is derived. In [14], a space diversity technique, namely multiple-input single-output (MISO), was applied for NOMA-based SPC systems using wireless power transfer (WPT) scheme with the aim of evaluating the outage probability of the considered systems. The study conducted in reference [15] explored the application of a multiple-input multiple-output (MIMO) scheme to NOMA-based SPC. This investigation involved the derivation of an upper bound for the probability of violating the delay target utilized for optimizing the transmission power. The scenario considered in [15] involved each pair of NOMA users being served by an individual transmit antenna. Additionally, the work described in [16] explored the marriage of transmit antennas based on maximum ratio transmission (MRT) for serving NOMA users with the aim of reducing BLER for SPC in MISO NOMA systems. In [17], Hashemi *et al.* carried out the performance analysis based on BLER and average rate for SPC assisted by reconfigurable intelligent surface (RIS). In [18], the application of RIS to SPC in

NOMA systems was examined by deriving the average BLER expression for performance evaluation.

For SPC optimization, the work in [19,20] investigated the problem of optimizing transmission energy for SPC-assisted NOMA systems subject to heterogeneous delay conditions. The work in [21] designed an optimal transmission strategy aimed at optimizing the average rate for SPC in a MISO network. In addition, another study in [22] explored the application of machine learning to SPC in WPT-enabled multi-hop systems. The authors in this work first derived a closed-form expression of BLER and performed the asymptotic analysis to evaluate system performance. They then formulated a problem of optimizing throughput and designed an effective deep learning framework toward real-time settings using deep convolutional neural network (CNN) to address effectively this problem.

### 1.1.2 Resource Allocation Optimization for URLLC Systems

There have been several different optimization problems considered in the literature to design URLLC protocols [23–31]. Specifically, the work in [23] investigated joint communication and computation offloading for hierarchical edge-cloud systems with URLLC, where a latency minimization problem of computational tasks among multiple industrial Internet of Things (IIoT) devices was examined. In [24,25], the authors studied the minimization of decoding error probability while adhering to latency conditions for URLLC-enabled UAV relay and factory automation systems. In [26], Sun *et al.* optimized resource management for URLLC, where the network energy efficiency (EE) is maximized by optimizing antenna configuration, bandwidth allocation, and power control under the constraints on reliability and latency.

Taking intelligent features into account, the application of reinforcement learning (RL) to resource management in URLLC-enabled systems has been conducted in [27–31]. In particular, the work in [32] studied the dynamic channel allocation for URLLC traffic in a multi-user multi-channel wireless network, guaranteeing that urgent packets have to be successfully received in a timely manner. A Q-learning algorithm where the controller learns the optimal policy under the absence of channel state information (CSI) and the channel statistics was proposed for channel allocation problem. However, Q-learning algorithm needs to build a Q-table for all network space (state-action space), hence it has low convergence speed and is only applicable to small network, making it difficult to apply for complicated issues in ultra-dense networks. To mitigate this challenge, Yang *el al.* [28] applied decentralized actor-critic RL model to resolve the resource management problem in URLLC internet of vehicles communication networks with the purpose of maximizing the sum capacity while ensuring the URLLC requirements. In [29], the authors investigated a power consumption minimization issue under reliability and latency conditions in an orthogonal frequency division multiple access (OFDMA) system. They then proposed a

DRL-based solution to resolve the problem by using generative adversarial networks. Considering grant-free transmissions as a promising method to guarantee URLLC requirements in uplink massive access scenarios, the works in [30, 31] proposed methods based on deep RL (DRL) techniques to attain optimal average throughput (i.e., number of users served successfully) in massive URLLC (mURLLC) scenarios using grant-free NOMA.

### 1.1.3 Coexistence of eMBB, mMTC, and URLLC Heterogeneous Services

Multiplexing eMBB, mMTC, and URLLC enables the coexistence of these services on the same network. This is a major problem to be tackled in future wireless networks to support diverse applications satisfying the high demands from these services. However, they are considered as heterogeneous services due to their diverse requirements. This leads to a crucial challenge in terms of resource management to ensure the coexistence of these different services. To overcome this issue, network slicing can be utilized to accommodate these services on the same network architecture [33–36]. In particular, Alsenwi *et al.* [33] investigated a scenario of risk-sensitive eMBB-URLLC network slicing in downlink OFDMA transmissions. In [34], a radio access network (RAN) eMBB-URLLC resource slicing within OFDMA-based 5G networks was examined, where a sum-rate optimization subject to data rate and URLLC requirements are considered. In [35], a coexistence between eMBB and URLLC based on puncturing technique in downlink 5G networks was examined, where an optimization problem of the minimum expected achieved rate according to eMBB service while maintaining the provisions of the URLLC traffic is investigated. In addition, the coexistence of different services in downlink OFDMA systems was considered in [36].

Network slicing is usually conducted based on OMA scenario. However, the exponential growth of the number of users leads to a demand for a more flexible and efficient multiplexing method. In this regard, NOMA-based multiplexing scheme is demonstrated as a potential method for enhancing network efficiency in uplink transmissions, as discussed in [37–40]. Specifically, in [37], a NOMA-based network slicing scheme for eMBB, URLLC, and mMTC services was investigated and indicated that NOMA-based slicing can outperform OMA-based one in guaranteeing heterogeneous requirements under some considered scenarios. In [38], the authors examined NOMA-based slicing for eMBB and URLLC, where the power minimization problem was considered. In [39], the coexistence of eMBB and URLLC in MIMO NOMA scenarios was investigated. In [40], a network slicing method for multiplexing eMBB, URLLC, and mMTC using rate-splitting-based NOMA scheme was proposed.

In addition, machine learning-based multiplexing has been explored in [41–44] to develop intelligently dynamic resource allocation mechanisms fulfilling the various QoS requirements from different services and adapting to the wireless channel varying unpredictably over the time. Given this context, the works [41–44] studied the coexistence

scheme based on DRL techniques for different services in OFDMA-based [41, 42] and NOMA-based [43, 44] systems.

## 1.2 Limitations of Existing Works

In this section, we delve into the drawbacks and unaddressed challenges in the existing literature that serve as direct motivators for this dissertation. To provide a brief overview, we outline the principal limitations of the existing URLLC-related studies as follows:

- Although there have been many studies on SPC for URLLC-enabled systems, further investigation is still necessary to attain a comprehensive insights for SPC. In particular, the analysis of BLER and blocklength minimization for NOMA-based SPC considering MIMO schemes and Nakagami-$m$ fading - a general channel model, is missing in the related literature [11–16]. The missing scenario can achieve a significant improvement in network performance related to BLER and minimum blocklength aspects while providing more general understanding of the manners of SPC-enabled NOMA systems based on space diversity.

- Taking intelligent features into account, the works [27–31] considered different RL algorithms for resource allocation in URLLC systems under various optimization problems. However, energy-efficient maximization for URLLC-enabled grant-free NOMA systems has not been exploited in these works.

- Despite the coexistence of eMMB, mMTC, and URLLC services has been investigated under different scenarios in the existing literature, there are only a few works investigating intelligent multiplexing methods for different services based on uplink NOMA communications [43, 44]. Therefore, it is still an open research direction and needs more studies in future research.

Given the aforementioned constraints of URLLC-related scenarios, this dissertation aims to address these challenges. The following section provides a concise summary of the contributions made in this thesis.

## 1.3 Thesis Contributions

In the following, the objective of each chapter is provided with the purpose of highlighting the primary contributions achieved within this dissertation.

Chapter 3 investigates the SPC in MIMO NOMA systems over Nakagami-$m$ fading. Our objective is to analyze the average BLER parameter to evaluate the SPC performance in the considered system, based on which the blocklength minimization problem subject

to reliability (BLER) constraint is considered. The main contributions of this chapter are summarized as follows:

- Firstly, we introduce an innovative framework for analyzing the efficiency of SPC in NOMA-based MIMO systems under Nakagami-$m$ fading. This framework explores various MIMO strategies, including transmit antenna selection (TAS), maximal ratio combining (MRC), selection combining (SC).
- Secondly, we analyze the system performance by deriving the average BLER expressions for users. We then perform the analysis of its asymptotic behavior in the high signal-to-noise ratio (SNR), based on which we formulate and solve the blocklength minimization problem under reliability (BLER) constraint.
- Finally, we provide numerical results to evaluate BLER and blocklength performance of the considered system. Furthermore, we compare MIMO NOMA and MIMO OMA in terms of blocklength performance with the aim of highlighting the advantages of MIMO NOMA for low-latency communications.

The outputs of this chapter are published in:

**[J1]** **D. D. Tran**, S. K. Sharma, S. Chatzinotas, I. Woungang and B. Ottersten, "Short-Packet Communications for MIMO NOMA Systems Over Nakagami-m Fading: BLER and Minimum Blocklength Analysis," in *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3583-3598, April 2021, doi: 10.1109/TVT.2021.3066367.

In chapter 4, we investigate grant-free NOMA scheme for uplink URLLC, where energy-efficient (EE) resource management method is considered by formulating the EE maximization problem subject to users' URLLC requirements. We then design an efficient solution based on multi-agent (MA) DRL to solve the examined problem. The summary of the main contributions achieved in this chapter is outlined as follows:

- We investigate a URLLC-enabled grant-free NOMA system and formulate an average EE maximization problem under users' URLLC requirements. This requires the development of a swift and effective communication protocol.
- We design a decentralized resource management strategy based on MADRL to solve the considered issue by applying three different DRL techniques: Deep Q Network (DQN), Double DQN (DDQN), and Dueling DDQN (3DQN).
- We then carry out the performance evaluation, comparing the proposed methods and established benchmark approaches. This analysis aims to show the superiority of our methods with respect to convergence characteristics and network EE gain. The obtained simulation results highlight the superior performance achieved by our methods as compared to investigated benchmark approaches under network EE, convergence property, and signaling overhead aspects.

The achievements of this chapter are published in the following venues:

**[J2]** **D. D. Tran**, S. K. Sharma, V. N. Ha, S. Chatzinotas and I. Woungang, "Multi-Agent DRL Approach for Energy-Efficient Resource Allocation in URLLC-Enabled Grant-Free NOMA Systems," in *IEEE Open Journal of the Communications Society*, vol. 4, pp. 1470-1486, 2023, doi: 10.1109/OJCOMS.2023.3291689.

**[C1]** **D. D. Tran**, S. K. Sharma and S. Chatzinotas, "BLER-based Adaptive Q-learning for Efficient Random Access in NOMA-based mMTC Networks," *IEEE Vehicular Technology Conference (VTC2021-Spring)*, Helsinki, Finland, 2021, pp. 1-5, doi: 10.1109/VTC2021-Spring51267.2021.9448787.

**[C2]** **D. D. Tran**, S. K. Sharma, S. Chatzinotas and I. Woungang, "Q-Learning-Based SCMA for Efficient Random Access in mMTC Networks With Short Packets," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Finland, 2021, pp. 1334-1338, doi: 10.1109/PIMRC50174.2021.9569713.

**[C3]** **D. D. Tran**, V. N. Ha and S. Chatzinotas, "Novel Reinforcement Learning based Power Control and Subchannel Selection Mechanism for Grant-Free NOMA URLLC-Enabled Systems," *IEEE Vehicular Technology Conference: (VTC2022-Spring)*, Helsinki, Finland, 2022, pp. 1-5, doi: 10.1109/VTC2022-Spring54318.2022.9860574.

In chapter 5, we consider the coexistence of heterogeneous services such as eMBB, mMTC, and URLLC based on NOMA scheme. In this regard, we develop a novel resource allocation strategy based on a joint optimization and cooperative MADRL approach to maximize the network EE under heterogeneous QoS requirements from different users. The main contributions of this chapter are given as follows:

- We investigate the coexistence of diverse services in a NOMA-based uplink network, where eMBB and URLLC users are assigned orthogonally to a number of sub-channels (SCs) to fulfill their stringent QoS requirements on high reliability, low latency, and high data rate. Meanwhile, mMTC users can access any SCs freely and quickly without any admission approval from BS to improve the spectrum access efficiency and connectivity density.
- We formulate an EE maximization problem for the considered network under constraints on various QoS requirements of users.
- We design a novel learning-based resource allocation strategy to address the proposed problem. In particular, a cooperative MADDQN (CMADDQN) scheme centralized at the BS is utilized for SC assignment based on which a dynamic power allocation solution is developed to obtain optimal transmission power for users.

- We compare algorithm convergence and network EE of our proposed methods with other benchmark approaches to clarify the advantages of our innovative solutions.

The achievements of this chapter are published in the following venues:

**[J3]** **D. D. Tran**, V. N. Ha, S. K. Sharma, T. T. Nguyen, S. Chatzinotas, and P. Popovski, "Energy-Efficient NOMA for 5G Heterogeneous Services: A Joint Optimization and Deep Reinforcement Learning Approach", submitted to IEEE Transactions on Communications.

**[C4]** **D. D. Tran**, S. K. Sharma, S. Chatzinotas and I. Woungang, "Learning-Based Multiplexing of Grant-Based and Grant-Free Heterogeneous Services with Short Packets," *IEEE Global Communications Conference (GLOBECOM)*, Madrid, Spain, 2021, pp. 01-06, doi: 10.1109/GLOBECOM46510.2021.9685321.

**[C5]** **D.D. Tran**, V. N. Ha, S. Chatzinotas, and T. T. Nguyen, "A hybrid optimization and deep RL approach for resource allocation in semi-GF NOMA networks," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Toronto, ON, Canada, 2023.

Finally, chapter 6 provides the conclusions of this dissertation and a discussion about the potential avenues for future research.

## 1.4 Other contributions beyond the scope of the thesis

Throughout my doctoral studies, I have made contributions to several publications that are not included in this thesis. The specifics of these contributions are outlined below:

### Journal Papers

**[J4]** V. L. Nguyen, D. B. Ha, T. V. Truong, **D. D. Tran**, and S. Chatzinotas, "Secure Communication for RF Energy Harvesting NOMA Relaying Networks with Relay-User Selection Scheme and Optimization," in *Mobile Networks and Applications*, pp. 1719-1733, Apr. 2022.

**[J5]** V. H. Dang, T. D. Ho, H. Tran, **D. D. Tran**, H. L. Quoc, C. So-In, S. Chatzinotas, and V. N. Vo, "Performance Optimization for Hybrid TS/PS SWIPT UAV in Cooperative NOMA IoT Networks," submitted to IEEE Transactions on Green Communications and Networking.

### Conference Papers

**[C6]** T. M. Kebedew, V. N. Ha, E. Lagunas, **D. D. Tran**, J. Grotz, and S. Chatzinotas, "Reinforcement Learning for QoE-Oriented Flexible Bandwidth Allocation in Satellite Communication Networks," *2023 IEEE Global Communications Conference (GLOBECOM)*, accepted.

Chapter 2

# Background

In this chapter, we provide fundamental knowledge about different techniques utilized in this dissertation, such as short packet communications (SPC), grant-free access method, machine learning for wireless communications, and non-orthogonal multiple access (NOMA). Specifically, we present the basic concepts related to the above techniques to give a general insights about the scenarios considered in this thesis. The specific problems under investigation and the corresponding proposed solutions are presented in detail in the forthcoming chapters.

URLLC is expected to support innovative applications characterized by unparalleled demands for both reliability and latency [2]. Fig. 2.1 shows the potential URLLC applications. Historically, many applications such as industrial control systems are designed based on wired systems to guarantee the required reliability due to the limitations of existing wireless networks. The integration of URLLC leverages robust wireless links, bringing transformative advantages such as enhanced flexibility and reduced installation and maintenance expenses. Nevertheless, diverse application domains impose unique prerequisites concerning both reliability and latency levels. Each sector demands a tailored approach to ensure that URLLC's capabilities align precisely with their specific operational demands. Table 2.1 provides the general scenarios concerning the requirements of different applications in terms of reliability and latency [2].

Given the above context, there have been several key enablers considered promising technologies to fulfill different URLLC requirements from heterogeneous applications [45, 46], as shown in Fig. 2.2. In the following, we present an overview of some key enablers for URLLC applied in this dissertation, such as SPC, grant-free access, machine learning, and NOMA.

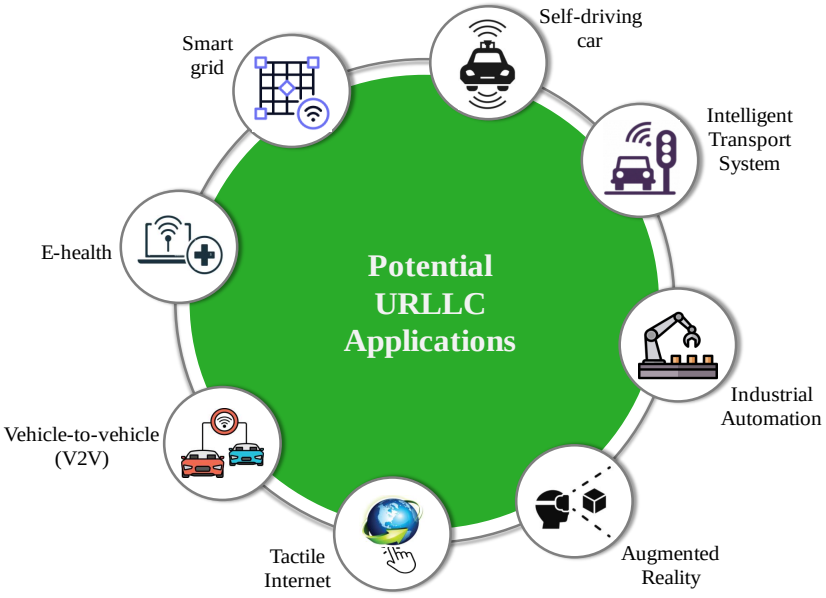



FIGURE 2.1: Potential URLLC applications

## 2.1 Short Packet Communications

SPC represent a crucial component within the realm of URLLC. Unlike traditional data transmission designed to target Shannon's channel capacity, which often involves larger packets of information, SPC is tailored to meet the requirements of mission-critical applications that rely on the rapid and reliable exchange of small, time-sensitive data packets. Let $N$, $C$, and $\varepsilon$ as blocklength, Shannon capacity, and block error rate (BLER), respectively. The achievable rate of SPC in finite blocklength (FBL) regime for a quasi-static flat fading channel can be expressed as [8, 9]

$$R = C - \sqrt{\frac{v}{N}} Q^{-1}(\varepsilon) + O\left(\frac{\log_2 N}{N}\right), \tag{2.1}$$

where $C = \log_2(1+\gamma)$, $v = 1 - 1/(1+\gamma)^2$ represents the channel dispersion, $\gamma$ is the signal-to-interference-plus-noise ratio (SINR), $Q^{-1}(x)$ is the inverse of the Gaussian Q-function, $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$, and $O\left(\frac{\log_2 N}{N}\right)$ is the remainder terms of order $\frac{\log_2 N}{N}$. In SPC, an important performance metric is BLER, which is determined approximately based on (2.1) as follows:

$$\varepsilon \approx Q\left(\frac{C - n/N}{\sqrt{v/N}}\right), \tag{2.2}$$

TABLE 2.1: Reliability and latency requirements for different applications

| Applications | Latency (ms) | Reliability (%) | Data Size (bytes) | Communication Range (m) |
|---|---|---|---|---|
| Smart grid | 3 - 20 | 99.999 | 80 - 1000 | 10 - 1000 |
| Self-driving car | 1 | 99 | 144 | 400 |
| ITS | 10 - 100 | 99.999 | 50 - 200 | 300 - 1000 |
| Industrial automation | 0.25 - 10 | 99.9999999 | 10 - 300 | 50 - 100 |
| Augmented reality | 0.4 - 2 | 99.999 | 12k - 16k | 100 - 400 |
| Tactile internet | 1 | 99.99999 | 250 | 100000 |
| V2V | 5 | 99.999 | 1600 | 300 |
| E-health | 30 | 99.999 | 28 - 1400 | 300 - 500 |

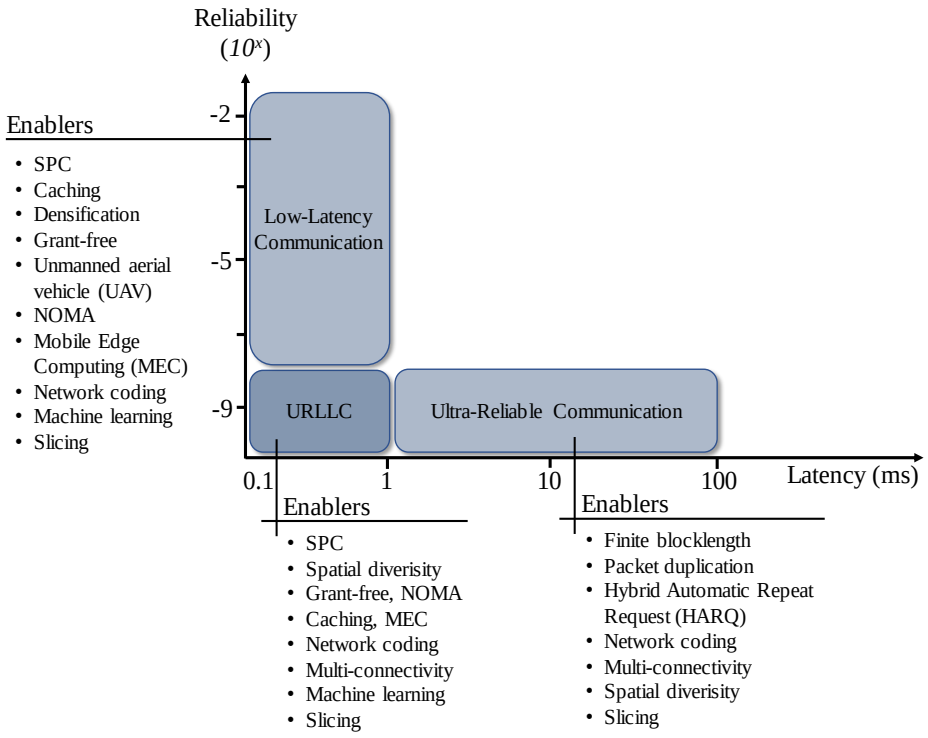



FIGURE 2.2: Key enablers for URLLC scenarios.

where $R = n/N$, $n$ is the number of information bits. It is noteworthy that the approximation in (2.2) is attained by omitting the term $O\left(\frac{\log_2 N}{N}\right)$ when $N \leq 100$ [9]. From (2.2),

the average BLER can be derived as

$$\bar{\varepsilon} = \int_0^\infty \varepsilon f_\gamma(x) dx, \tag{2.3}$$

where $f_X(x)$ denotes the probability density function (PDF) of a random variable $X$. However, direct derivation of $\bar{\varepsilon}$ in (2.3) is challenging due to the Gaussian Q-function of the instantaneous BLER $\varepsilon$. Therefore, an approximation method of $\varepsilon$ can be applied, which is given by [47]

$$\varepsilon \approx \begin{cases} 1, & \gamma \leq \nu \\ A, & \nu < \gamma < \mu \\ 0, & \gamma \geq \mu \end{cases}, \tag{2.4}$$

where $A = 0.5 - \chi\sqrt{N}(\gamma - \beta)$, $\chi = \sqrt{\frac{1}{2\pi\left(2^{2n/N}-1\right)}}$, $\nu = \beta - 1/2\chi\sqrt{N}$, $\mu = \beta + 1/2\chi\sqrt{N}$, and $\beta = 2^{n/N} - 1$. The average BLER in (2.3) then can be derived more easily by substituting (2.4) into (2.3) which is represented as

$$\bar{\varepsilon} \approx \chi\sqrt{N} \int_\nu^\mu F_\gamma(x) dx, \tag{2.5}$$

where $F_X(x)$ denotes the cumulative distribution function (CDF) of a random variable $X$.

## 2.2 Grant-Free Access

Grant-free (GF) access is a promising transmission method to reduce latency for URLLC, especially in massive access scenarios [48, 49]. It holds the promise of transforming the way critical communications are established in 5G and beyond networks. Specifically, unlike grant-based (GB) traditional access methods that rely on a central authority to grant access to users, GF access allows users to enter networks in a more flexible and immediate manner. In this access method, users can transmit their data freely without the need for prior authorization from the central controllers (e.g., base station, access point, etc.), streamlining the process of initiating real-time and mission-critical communications. This is particularly crucial for novel applications, such as autonomous vehicles, industrial automation, and emergency response systems, where instant access to the network can make a life-saving difference. In the following, the access protocols based on GF and conventional GB transmission strategies are presented to clarify the advantages of GF access method.

### Grant-Based Transmission Protocol

Fig. 2.3(a) depicts the conventional GB access protocol. In particular, its process can be indicated through the following handshake steps between users and base station (BS).

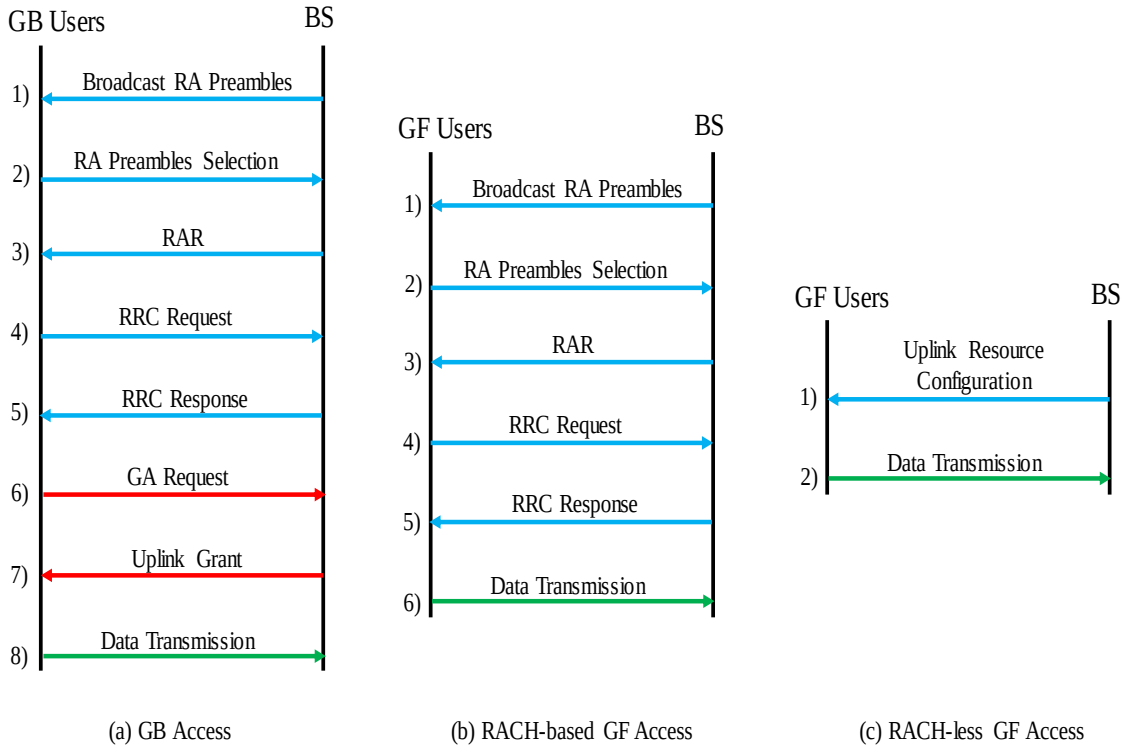



FIGURE 2.3: Handshake procedures of GB and GF access methods

1. In the initial step, the BS disseminates a list of its accessible random access (RA) preambles to the GB users within the network.

2. The GB users select one of the RA preambles and then upload their selections to the BS, a process aimed at identifying the channels they are occupying.

3. The BS transmits RA response (RAR), consisting of the information on optimal data rates, synchronization signals, and resource management.

4. The GB users send radio resource control (RRC) request, awaiting a response from the BS to determine their temporary identity.

5. The GB users are then granted to specific resource blocks. If there are no collisions, the GB users can utilize the assigned channels and send a connection request. Otherwise, they can go to the idle mode and wait until next time slot to perform the handshake procedure again.

6. The GB users can transmit their data after achieving the BS's permission.

It is noteworthy that the sequence of steps from one to five is referred to as RA process, while steps six and seven are known as grant acquisition (GA) process. Thus, although uplink grants conducted in GB access scheme can reduce the collision situations, it leads to long latency.

### Grant-Free Transmission Protocol

Basically, GF access is designed to reduce latency caused by the uplink grants implemented in GB access by mitigating the GA process. Specifically, GF access can be divided into two main categories: RA channel based (RACH-based) and RA channel less (RACH-less) GF transmissions [50]. Their handshake procedures are shown in Figs. 2.3(b) and 2.3(c), respectively. In RACH-based GF access method, the RA procedure is carried out, but after establishing synchronization with the BS, GF users can transmit their data immediately without conducting the GA process. Meanwhile, using RACH-less GF access method, GF users can begin their data transmission without performing any RA and GA procedures. Given this context, the GF access can offer a notable reduction in uplink transmission latency when compared to GB access, but the probability of collisions gets higher.

## 2.3 Machine Learning for Wireless Communications

Machine learning (ML) has emerged as a potential technology solution for next generation of wireless communications, i.e., 5G and beyond, changing how we design, manage, and optimize networks [51]. By leveraging data-driven insights and adaptive algorithms, ML techniques can bring superior capabilities to 5G and beyond wireless networks. Various machine learning approaches which is classified in Fig. 2.4, mainly including supervised learning (SL), unsupervised learning (UL), and reinforcement learning (RL), have been deployed to solve different problems in these networks, such as channel prediction [52], clustering and anomaly detection [53], network optimization and dynamic resource allocation [54]. Additionally, deep learning, a subfield of ML, has been exploited for complex tasks such as spectrum sensing, modulation recognition, and beamforming [55]. These techniques not only enhance the efficiency and reliability of wireless networks but also open the way for more adaptive and intelligent communication systems, making them crucial tools in the evolving landscape of wireless technologies.

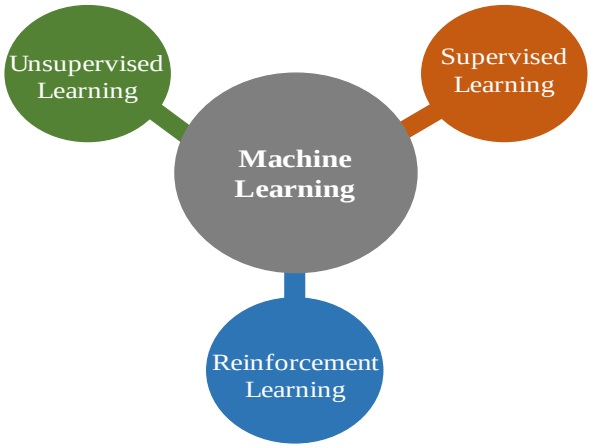



FIGURE 2.4: Machine learning approaches.

Taking different ML approaches into account, in SL method, models are trained on labeled data, where the correct answers are provided. The objective is to find a mapping or relationship between the inputs and their corresponding labels, allowing the model to make accurate predictions on new, unseen data. SL method can be broadly categorized into two main types: classification and regression, whose typical algorithms are provided in Fig. 2.5. In classification, the model is trained to assign input data to predefined categories or classes, while in regression, the model predicts a continuous numerical value based on the input.

On the other hand, UL methods deals with unlabeled data, where the model identifies hidden patterns or structures within the data. It is commonly applied for clustering problems. Its typical algorithms are shown in Fig. 2.6. Clustering algorithms aim to group similar data points into clusters, uncovering natural groupings in the data, which is valuable for tasks like customer segmentation and image categorization.

Finally, RL is a prominent subfield of ML enabling agents to make sequential decisions in an environment in order to maximize a cumulative reward. In RL, an agent interacts with an environment, where it takes actions and subsequently receives feedback, i.e., rewards or penalties, depending on the consequences of its decisions. In this regard, the objective is to learn a strategy or policy leading to a sequence of actions that bring the highest possible cumulative reward. RL problems are typically classified into two broad categories [56]: model-free and model-based methods. In model-free RL, agents learn directly from their interactions with the environment, aiming to optimize their policies through trial-and-error manner. In contrast, model-based RL involves constructing an internal model of the environment and using it to plan actions more effectively. Deep RL (DRL) is an exciting

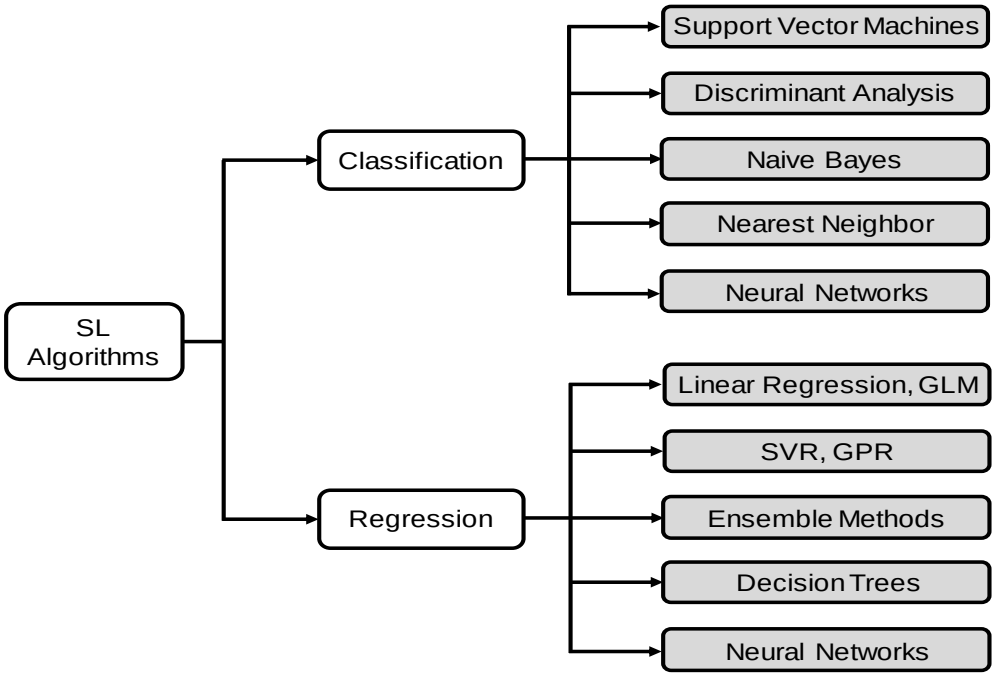



FIGURE 2.5: Supervised learning algorithms classification.

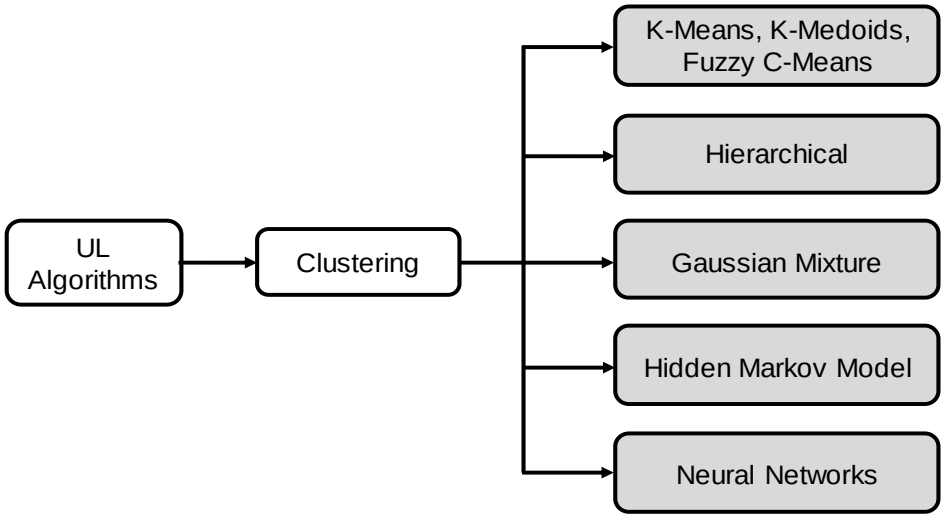



FIGURE 2.6: Unsupervised learning algorithms classification.

advancement within RL that incorporates deep neural networks (DNN) to handle high-dimensional state spaces and complex decision-making tasks [51].

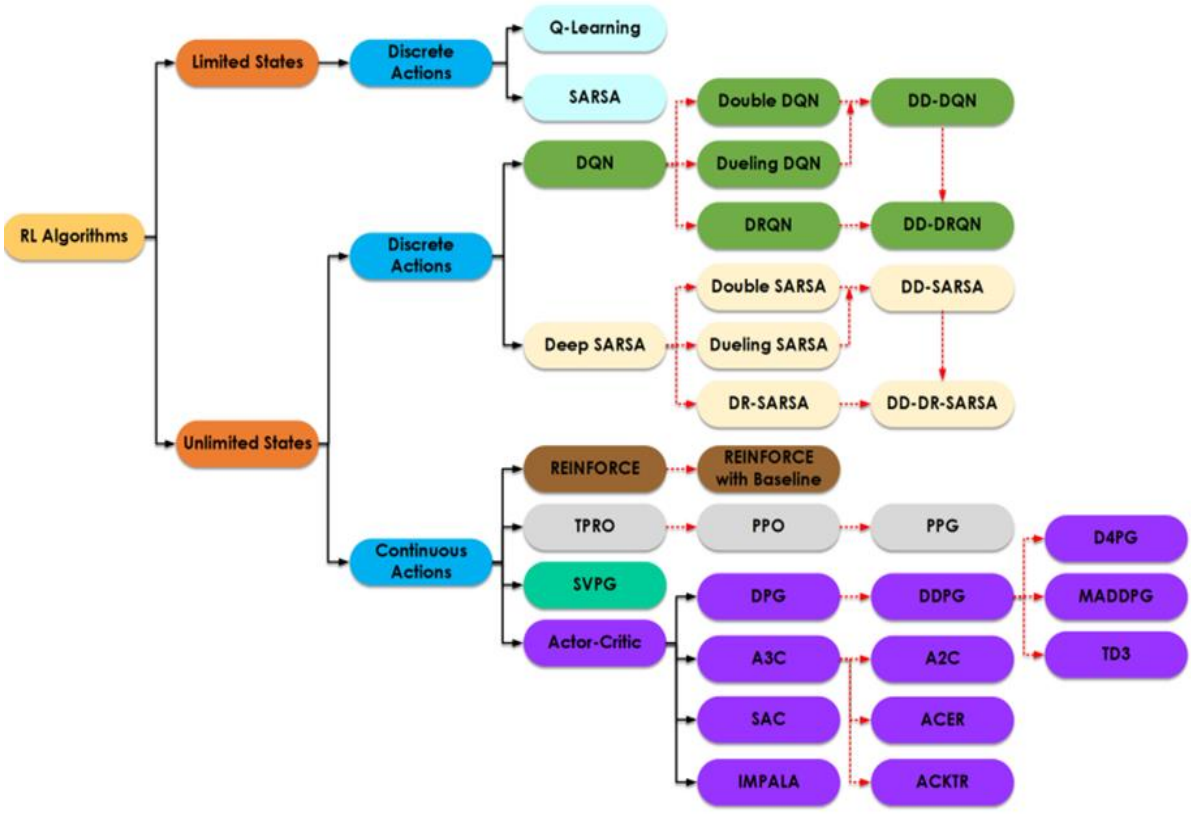



FIGURE 2.7: Reinforcement learning algorithms classification.

There are a lot of different RL algorithms have been proposed in the literature [57]. A general picture regarding RL algorithms is provided in Fig. 2.7. For example, one of typical RL algorithms is Q-learning that belongs to the class of model-free RL method and is particularly suited for problems with discrete state-action spaces [56]. The essence of Q-learning lies in estimating a value function known as the Q-function, which indicates the expected cumulative rewards that an agent can achieve by taking a specific action in a given state and following an optimal policy. Using Q-learning, the agent constructs a dedicated Q-table for the purpose of storing Q-values corresponding to every conceivable state-action pairing. Through an iterative process, Q-learning updates these Q-values based on the agent's experiences in the environment, gradually improving its policy to maximize long-term rewards. In addition, one of popular DRL algorithms is Deep Q-network (DQN) that

combines the principles of Q-learning with the capacity of DNN to handle complex and high-dimensional state spaces. The need for DQN arises from the limitations of traditional Q-learning, which struggles when dealing with environments that have large state spaces and require extensive exploration to learn optimal policies. By using DNN architecture, DQN can approximate the Q-values of different state-action pairs, allowing it to tackle challenging tasks in real-world systems [51].

## 2.4 Non-Orthogonal Multiple Access

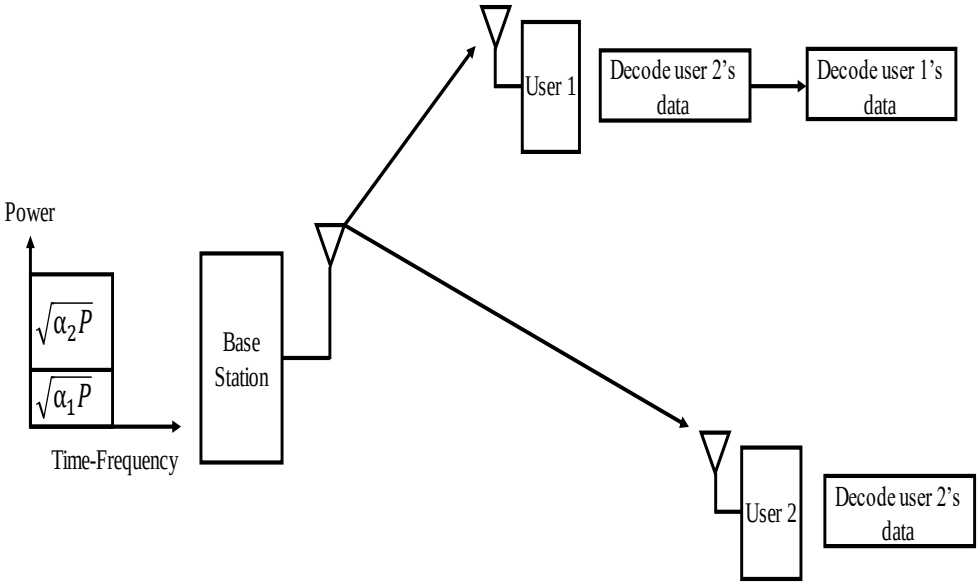



FIGURE 2.8: Illustration of two-user power-domain NOMA

NOMA has emerged as a potential next generation multiple access candidate for beyond 5G wireless networks [58]. In NOMA, multiple users share the same time and frequency resources simultaneously, resulting in a non-orthogonal overlap of signals. This approach enables higher spectral efficiency and better utilization of available resources. NOMA is classified into two primary categories: power-domain NOMA and code-domain NOMA. Specifically, power-domain NOMA assigns different power levels to users sharing the same resource block (time/frequency), while code-domain NOMA employs unique spreading codes to separate users' signals. Considering a typical downlink power-domain NOMA system, which consists of two users including one near user $U_1$ and one far user $U_2$, as depicted in Fig. 2.8. Following downlink NOMA principle, BS first combines the

transmitted messages of two users as a superposition message and then transmits it to two users. In this regard, the superposition message can be formulated as [59]

$$x = \sqrt{P\alpha_1}x_1 + \sqrt{P\alpha_2}x_2, \tag{2.6}$$

where $P$ is the transmission power, $\alpha_i$ ($\alpha_1 + \alpha_2 = 1$, $1 \leq i \leq 2$) and $x_i$ denote the power allocation coefficient and message according to user $U_i$, respectively. Thus, the received signal at user $U_i$ ($1 \leq i \leq 2$) can be represented as [59]

$$y_i = h_i x + w_i, \tag{2.7}$$

where $h_i$ is the channel coefficient of the link from the BS to user $U_i$ and $w_i \sim \mathcal{CN}(0, N_0)$ denotes the additive white Gaussian noise (AWGN) at user $U_i$. Since user $U_1$ is a near user to the BS, hence it usually has channel conditions better than far user $U_2$. Given this context, it is assumed that user $U_1$ is allocated with lower power than user $U_2$, i.e., $\alpha_1 \leq \alpha_2$. Therefore, user $U_2$ can directly decode its message $x_2$ by considering the message from user $U_1$ as interfering noise. Consequently, the instantaneous signal-to-interference-plus-noise ratio (SINR) at user $U_2$ for the detection of $x_2$ is given by

$$\gamma_2 = \frac{\alpha_2\gamma_0 \left|h_2\right|^2}{\alpha_1\gamma_0 \left|h_2\right|^2 + 1}, \tag{2.8}$$

where $\gamma_0 = \frac{P}{N_0}$ denotes the average transmit signal-to-noise ratio (SNR). Conversely, user $U_1$ first needs to decode $x_2$ due to $\alpha_1 \leq \alpha_2$ and then removes this component from its received signal by applying successive interference cancellation (SIC) technique [60]. After that, it can detect its own message $x_1$. Based on the above discussion, the instantaneous SINRs at user $U_1$ for the detection of $x_2$ and $x_1$ are respectively written by

$$\gamma_{12} = \frac{\alpha_2\gamma_0 \left|h_1\right|^2}{\alpha_1\gamma_0 \left|h_1\right|^2 + 1}, \tag{2.9}$$

and

$$\gamma_{11} = \alpha_1\gamma_0 \left|h_1\right|^2. \tag{2.10}$$

In addition, NOMA standardization for both uplink and downlink transmissions has been also studied in the third Generation Partnership Project (3GPP) frameworks in recent years [61–63]. Its recent developments, such as grant-free NOMA [31] and rate-splitting-based NOMA [40], is opening the door to make it a promising candidate for addressing the diverse needs of future wireless networks, especially in massive access scenario [58].

Chapter 3

# Short-Packet Communications in URLLC-Enabled Systems: BLER and Minimum Blocklength Analysis

Recently, ultra-reliable and low-latency communications (URLLC) using short-packets has been proposed to fulfill the stringent requirements regarding reliability and latency of emerging applications in 5G and beyond networks. In addition, multiple-input multiple-output non-orthogonal multiple access (MIMO NOMA) is a potential candidate to improve the spectral efficiency, reliability, latency, and connectivity of wireless systems. In this chapter, we investigate short-packet communications (SPC) in a multi-user downlink MIMO NOMA system over Nakagami-$m$ fading, and propose two antenna-user selection methods considering two clusters of users having different priority levels. In contrast to the widely-used long data-packet assumption, the SPC analysis requires the redesign of the communication protocols and novel performance metrics. Given this context, we analyze the SPC performance of MIMO NOMA systems using the average block error rate (BLER) and minimum blocklength, instead of the conventional metrics such as ergodic capacity and outage capacity. More specifically, to characterize the system performance regarding SPC, asymptotic (in the high signal-to-noise ratio regime) and approximate closed-form expressions of the average BLER at the users are derived. Based on the asymptotic behavior of the average BLER, an analysis of the diversity order, minimum blocklength, and optimal power allocation is carried out. The achieved results show that MIMO NOMA can serve multiple users simultaneously using a smaller blocklength compared with MIMO OMA, thus demonstrating the benefits of MIMO NOMA for SPC in minimizing the transmission latency. Furthermore, our results indicate that the proposed methods not only improve the BLER performance, but also guarantee full diversity gains for the respective users.

The chapter is organized as follows. Introduction to the current state of the art is discussed in Section 3.1. Section 3.2 introduces the system model. The proposed approach for BLER performance analysis with SPC is presented in Section 3.3. Section 3.4 shows the the proposed analytical framework for optimal power allocation and minimum blocklength. Section 3.5 describes the simulation results. Finally, Section 3.6 provides the summary and concluding remarks of this chapter.

## 3.1 Introduction

URLLC has recently been considered as a promising technology for the 5th generation (5G) and beyond wireless networks to support novel applications with unprecedented requirements of reliability and latency [6, 7, 64]. Furthermore, it is a potential solution for mission-critical Internet of Things (IoT) applications such as industrial automation, remote surgery, and vehicle-to-everything (V2X) communications, which require high reliability and low latency [65, 66]. URLLC systems should be designed to meet the requirements of high reliability (99.999%) and low latency (1 ms) [8]. To achieve such stringent requirements, a new transmission approach, i.e., SPC, could be a promising solution. This is different from the traditional analytic methods designed to target Shannon's channel capacity using long data-packets, which are no longer suitable for low latency systems [8]. To characterize the performance of SPC, new performance metrics including BLER and overhead ratio (i.e., ratio of pilots to the information payload), have been introduced in the literature [9, 10, 67].

Besides, NOMA has recently emerged as a promising technology to improve the spectral efficiency and user fairness for wireless networks [68, 69]. In contrast to the orthogonal multiple access (OMA) which utilizes orthogonal resources (e.g., time and frequency) to support multiple users, this technique can serve them at the same time/frequency/code by using the power domain and effective interference management methods, such as SIC [68]. Therefore, NOMA can more effectively support massive connectivity and further improve the reliability and latency for wireless systems [70, 71]. With its potential advantages, NOMA standardization has been recently studied in 3GPP frameworks [61–63] including the 3GPP Release 16 [63]. Also, the latest trend is to employ NOMA in the uplink due to the emergence of IoT and machine-type communication systems [63, 64, 72]. Thus, NOMA and its variations are expected to be employed in various 5G and beyond application scenarios [71, 73, 74].

In addition, the combination of NOMA and multiple-input multiple-output (MIMO) systems (so-called MIMO NOMA), which can significantly enhance the spectral efficiency and performance of NOMA systems, has also been investigated in recent years [75, 76]. The ergodic capacity analysis of MIMO NOMA systems has been considered in [77], where

the authors have proved the superiority of MIMO NOMA over MIMO OMA in terms of capacity. To exploit the spatial degrees of freedom, some MIMO NOMA schemes have been proposed in the literature [78, 79]. Specifically, the authors in [23] have considered a multi-beam MIMO NOMA scenario, where multiple analog beams are formed for downlink transmission of a NOMA user group by exploiting the channel sparsity and a large scale antenna array. Meanwhile, the work in [24] has investigated a space-time coded MIMO NOMA system, where two users' signals are mapped into $n$-dimensional constellations corresponding to the same algebraic lattices from a number field. Although the system performance can be significantly improved with the increase in the utilized number of antennas in MIMO systems, this requires large power consumption and high complexity of signal processing [80]. To mitigate these issues while ensuring the diversity and capacity benefits from MIMO, transmit antenna selection (TAS) scheme has been proposed as a promising solution to improve the performance gain of MIMO NOMA systems [60, 80, 81]. It is noteworthy that the above works on MIMO NOMA have been conducted under the assumption of long data-packet transmissions, which is no longer applicable for emerging URLLC applications with short data-packets in 5G and beyond networks [2, 3, 64].

To overcome this challenge, in this chapter, we propose to utilize SPC for MIMO NOMA systems to improve the reliability and latency as well as enhance the spectral efficiency and connectivity for wireless systems. As stated earlier, the large power consumption and high computational complexity of MIMO systems are putting a crucial challenge in designing effective communication protocols for SPC-based MIMO NOMA systems. Therefore, we consider a scenario, where TAS is used at the transmitter, and selection combining (SC) and maximal ratio combining (MRC) are utilized at the users with the purpose of improving the performance and reducing the complexity for MIMO NOMA systems with SPC. Herein, suitable performance metrics for SPC including average BLER and minimum blocklength, are utilized instead of the conventional ones such as ergodic capacity and outage capacity.

Recently, there have been a few works on SPC in NOMA systems, which is considered as a promising solution to enhance the reliability, latency, and connectivity for wireless networks [11–16, 19, 20]. In particular, in [11], a two-user NOMA system with short-packets over Rayleigh fading channels was considered, in which the average BLER at users is derived to evaluate the system performance. In [12], the BLER performance of a NOMA system was addressed, where stochastic geometry and Nakagami-$m$ fading channels are considered. In [13], X. Lai *et al.* analyzed the performance of a cooperative NOMA SPC system over Rayleigh fading channels. Furthermore, the transmission energy minimization problem and packet scheduling for two-user downlink NOMA systems with strictly heterogeneous latency constraints were investigated in [19, 20]. However, the works [11–13, 19, 20] only considered single-input single-output (SISO) systems.

To exploit the benefits of multiple antennas in improving the reliability and reducing the latency for SPC in NOMA systems, the work in [14] investigated a multiple-input single-output (MISO) scheme to evaluate the outage performance of a URLLC NOMA system with wireless power transfer. In [15], MIMO NOMA for URLLC systems was considered to enhance the reliability and latency performance of the system. In this regard, a closed-form upper bound for the delay target violation probability was derived in [15] to identify the sufficient and necessary condition for the optimal transmit power. However, the analysis of average BLER and minimum blocklength was not considered in [15]. The works in both [14] and [15] investigated a scenario where an $N$-antenna base station (BS) provides services to $N$ pairs of NOMA users, in which each pair of users is served by a distinct single transmit antenna. In contrast to this scenario, in [16], the combination of transmit antennas to serve a pair of users was examined in order to enhance the BLER performance of short-packet NOMA systems by utilizing the maximum ratio transmission (MRT), in which only the MISO scenario was considered.

Although MRT can significantly improve the system performance by combining all transmit antennas for transmission, it leads to high complexity of the signal processing and feedback overhead [82]. Against this context, TAS has been proposed as a low-complexity and power-efficient solution for multi-antenna transmitters to enhance the performance of NOMA systems by selecting a best transmit antenna for transmission that maximizes the signal-to-noise ratio (SNR) at the receiver side [60, 80, 81]. Nevertheless, the short-packet transmission in MIMO NOMA systems considering the TAS solution, average BLER, and minimum blocklength has not yet been analyzed. Furthermore, it is noted that most of these existing studies [11–16] only investigated Rayleigh fading channels. Research on SPC for MIMO NOMA systems applying TAS for the transmitter, selection combining (SC) and maximal ratio combining (MRC) for the receiver, over a generic fading channel, i.e., Nakagami-$m$, to improve the system performance more effectively and bring more general insights of the system behavior has not yet been conducted, and thus is the focus of this chapter.

In contrast to the existing related works, in this chapter, we propose a new framework to analyze the system performance of utilizing SPC in a NOMA network, in which MIMO and Nakagami-$m$ distribution are considered. Most existing works on NOMA are conducted under the assumption that NOMA is carried out based on the difference in users' channel conditions [11–16,19,20,60,68–70,74–77,81]. More precisely, in a two-user downlink NOMA system, a BS transmits information to the users by superimposing users' messages with different transmit power levels [68]. The user having worse channel quality is allocated with the higher power level compared with the user having a better channel condition. However, in practice, users may have similar channel conditions but require different quality of service (QoS) as discussed in [83–85]. For example, some users may need to be served faster with

low targeted data-rate, i.e, incident alerts, while some users can be served with the best effort, i.e., downloading of multimedia files [84]. In such a heterogeneous scenario, NOMA scheme becomes advantageous as compared to the conventional OMA as it can concurrently serve users having different QoS priorities with the same resources (time/frequency/code).

Given this context, we examine a scenario, in which a BS communicates with two user clusters having different priority levels over Nakagami-$m$ fading channels, where the BS and all users are equipped with multiple antennas. Note that Nakagami-$m$ with parameter $m$ is described as a general distribution that can include the well-known Rayleigh distribution and approximate the Rician one with the parameter $K$, where $m = (K+1)^2 / (2K+1)$ [86]. In contrast to [12], which considers Nakagami-$m$ fading channels for the BLER derivation in SISO case, our analysis derives the BLER expression for a more general scenario, i.e., MIMO. Herein, different MIMO schemes are investigated to reduce the complexity of signal processing and exploit the benefits of multiple antennas in improving the system performance. Particularly, at the BS, TAS is utilized to select the best transmit antenna for transmission that maximizes the post-processed SNR at the receiver [82]. Besides, at the user-side, two different diversity techniques are investigated: 1) SC, which selects the best received signal branch for further processing; and 2) MRC, which combines all the received signal branches from receive antennas to maximize the output SNR.

In addition, as discussed in [60, 78, 79, 81], assigning all users in a system for the implementation of NOMA is difficult due to the strong co-channel interference, leading to large complexity and high decoding delay. To overcome this issue, hybrid NOMA has recently been considered as a promising solution for 5G and beyond networks [87, 88]. Particularly, in this solution, all users in a network are divided into multiple small groups. Herein, the users in each group are served by NOMA, whereas the different groups are assigned to different orthogonal resource blocks (e.g., time or frequency). Therefore, in this chapter, we consider a scenario, where users are paired[1] to perform NOMA with the purpose of decreasing the strong co-channel interference in NOMA systems [59, 89, 90]. This is a common assumption widely adopted in the NOMA literature to reduce the computational complexity and time delay of SIC decoding [60, 78, 79, 81]. It is noted that the achieved results from this analysis can be straightforwardly applied to different groups, which are incorporated into the network in an orthogonal manner.

Therefore, the main contributions of this chapter are summarized as follows:

1. Firstly, we propose a novel framework to analyze the performance of an SPC-based NOMA system, where MIMO transmission and Nakagami-$m$ fading are considered.

[1]It is noted that the proposed schemes can be applied to the general scenario with more than two users within a NOMA group, which, however, results in higher computational complexity and larger time delay of SIC decoding, and is thus left for future work.

To achieve a general insight into the system behavior, we investigate two different cases of applying MIMO schemes for the transmitter and receiver sides including TAS/SC and TAS/MRC. Moreover, we investigate two antenna-user selection methods, namely high-priority cluster selection (HCS) and low-priority cluster selection (LCS), to design the effective communication protocols for SPC in a MIMO NOMA system.

2. Secondly, we derive closed-form expressions for the average BLER of users in all considered cases. It should be noted that this work analyzes the performance in terms of average BLER, which is more suitable for SPC than widely-used performance metrics such as ergodic capacity and outage capacity [8, 9].

3. Thirdly, we derive asymptotic expressions for the average BLER in the high SNR regime and carry out an analysis of diversity order, minimum blocklength and optimal power allocation for SPC-based MIMO NOMA system based on the asymptotic average BLER.

4. Finally, we perform the blocklength comparison between MIMO NOMA and MIMO OMA systems to clarify the superiority of MIMO NOMA compared to MIMO OMA in terms of low-latency transmission when considering SPC.

## 3.2 System Model

In this chapter, the SPC in a multiuser downlink MIMO NOMA system over Nakagami-$m$ fading channels is considered, as depicted in Fig. 3.1. The network consists of one base station (BS), denoted by $S$, two cluster of users, denoted by $H = \{H_1, \ldots, H_I\}$ and $L = \{L_1, \ldots, L_J\}$. In addition, the BS and the users in both clusters $H$ and $L$ are equipped with $K_S$, $K_H$, and $K_L$ antennas, respectively. As reported earlier in Section 3.1, it is assumed that the users' QoS requirements are taken into account in the design of the MIMO NOMA transmission in SPC instead of their channel conditions. More precisely, we consider the scenario where the users in clusters $H$ and $L$ are treated as high-priority and low-priority ones, respectively. Furthermore, the users are paired to perform NOMA with the purpose of decreasing the strong co-channel interference in NOMA systems [59, 89, 90]. Specifically, each user pair consists of two users having different priorities selected from both the clusters H and L. Moreover, as mentioned earlier in Section 3.1, to exploit the benefits of multiple antennas, we consider the scenario where TAS is employed at BS $S$ whereas SC or MRC is utilized at the users' side (i.e., TAS/SC or TAS/MRC).

Regarding channel model, let $h_{S_k H_{i,r}}$ $\left(h_{S_k L_{j,s}}\right)$ ($1 \leq k \leq K_S$, $1 \leq i \leq I$, $1 \leq j \leq J$, $1 \leq r \leq K_H$, $1 \leq s \leq K_L$) denote the channel coefficient of the link from antenna $k$ at BS $S$ to antenna $r$ ($s$) at the user $H_i$ ($L_j$). Herein, $h_{S_k H_{i,r}}$ $\left(h_{S_k L_{j,s}}\right)$ is an independent identically

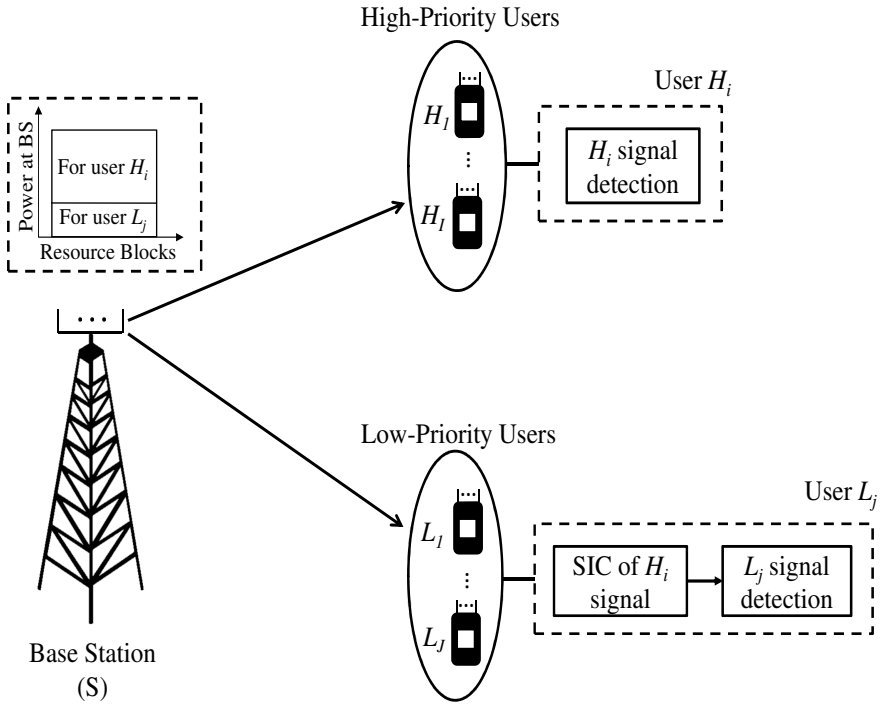



FIGURE 3.1: Model of a MIMO NOMA system under SPC over Nakagami-$m$ fading.

distributed (i.i.d) random variable following Nakagami-$m$ distribution with parameter $m_H$ ($m_L$) and mean value $\Omega_H = \mathbb{E}\left[\left|h_{S_k H_{i,r}}\right|^2\right]$ $\left(\Omega_L = \mathbb{E}\left[\left|h_{S_k L_{j,s}}\right|^2\right]\right)$. Thus, the Nakagami-$m$ distributions of $h_{S_k H_{i,r}}$ and $h_{S_k L_{j,s}}$ are, respectively, given by [86]

$$f_{h_{S_k H_{i,r}}}(x) = \frac{2m_H^{m_H} x^{2m_H-1}}{\Gamma(m_H)\,\Omega_H} e^{-\frac{m_H x^2}{\Omega_H}}, \tag{3.1}$$

and

$$f_{h_{S_k L_{j,s}}}(x) = \frac{2m_L^{m_L} x^{2m_L-1}}{\Gamma(m_L)\,\Omega_L} e^{-\frac{m_L x^2}{\Omega_L}}. \tag{3.2}$$

### 3.2.1 Antenna and User Selection

In this subsection, we present the proposed solutions of selecting antennas and users. As stated earlier, the user pairing is utilized for designing the MIMO NOMA. Specifically, the best user in each cluster is selected to perform NOMA based on the channel gains of the link

from BS S to the users in order to improve the performance of NOMA implementation[2] [91]. Furthermore, we investigate two different antenna-user selection methods, i.e., HCS and LCS, which aim to improve the performance for the users selected from clusters $H$ and $L$, respectively. It is noted that this selection process can be carried out prior to information transmission through a suitable signaling and channel state information (CSI) estimation method [60]. In addition, as in [60, 75, 81], the perfect CSI scenario is considered and the required partial CSI, i.e., the instantaneous channel power gains, for each method is assumed to be available at the BS.

### HCS Method

Since the users in cluster $H$ has higher priority than those in cluster $L$, this method focuses on improving the performance of the selected user in cluster $H$. In particular, HCS method aims to jointly select a transmit antenna and a user in cluster $H$ to maximize the channel power gain of the link from the BS $S$ to the selected user.

For the TAS/SC scheme, the indices of selected transmit antenna, $\hat{k}$, user and receive antenna selected from cluster $H$, $\hat{i}$ and $\hat{r}_H$, are given by [82, 85]

$$\left(\hat{k}, \hat{i}, \hat{r}_H\right) = \arg \max_{1\leq k\leq K_S, 1\leq i\leq I, 1\leq r\leq K_H} \left\{\left|h_{S_k H_{i,r}}\right|^2\right\}, \tag{3.3}$$

and the indices of user and receive antenna selected from cluster $L$, $\hat{j}$ and $\hat{r}_L$, are expressed as

$$\left(\hat{j}, \hat{r}_L\right) = \arg \max_{1\leq j\leq J, 1\leq r\leq K_L} \left\{\left|h_{S_{\hat{k}} L_{j,r}}\right|^2\right\}. \tag{3.4}$$

For TAS/MRC, $\hat{k}$, $\hat{i}$, and $\hat{j}$ are given by [82]

$$\left(\hat{k}, \hat{i}\right) = \arg \max_{1\leq k\leq N_S, 1\leq i\leq I} \left\{\|\mathbf{h}_{S_k H_i}\|^2\right\}, \tag{3.5}$$

and

$$\hat{j} = \arg \max_{1\leq j\leq J} \left\{\left\|\mathbf{h}_{S_{\hat{k}} L_j}\right\|^2\right\}, \tag{3.6}$$

where $\mathbf{h}_{S_k H_i}$ $\left(\mathbf{h}_{S_k L_j}\right)$ represents the $K_H \times 1$ ($K_L \times 1$) channel vector of the link from antenna $k$ at BS $S$ to user $H_i$ ($L_j$).

[2] In fact, some other sophisticated user pairing methods may further improve the performance of SPC-based MIMO NOMA systems. However, it is beyond the scope of this chapter.

**LCS Method**

To improve the performance of the selected user in cluster $L$ which has a lower priority, an antenna at BS $S$ and a user in cluster $L$ are jointly chosen for transmission to provide the best channel power gain of the link from BS $S$ to the selected user. Mathematically, $\hat{k}$, $\hat{i}$, $\hat{j}$, $\hat{r}_H$, and $\hat{r}_L$ in this method can be expressed as follows:

For TAS/SC:

$$\begin{cases} \left(\hat{k},\hat{j},\hat{r}_L\right) = \arg \max\limits_{1\le k\le K_S, 1\le j\le J, 1\le r\le K_L} \left\{\left|h_{S_k L_{j,r}}\right|^2\right\}, \\ \left(\hat{i},\hat{r}_H\right) = \arg \max\limits_{1\le i\le I, 1\le r\le K_H} \left\{\left|h_{S_{\hat{k}} H_{i,r}}\right|^2\right\}, \end{cases} \tag{3.7}$$

and for TAS/MRC:

$$\begin{cases} \left(\hat{k},\hat{j}\right) = \arg \max\limits_{1\le k\le N_S, 1\le j\le J} \left\{\left\|\mathbf{h}_{S_k L_j}\right\|^2\right\}, \\ \hat{i} = \arg \max\limits_{1\le i\le I} \left\{\left\|\mathbf{h}_{S_{\hat{k}} H_i}\right\|^2\right\}, \end{cases} . \tag{3.8}$$

### 3.2.2 Information Transmission Process and Channel Statistics

With the NOMA protocol, BS $S$ transmits the mixed message [59]

$$x = \sqrt{P_S \alpha_{H_{\hat{i}}}} x_{H_{\hat{i}}} + \sqrt{P_S \alpha_{L_{\hat{j}}}} x_{L_{\hat{j}}} \tag{3.9}$$

to users $H_{\hat{i}}$ and $L_{\hat{j}}$. Herein, $P_S$ is the total transmit power, $\alpha_{H_{\hat{i}}}$ and $\alpha_{L_{\hat{j}}}$ ($\alpha_{H_{\hat{i}}} + \alpha_{L_{\hat{j}}} = 1$) denote the power allocation coefficients, as well as $x_{H_{\hat{i}}}$ and $x_{L_{\hat{j}}}$ represent the messages for users $H_{\hat{i}}$ and $L_{\hat{j}}$, respectively. It is noted that $\alpha_{H_{\hat{i}}} > \alpha_{L_{\hat{j}}} > 0$ due to higher priority of user $H_{\hat{i}}$. Thus, the received signal at user $U$ $\left(U \in \left\{H_{\hat{i}}, L_{\hat{j}}\right\}\right)$ is given by

$$y_U = \mathbf{u}_U \mathbf{h}_{S_{\hat{k}} U} \sqrt{P_S} \left(\sqrt{\alpha_{H_{\hat{i}}}} x_{H_{\hat{i}}} + \sqrt{\alpha_{L_{\hat{j}}}} x_{L_{\hat{j}}}\right) + \mathbf{u}_U \mathbf{w}_U, \tag{3.10}$$

where $\mathbf{w}_U \sim \mathcal{CN}(0, N_0)$ denotes the additive white Gaussian noise (AWGN) at user $U$, and $\mathbf{u}_U$ represents the signal processing operation at user $U$, which is defined as in [92]

$$\mathbf{u}_U = \begin{cases} \mathbf{e}_{K_U, \hat{r}_U}, & \text{for TAS/SC} \\ \dfrac{\mathbf{h}^{\dagger}_{S_{\hat{k}} U}}{\left\|\mathbf{h}_{S_{\hat{k}} U}\right\|}, & \text{for TAS/MRC} \end{cases} , \tag{3.11}$$

where $\mathbf{e}_{K,i}$ is a $1 \times K$ vector whose the $i$-th element is equal to 1, and the others are zeros.

According to NOMA principle, the user $H_{\hat{i}}$ can directly decode its own message, $x_{H_{\hat{i}}}$, since it is allocated with larger transmit power (i.e., $\alpha_{H_{\hat{i}}} > \alpha_{L_{\hat{j}}}$), hence, the interference generated by the signal of the user $L_{\hat{j}}$, $x_{L_{\hat{j}}}$, can be treated as noise [60]. Thus, the instantaneous signal-to-interference-plus-noise ratio (SINR) at the user $H_{\hat{i}}$ to detect $x_{H_{\hat{i}}}$ is written as

$$\gamma_{H_{\hat{i}}}^{x_{H_{\hat{i}}}} = \frac{\alpha_{H_{\hat{i}}}\gamma_0 g_{SH}}{\alpha_{L_{\hat{j}}}\gamma_0 g_{SH} + 1}, \tag{3.12}$$

where $\gamma_0 = \frac{P_S}{N_0}$ denotes the average transmit SNR and $g_{SH}$ is defined as

$$g_{SH} = \begin{cases} \left|h_{S_{\hat{k}}H_{\hat{i},\hat{r}_H}}\right|^2, & \text{for TAS/SC} \\ \left\|\mathbf{h}_{S_{\hat{k}}H_{\hat{i}}}\right\|^2, & \text{for TAS/MRC} \end{cases}. \tag{3.13}$$

Meanwhile, since the user $H_{\hat{i}}$ is served with higher priority than the user $L_{\hat{j}}$ (i.e., $\alpha_{H_{\hat{i}}} > \alpha_{L_{\hat{j}}}$), the user $L_{\hat{j}}$ first needs to decode $x_{H_{\hat{i}}}$ and then remove this component from the received signal by using SIC before detecting its own message, $x_{L_{\hat{j}}}$, [60]. Unlike [11–13, 60] considering the perfect SIC (PSIC), in this chapter, we consider the imperfect SIC (ISIC) scenario[3] to achieve more practical insights, where there exists a residual interference component due to the ISIC process [93]. Thus, the instantaneous SINRs at the user $L_{\hat{j}}$ to detect $x_{H_{\hat{i}}}$ and $x_{L_{\hat{j}}}$ are respectively expressed as

$$\gamma_{L_{\hat{j}}}^{x_{H_{\hat{i}}}} = \frac{\alpha_{H_{\hat{i}}}\gamma_0 g_{SL}}{\alpha_{L_{\hat{j}}}\gamma_0 g_{SL} + 1}, \tag{3.14}$$

and

$$\gamma_{L_{\hat{j}}}^{x_{L_{\hat{j}}}} = \frac{\alpha_{L_{\hat{j}}}\gamma_0 g_{SL}}{\psi\alpha_{H_{\hat{i}}}\gamma_0 g_{SL} + 1}, \tag{3.15}$$

where $\psi = \mathbb{E}\left[\left|x_{H_{\hat{i}}} - \hat{x}_{H_{\hat{i}}}\right|^2\right]$ denotes the level of residual interference caused by the ISIC process at user $L_{\hat{j}}$, which indicates the difference between the actual signal $x_{H_{\hat{i}}}$ and the estimated signal $\hat{x}_{H_{\hat{i}}}$. Specifically, $\psi = 0$ means perfect SIC and $0 < \psi \leq 1$ denotes ISIC. In (3.14) and (3.15), $g_{SL}$ is given by

$$g_{SL} = \begin{cases} \left|h_{S_{\hat{k}}L_{\hat{j},\hat{r}_L}}\right|^2, & \text{for TAS/SC} \\ \left\|\mathbf{h}_{S_{\hat{k}}L_{\hat{j}}}\right\|^2, & \text{for TAS/MRC} \end{cases}. \tag{3.16}$$

[3]It is noteworthy to mention that in this chapter, we consider perfect CSI to evaluate the effects of ISIC on the performance of SPC-based MIMO NOMA systems. However, analyzing the impact of imperfect CSI on the performance of the SPC systems is also an important problem to be investigated in future works.

## 3.3 Proposed approach for BLER performance analysis with SPC

In this section, we present some preliminaries on SPC and average BLER calculation, the derivation of CDF of channel power gains, and the average BLER analysis by utilizing HCS and LCS methods with TAS/SC and TAS/MRC schemes, specified in Section 3.2.1.

### 3.3.1 Preliminaries

Considering SPC with blocklength $N$, the Shannon capacity $C$, and the BLER $\varepsilon$, the maximum achievable rate can be expressed as [11]:

$$R = C - \sqrt{\frac{v}{N}} Q^{-1}(\varepsilon) + O\left(\frac{\log_2 N}{N}\right), \tag{3.17}$$

where $Q^{-1}(x)$ is the inverse of the Gaussian Q-function, $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$, $C = log_2(1+\gamma)$, $v = (\log_2 e)^2 \left[1 - \frac{1}{(1+\gamma)^2}\right]$ represents the channel dispersion, $\gamma$ is the SNR or SINR, and $O\left(\frac{\log_2 N}{N}\right)$ is the remainder terms of order $\frac{\log_2 N}{N}$. From (3.17), an approximation method, which is commonly referred to as normal approximation [9], is utilized to compute the instantaneous BLER as follows:

$$\varepsilon \approx Q\left(\frac{C - n/N}{\sqrt{v/N}}\right), \tag{3.18}$$

where $R = n/N$, $n$ denotes the number of information bits, and the approximation is achieved by omitting the term $O\left(\frac{\log_2 N}{N}\right)$ when $N \geq 100$ as in [9].

Based on (3.18), the instantaneous BLER of decoding the message of user $V$, $x_V$ $\left(V \in \left\{H_{\hat{i}}, L_{\hat{j}}\right\}\right)$, at user $U$ $\left(U \in \left\{H_{\hat{i}}, L_{\hat{j}}\right\}\right)$ is given by [11]:

$$\varepsilon_U^{x_V} \approx Q\left(\frac{\log_2(1+\gamma_U^{x_V}) - n_V/N_V}{\sqrt{v_U^{x_V}/N_V}}\right), \tag{3.19}$$

where $v_U^{x_V} = (\log_2 e)^2 \left[1 - \frac{1}{\left(1+\gamma_U^{x_V}\right)^2}\right]$, $n_V$ and $N_V$ denote the number of information bits and blocklength to user $V$, respectively. Thus, the instantaneous BLER is calculated through the received SINR, the Shannon capacity, the number of information bits, and the blocklength; and specific encoding and modulation methods are not considered. From

(3.19), the average BLER $\bar{\varepsilon}_U^{x_V}$ has the following form

$$\bar{\varepsilon}_U^{x_V} \approx \int_0^{\infty} \varepsilon_U^{x_V} f_{\gamma_U^{x_V}}(x)\, dx, \tag{3.20}$$

where $f_X(x)$ is the probability density function (PDF) of a random variable $X$. It is challenging to derive $\bar{\varepsilon}_U^{x_V}$ in (3.20). Therefore, an approximation[4] of $\varepsilon_U^{x_V}$ is utilized as discussed in [47], i.e.,

$$\varepsilon_U^{x_V} \approx \begin{cases} 1, & \gamma_U^{x_V} \leq v_V \\ A_U^{x_V}, & v_V < \gamma_U^{x_V} < \mu_V \\ 0, & \gamma_U^{x_V} \geq \mu_V \end{cases}, \tag{3.21}$$

where $A_U^{x_V} = 0.5 - \chi_V \sqrt{N_V}\,(\gamma_U^{x_V} - \beta_V)$, $\chi_V = \sqrt{\frac{1}{2\pi\left(2^{\frac{2n_V}{N_V}} - 1\right)}}$, $v_V = \beta_V - \frac{1}{2\chi_V\sqrt{N_V}}$, $\mu_V = \beta_V + \frac{1}{2\chi_V\sqrt{N_V}}$, and $\beta_V = 2^{\frac{n_V}{N_V}} - 1$. By substituting (3.21) into (3.20), $\bar{\varepsilon}_U^{x_V}$ can be given by

$$\bar{\varepsilon}_U^{x_V} \approx \chi_V \sqrt{N_V} \int_{v_V}^{\mu_V} F_{\gamma_U^{x_V}}(x)\, dx. \tag{3.22}$$

For user $H_{\hat{i}}$, from (3.12) and (3.22), its average BLER is expressed as

$$\begin{aligned} \bar{\varepsilon}_{H_{\hat{i}}} &= \bar{\varepsilon}_{H_{\hat{i}}}^{x_{H_{\hat{i}}}} \\ &\approx \chi_{H_{\hat{i}}} \sqrt{N_{H_{\hat{i}}}} \int_{v_{H_{\hat{i}}}}^{\mu_{H_{\hat{i}}}} F_{\gamma_{H_{\hat{i}}}^{x_{H_{\hat{i}}}}}(x)\, dx. \end{aligned} \tag{3.23}$$

For user $L_{\hat{j}}$, it first needs to remove the message of user $H_{\hat{i}}$, i.e., $x_{H_{\hat{i}}}$, by using ISIC before detecting its own message, i.e., $x_{L_{\hat{j}}}$. Therefore, user $L_{\hat{j}}$ cannot decode $x_{L_{\hat{j}}}$ if it decodes $x_{H_{\hat{i}}}$ unsuccessfully. This will affect its BLER performance. Given this context, the average BLER at user $L_{\hat{j}}$ is given by

$$\bar{\varepsilon}_{L_{\hat{j}}} = \bar{\varepsilon}_{L_{\hat{j}}}^{x_{H_{\hat{i}}}} + \left(1 - \bar{\varepsilon}_{L_{\hat{j}}}^{x_{H_{\hat{i}}}}\right) \bar{\varepsilon}_{L_{\hat{j}}}^{x_{L_{\hat{j}}}}, \tag{3.24}$$

[4]In this chapter, we consider an approximation method for BLER, as discussed in [47], to analyze the performance of SPC-based MIMO NOMA systems in terms of BLER. Deriving the error bound for BLER based on Jensen's inequality [94], which is more challenging, could be an interesting problem to investigate in future work.

where

$$\bar{\varepsilon}_{L_{\hat{j}}}^{x_{H_{\hat{i}}}} \approx \chi_{H_{\hat{i}}} \sqrt{N_{H_{\hat{i}}}} \int_{v_{H_{\hat{i}}}}^{\mu_{H_{\hat{i}}}} F_{\gamma_{L_{\hat{j}}}}^{x_{H_{\hat{i}}}}(x)\, dx,$$

and

$$\bar{\varepsilon}_{L_{\hat{j}}}^{x_{L_{\hat{j}}}} \approx \chi_{L_{\hat{j}}} \sqrt{N_{L_{\hat{j}}}} \int_{v_{L_{\hat{j}}}}^{\mu_{L_{\hat{j}}}} F_{\gamma_{L_{\hat{j}}}}^{x_{L_{\hat{j}}}}(x)\, dx.$$

### 3.3.2 Derivation for Cumulative Distribution Function (CDF) of Channel Power Gains

To derive the average BLER at users $H_{\hat{i}}$ and $L_{\hat{j}}$, we first need to calculate the CDFs of $g_{SH}$ and $g_{SL}$ with TAS/SC and TAS/MRC schemes in both HCS and LCS methods. Note that these derivations are based on the MIMO diversity techniques, channel distribution, and the antenna-user selection methods utilized for the analysis, regardless of the types of transmission (e.g., SPC or long data-packet transmissions). This is described as follows:

#### HCS Method

The CDFs of $g_{SH}$ and $g_{SL}$ with HCS method are derived in the following propositions.

**Proposition 1.** *Under HCS method and Nakagami-m fading, the CDF of $g_{SH}$ with TAS/SC and TAS/MRC schemes is given by*

$$F_{g_{SH}}^{HCS}(x) = 1 + \sum_{p=1}^{a_{H,I}} \sum_{\Delta_H = p} \Phi_H c_{H,I} x^{\varphi_H} e^{-\frac{p m_H x}{\lambda_{SH}}}, \tag{3.25}$$

*where* $\Delta_H = \sum_{q=0}^{b_H-1} \delta_{H,q}$, $\varphi_H = \sum_{q=0}^{b_H-1} q\delta_{H,q}$, $\Phi_H = (-1)^p \left[ \prod_{q=0}^{b_H-1} \left( \frac{m_H^q}{q! \lambda_{SH}^q} \right)^{\delta_{H,q}} \right]$, $\lambda_{SH} = \Omega_H d_{SH}^{-\theta}$, $a_{H,I} = \begin{cases} K_S K_H I, & \text{for TAS/SC} \\ K_S I, & \text{for TAS/MRC} \end{cases}$, $b_H = \begin{cases} m_H, & \text{for TAS/SC} \\ m_H K_H, & \text{for TAS/MRC} \end{cases}$, *and* $c_{H,I} = \begin{pmatrix} a_{H,I} \\ p \end{pmatrix} \begin{pmatrix} p \\ \delta_{H,0}, \ldots, \delta_{H,b_H-1} \end{pmatrix}$, *$d_{SH}$ and $\theta$ denote the distance and path loss exponent of the link from BS $S$ to user $H_{\hat{i}}$, respectively.*

*Proof.* See Appendix A.1. □

**Proposition 2.** *Under HCS method and Nakagami-m fading, the CDF of $g_{SL}$ with TAS/SC and TAS/MRC schemes is expressed as*

$$F_{g_{SL}}^{HCS}(x) = 1 + \sum_{p=1}^{a_{L,I}} \sum_{\Delta_L = p} \Phi_L c_{L,I} x^{\varphi_L} e^{-\frac{p m_L x}{\lambda_{SL}}}, \tag{3.26}$$

*where* $\Delta_L = \sum_{q=0}^{b_L-1} \delta_{L,q}$, $\varphi_L = \sum_{q=0}^{b_L-1} q\delta_{L,q}$, $\Phi_L = (-1)^p \left[ \prod_{q=0}^{b_L-1} \left( \frac{m_L^q}{q! \lambda_{SL}^q} \right)^{\delta_{L,q}} \right]$, $\lambda_{SL} = \Omega_L d_{SL}^{-\theta}$, $a_{L,I} = \begin{cases} K_L J, & \text{for TAS/SC} \\ J, & \text{for TAS/MRC} \end{cases}$, $b_L = \begin{cases} m_L, & \text{for TAS/SC} \\ m_L K_L, & \text{for TAS/MRC} \end{cases}$, *and* $c_{L,I} = \begin{pmatrix} a_{L,I} \\ p \end{pmatrix} \begin{pmatrix} p \\ \delta_{L,0}, \ldots, \delta_{L,b_L-1} \end{pmatrix}$, *$d_{SL}$ and $\theta$ denotes the distance of the link from BS $S$ to user $L_{\hat{j}}$, respectively.*

*Proof.* It is noted that TAS is used to select the best transmit antenna for user $H_{\hat{i}}$ in this case, hence, it is considered as a random solution for user $L_{\hat{j}}$. As such, using (3.3), (3.4), (3.5), and (3.6), the CDF of $g_{SL}$ is given by [82, 95]

$$F_{g_{SL}}(x) = \left( 1 - \sum_{p=0}^{b_L-1} \frac{m_L^p}{p! \lambda_{SL}^p} x^p e^{-\frac{m_L x}{\lambda_{SL}}} \right)^{a_{L,I}}. \tag{3.27}$$

By using binomial expansion and multinomial theorem similar to the proof of Proposition 1 in Appendix A.1, we obtain the final expression of $F_{g_{SL}}(x)$ as in (3.26) and the proof is completed. □

### LCS Method

Utilizing (3.7), (3.8), and algebraic manipulations similar to the proof of Proposition 1 in Appendix A.1, the CDF of $g_{SH}$ and $g_{SL}$ in this case are expressed as

$$F_{g_{SH}}^{LCS}(x) = 1 + \sum_{p=1}^{a_{H,II}} \sum_{\Delta_H = p} \Phi_H c_{H,II} x^{\varphi_H} e^{-\frac{p m_H x}{\lambda_{SH}}}, \tag{3.28}$$

and

$$F_{g_{SL}}^{LCS}(x) = 1 + \sum_{p=1}^{a_{L,II}} \sum_{\Delta_L = p} \Phi_L c_{L,II} x^{\varphi_L} e^{-\frac{p m_L x}{\lambda_{SL}}}, \tag{3.29}$$

where $a_{H,II} = \begin{cases} K_H I, & \text{for TAS/SC} \\ I, & \text{for TAS/MRC} \end{cases}$, $c_{H,II} = \begin{pmatrix} a_{H,II} \\ p \end{pmatrix} \begin{pmatrix} p \\ \delta_{H,0}, \ldots, \delta_{H,b_H-1} \end{pmatrix}$, $a_{L,II} = \begin{cases} K_S K_L J, & \text{for TAS/SC} \\ K_S J, & \text{for TAS/MRC} \end{cases}$, and $c_{L,II} = \begin{pmatrix} a_{L,II} \\ p \end{pmatrix} \begin{pmatrix} p \\ \delta_{L,0}, \ldots, \delta_{L,b_L-1} \end{pmatrix}$.

### 3.3.3 Average BLER Analysis of HCS Method

The derivation of the average BLER at users $H_{\hat{i}}$ and $L_{\hat{j}}$ in case of using the TAS/SC or TAS/MRC scheme with HCS method are provided in the following theorems.

**Theorem 1.** *Under HCS method and Nakagami-m fading, the average BLER at user $H_{\hat{i}}$ utilizing TAS/SC or TAS/MRC is expressed as*

$$\bar{\varepsilon}_{H_{\hat{i}}}^{HCS} \approx 1 + \frac{\chi_{H_{\hat{i}}} \alpha_{H_{\hat{i}}} \sqrt{N_{H_{\hat{i}}}}}{\gamma_0 \alpha_{L_{\hat{j}}}^2} \sum_{p=1}^{a_{H,I}} \sum_{\Delta_H = p} \sum_{q=0}^{\varphi_H} \begin{pmatrix} \varphi_H \\ q \end{pmatrix} \left( -\frac{1}{\gamma_0 \alpha_{L_{\hat{j}}}} \right)^q \Phi_H c_{H,I} e^{\frac{\omega_H}{\gamma_0 \alpha_{L_{\hat{j}}}}} \mathcal{A}_H, \quad (3.30)$$

*where*

$$\mathcal{A}_H = \begin{cases} \omega_H \Xi_{H,1} + \Xi_{H,2}, & \hat{\varphi}_H = -2 \\ -\Xi_{H,1}, & \hat{\varphi}_H = -1 \\ \omega_H^{-\hat{\varphi}_H - 1} \Xi_{H,3}, & \hat{\varphi}_H \geq 0 \end{cases},$$

$\omega_H = \frac{p m_H}{\lambda_{SH}}$, $\Xi_{H,1} = \text{Ei}\left(-\omega_H \phi_{H_{\hat{i}}}\right) - \text{Ei}\left(-\omega_H \kappa_{H_{\hat{i}}}\right)$, $\Xi_{H,2} = \frac{e^{-\omega_H \phi_{H_{\hat{i}}}}}{\phi_{H_{\hat{i}}}} - \frac{e^{-\omega_H \kappa_{H_{\hat{i}}}}}{\kappa_{H_{\hat{i}}}}$, $\Xi_{H,3} = \Gamma\left(\hat{\varphi}_H + 1, \omega_H \phi_{H_{\hat{i}}}\right) - \Gamma\left(\hat{\varphi}_H + 1, \omega_H \kappa_{H_{\hat{i}}}\right)$, $\phi_{H_{\hat{i}}} = \frac{1}{\gamma_0 \alpha_{L_{\hat{j}}}} + B_{v_{H_{\hat{i}}}}$, $\kappa_{H_{\hat{i}}} = \frac{1}{\gamma_0 \alpha_{L_{\hat{j}}}} + B_{\mu_{H_{\hat{i}}}}$, $B_x = \frac{x}{\gamma_0 \left( \alpha_{H_{\hat{i}}} - \alpha_{L_{\hat{j}}} x \right)}$, *and* $\hat{\varphi}_H = \varphi_H - q - 2$.

*Proof.* See Appendix A.2. □

**Theorem 2.** *Under HCS method and Nakagami-m fading, the average BLER at user $L_{\hat{j}}$ utilizing TAS/SC or TAS/MRC is given by*

$$\bar{\varepsilon}_{L_{\hat{j}}}^{HCS} = \bar{\varepsilon}_{L_{\hat{j}}}^{x_{H_{\hat{i}}}, HCS} + \left(1 - \bar{\varepsilon}_{L_{\hat{j}}}^{x_{H_{\hat{i}}}, HCS}\right) \bar{\varepsilon}_{L_{\hat{j}}}^{x_{L_{\hat{j}}}, HCS}, \quad (3.31)$$

*where*

$$\bar{\varepsilon}_{L_{\hat{j}}}^{x_{H_{\hat{i}}}, HCS} \approx 1 + \frac{\chi_{H_{\hat{i}}} \alpha_{H_{\hat{i}}} \sqrt{N_{H_{\hat{i}}}}}{\gamma_0 \alpha_{L_{\hat{j}}}^2} \sum_{p=1}^{a_{L,I}} \sum_{\Delta_L = p} \sum_{q=0}^{\varphi_L} \begin{pmatrix} \varphi_L \\ q \end{pmatrix} \times \left( -\frac{1}{\gamma_0 \alpha_{L_{\hat{j}}}} \right)^q \Phi_L c_{L,I} e^{\frac{\omega_L}{\gamma_0 \alpha_{L_{\hat{j}}}}} \mathcal{A}_L \left( \phi_{H_{\hat{i}}}, \kappa_{H_{\hat{i}}} \right),$$

$$\bar{\varepsilon}_{L_{\hat{j}}}^{x_{L_{\hat{j}}},HCS} = \begin{cases} \bar{\varepsilon}_{L_{\hat{j}},1}^{x_{L_{\hat{j}}},HCS}, & \psi = 0 \\ \bar{\varepsilon}_{L_{\hat{j}},2}^{x_{L_{\hat{j}}},HCS}, & 0 < \psi \leq 1 \end{cases},$$

$$\bar{\varepsilon}_{L_{\hat{j}},1}^{x_{L_{\hat{j}}},HCS} \approx 1 + \chi_{L_{\hat{j}}} \sqrt{N_{L_{\hat{j}}}} \sum_{p=1}^{a_{L,I}} \sum_{\Delta_L = p} \frac{\Phi_L c_{L,I} \hat{\omega}_L^{-\varphi_L - 1}}{\left(\alpha_{L_{\hat{j}}} \gamma_0\right)^{\varphi_L}} \Xi_{L,4},$$

$$\bar{\varepsilon}_{L_{\hat{j}},2}^{x_{L_{\hat{j}}},HCS} \approx 1 + \frac{\chi_{L_{\hat{j}}} \alpha_{L_{\hat{j}}} \sqrt{N_{L_{\hat{j}}}}}{\gamma_0 \psi^2 \alpha_{H_{\hat{i}}}^2} \sum_{p=1}^{a_{L,I}} \sum_{\Delta_L = p} \sum_{q=0}^{\varphi_L} \binom{\varphi_L}{q} \times \left(-\frac{1}{\gamma_0 \psi \alpha_{H_{\hat{i}}}}\right)^q \Phi_L c_{L,I} e^{\frac{\omega_L}{\gamma_0 \eta \alpha_{H_{\hat{i}}}}} \mathcal{A}_L \left(\phi_{L_{\hat{j}}}, \kappa_{L_{\hat{j}}}\right),$$

$$\mathcal{A}_L(x,y) = \begin{cases} \omega_L \Xi_{L,1}^{(x,y)} + \Xi_{L,2}^{(x,y)}, & \hat{\varphi}_L = -2 \\ -\Xi_{L,1}^{(x,y)}, & \hat{\varphi}_L = -1 \\ \omega_L^{-\hat{\varphi}_L - 1} \Xi_{L,3}^{(x,y)}, & \hat{\varphi}_L \geq 0 \end{cases},$$

$\Xi_{L,1}^{(x,y)} = \mathrm{Ei}(-\omega_L x) - \mathrm{Ei}(-\omega_L y)$, $\Xi_{L,2}^{(x,y)} = \frac{e^{-\omega_L x}}{x} - \frac{e^{-\omega_L y}}{y}$, $\Xi_{L,3}^{(x,y)} = \Gamma(\hat{\varphi}_L + 1, \omega_L x) - \Gamma(\hat{\varphi}_L + 1, \omega_L y)$, $x \in \left\{\phi_{H_{\hat{i}}}, \hat{\phi}_{L_{\hat{j}}}\right\}$, $y \in \left\{\kappa_{H_{\hat{i}}}, \hat{\kappa}_{L_{\hat{j}}}\right\}$, $\phi_{L_{\hat{j}}} = \frac{1}{\gamma_0 \psi \alpha_{H_{\hat{i}}}} + \hat{B}_{v_{L_{\hat{j}}}}$, $\kappa_{L_{\hat{j}}} = \frac{1}{\gamma_0 \psi \alpha_{H_{\hat{i}}}} + \hat{B}_{\mu_{L_{\hat{j}}}}$, $\hat{B}_z = \frac{z}{\gamma_0 \left(\alpha_{L_{\hat{j}}} - \psi \alpha_{H_{\hat{i}}} z\right)}$, $\Xi_{L,4} = \Gamma\left(\varphi_L + 1, \hat{\omega}_L v_{L_{\hat{j}}}\right) - \Gamma\left(\varphi_L + 1, \hat{\omega}_L \mu_{L_{\hat{j}}}\right)$, $\omega_L = \frac{p m_L}{\lambda_{SL}}$, $\hat{\varphi}_L = \varphi_L - q - 2$, *and* $\hat{\omega}_L = \frac{p m_L}{\lambda_{SL} \alpha_{L_{\hat{j}}} \gamma_0}$.

*Proof.* See Appendix A.3. □

### 3.3.4 Average BLER Analysis of LCS Method

In this case, the average BLER at user $H_{\hat{i}}$ and $L_{\hat{j}}$ are derived through the following theorems.

**Theorem 3.** *Under LCS method and Nakagami-m fading, the average BLER at user $H_{\hat{i}}$ with TAS/SC or TAS/MRC is expressed as*

$$\bar{\varepsilon}_{H_{\hat{i}}}^{LCS} \approx 1 + \frac{\chi_{H_{\hat{i}}} \alpha_{H_{\hat{i}}} \sqrt{N_{H_{\hat{i}}}}}{\gamma_0 \alpha_{L_{\hat{j}}}^2} \sum_{p=1}^{a_{H,II}} \sum_{\Delta_H = p} \sum_{q=0}^{\varphi_H} \binom{\varphi_H}{q} \left(-\frac{1}{\gamma_0 \alpha_{L_{\hat{j}}}}\right)^q \Phi_H c_{H,II} e^{\frac{\omega_H}{\gamma_0 \alpha_{L_{\hat{j}}}}} \mathcal{A}_H. \quad (3.32)$$

*Proof.* To derive $\bar{\varepsilon}_{H_{\hat{i}}}^{LCS}$ in this theorem, the algebraic manipulations similar to the derivation of $\bar{\varepsilon}_{H_{\hat{i}}}^{HCS}$ in Appendix A.2 can be utilized, where (3.28) is employed instead of (3.25). □

**Theorem 4.** *Under LCS method and Nakagami-m fading, the average BLER at user $L_{\hat{j}}$ with TAS/SC or TAS/MRC is given by*

$$\bar{\varepsilon}_{L_{\hat{j}}}^{LCS} = \bar{\varepsilon}_{L_{\hat{j}}}^{x_{H_{\hat{i}}},LCS} + \left(1 - \bar{\varepsilon}_{L_{\hat{j}}}^{x_{H_{\hat{i}}},LCS}\right) \bar{\varepsilon}_{L_{\hat{j}}}^{x_{L_{\hat{j}}},LCS}, \tag{3.33}$$

*where*

$$\bar{\varepsilon}_{L_{\hat{j}}}^{x_{H_{\hat{i}}},LCS} \approx 1 + \frac{\chi_{H_{\hat{i}}}\alpha_{H_{\hat{i}}}\sqrt{N_{H_{\hat{i}}}}}{\gamma_0\alpha_{L_{\hat{j}}}^2} \sum_{p=1}^{a_{L,II}} \sum_{\Delta_L=p} \sum_{q=0}^{\varphi_L} \binom{\varphi_L}{q} \times \left(-\frac{1}{\gamma_0\alpha_{L_{\hat{j}}}}\right)^q \Phi_L c_{L,II} e^{\frac{\omega_L}{\gamma_0\alpha_{L_{\hat{j}}}}} \mathcal{A}_L\left(\phi_{H_{\hat{i}}}, \kappa_{H_{\hat{i}}}\right),$$

$$\bar{\varepsilon}_{L_{\hat{j}}}^{x_{L_{\hat{j}}},LCS} = \begin{cases} \bar{\varepsilon}_{L_{\hat{j}},1}^{x_{L_{\hat{j}}},LCS}, & \psi = 0 \\ \bar{\varepsilon}_{L_{\hat{j}},2}^{x_{L_{\hat{j}}},LCS}, & 0 < \psi \leq 1 \end{cases},$$

$$\bar{\varepsilon}_{L_{\hat{j}},1}^{x_{L_{\hat{j}}},LCS} \approx 1 + \chi_{L_{\hat{j}}}\sqrt{N_{L_{\hat{j}}}} \sum_{p=1}^{a_{L,II}} \sum_{\Delta_L=p} \frac{\Phi_L c_{L,II} \hat{\omega}_L^{-\varphi_L-1}}{\left(\alpha_{L_{\hat{j}}}\gamma_0\right)^{\varphi_L}} \Xi_{L,4},$$

*and*

$$\bar{\varepsilon}_{L_{\hat{j}},2}^{x_{L_{\hat{j}}},LCS} \approx 1 + \frac{\chi_{L_{\hat{j}}}\alpha_{L_{\hat{j}}}\sqrt{N_{L_{\hat{j}}}}}{\gamma_0\psi^2\alpha_{H_{\hat{i}}}^2} \sum_{p=1}^{a_{L,II}} \sum_{\Delta_L=p} \sum_{q=0}^{\varphi_L} \binom{\varphi_L}{q} \times \left(-\frac{1}{\gamma_0\psi\alpha_{H_{\hat{i}}}}\right)^q \Phi_L c_{L,II} e^{\frac{\omega_L}{\gamma_0\eta\alpha_{H_{\hat{i}}}}} \mathcal{A}_L\left(\phi_{L_{\hat{j}}}, \kappa_{L_{\hat{j}}}\right).$$

*Proof.* The proof of this theorem can be carried out in the same way as the proof of Theorem 2, where (3.29) is used instead of (3.26). □

## 3.4 Proposed analytical framework for optimal power allocation and minimum blocklength

By following the average BLER analysis presented in Section 3.3, this section provides the derivation of the optimal power allocation coefficients for a minimum blocklength[5] based

[5]In this chapter, we focus on minimizing the blocklength in NOMA-based SPC systems to reduce the latency for two users having the best channel conditions, which are selected from two predefined clusters $H$ and $L$. However, investigating suitable user pairing methods to guarantee latency requirements for the

on asymptotic average BLER in high SNR regime, and it also presents the analytical comparison of the minimum blocklength of NOMA with the OMA case.

### 3.4.1 Asymptotic Average BLER Analysis

As discussed in [11,12], the average BLER, $\bar{\varepsilon}_U^{x_V}$, in (3.22) can be simplified by utilizing the first-order Riemann integral approximation, i.e., $\int_a^b f(x)dx = (b-a)f\left(\frac{a+b}{2}\right)$, as follows:

$$\bar{\varepsilon}_U^{x_V} \approx \chi_V \sqrt{N_V}\,(\mu_V - v_V)\, F_{\gamma_U^{x_V}}\left(\frac{v_V+\mu_V}{2}\right). \tag{3.34}$$

By substituting $v_V$ and $\mu_V$ defined in (3.20) into (3.34), $\bar{\varepsilon}_U^{x_V}$ is rewritten as

$$\bar{\epsilon}_U^{x_V} \approx F_{\gamma_U^{x_V}}(\beta_V), \tag{3.35}$$

where $\beta_V$ is defined in (3.21).

By using the series representation of $e^x$ in [96, Eq. 1.211], i.e., $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, the asymptotic CDF of $\gamma_{H_{\hat{i}}}^{x_{H_{\hat{i}}}}$, $\gamma_{L_{\hat{j}}}^{x_{H_{\hat{i}}}}$, and $\gamma_{L_{\hat{j}}}^{x_{L_{\hat{j}}}}$ are respectively given by

$$F_{\gamma_{H_{\hat{i}}}}^{s,\infty}(x) = F_{g_{SH}}^{s,\infty}(B_x) \overset{\gamma_0\to\infty}{\approx} \frac{(m_H B_x)^{b_H a_{H,r}}}{(b_H!)^{a_{H,r}} \lambda_{SH}^{b_H a_{H,r}}}, \tag{3.36}$$

$$F_{\gamma_{L_{\hat{j}}}}^{s,\infty}(x) \overset{\gamma_0\to\infty}{\approx} \frac{(m_L B_x)^{b_L a_{L,r}}}{(b_L!)^{a_{L,r}} \lambda_{SL}^{b_L a_{L,r}}}, \tag{3.37}$$

and

$$F_{\gamma_{L_{\hat{j}}}}^{s,\infty}(x) \overset{\gamma_0\to\infty}{\approx} \frac{\left(m_L \hat{B}_x\right)^{b_L a_{L,r}}}{(b_L!)^{a_{L,r}} \lambda_{SL}^{b_L a_{L,r}}}, \tag{3.38}$$

where $s \in \{HCS, LCS\}$, $r = \begin{cases} I, & \text{if } s = HCS \\ II, & \text{if } s = LCS \end{cases}$, $B_x = \frac{x}{\gamma_0\left(\alpha_{H_{\hat{i}}} - \alpha_{L_{\hat{j}}} x\right)}$, and $\hat{B}_x = \frac{x}{\gamma_0\left(\alpha_{L_{\hat{j}}} - \psi\alpha_{H_{\hat{i}}} x\right)}$. From (3.34) - (3.38), the asymptotic average BLER at users $H_{\hat{i}}$ and $L_{\hat{j}}$

users having weak channel gains is an important research issue to be addressed in future works.

are respectively expressed as

$$\bar{\varepsilon}_{H_{\hat{i}}}^{s,\infty} \approx \frac{\left(m_H B_{\beta_{H_{\hat{i}}}}\right)^{b_H a_{H,r}}}{(b_H!)^{a_{H,r}} \lambda_{SH}^{b_H a_{H,r}}}, \tag{3.39}$$

and

$$\begin{aligned} \bar{\varepsilon}_{L_{\hat{j}}}^{s,\infty} &= \bar{\varepsilon}_{L_{\hat{j}},\infty}^{x_{H_{\hat{i}}},s} + \left(1 - \bar{\varepsilon}_{L_{\hat{j}},\infty}^{x_{H_{\hat{i}}},s}\right) \bar{\varepsilon}_{L_{\hat{j}},\infty}^{x_{L_{\hat{j}}},s} \\ &\approx \bar{\varepsilon}_{L_{\hat{j}},\infty}^{x_{H_{\hat{i}}},s} + \bar{\varepsilon}_{L_{\hat{j}},\infty}^{x_{L_{\hat{j}}},s} \\ &\approx \frac{\left(m_L B_{\beta_{H_{\hat{i}}}}\right)^{b_L a_{L,r}}}{(b_L!)^{a_{L,r}} \lambda_{SL}^{b_L a_{L,r}}} + \frac{\left(m_L \hat{B}_{\beta_{L_{\hat{j}}}}\right)^{b_L a_{L,r}}}{(b_L!)^{a_{L,r}} \lambda_{SL}^{b_L a_{L,r}}}. \end{aligned} \tag{3.40}$$

From (3.39) and (3.40), the diversity order at users $H_{\hat{i}}$ and $L_{\hat{j}}$ are respectively given by [89]:

$$\begin{aligned} D_{H_{\hat{i}}} &= -\lim_{\gamma_0 \to \infty} \frac{\log\left(\bar{\varepsilon}_{H_{\hat{i}}}^{s,\infty}\right)}{\log(\gamma_0)} \\ &= \begin{cases} m_H K_S K_H I, & \text{for HCS method} \\ m_H K_H I, & \text{for LCS method} \end{cases}, \end{aligned} \tag{3.41}$$

and

$$\begin{aligned} D_{L_{\hat{j}}} &= -\lim_{\gamma_0 \to \infty} \frac{\log\left(\bar{\varepsilon}_{L_{\hat{j}}}^{s,\infty}\right)}{\log(\gamma_0)} \\ &= \begin{cases} m_L K_L J, & \text{for HCS method} \\ m_L K_S K_L J, & \text{for LCS method} \end{cases}. \end{aligned} \tag{3.42}$$

**Remark 5.** *For both TAS/SC and TAS/MRC schemes, the diversity orders at users $H_{\hat{i}}$ and $L_{\hat{j}}$, denoted by $\left(D_{H_{\hat{i}}}, D_{L_{\hat{j}}}\right)$, are $(m_H K_S K_H I, m_L K_L J)$ for HCS method, and $(m_H K_H I, m_L K_S K_L J)$ for LCS method. This reveals that the users $H_{\hat{i}}$ and $L_{\hat{j}}$ have achieved full diversity order with HCS and LCS methods, respectively. Furthermore, the system performance of user $H_{\hat{i}}$ can be improved by increasing $m_H$, $K_S$, $K_H$, and $I$ with HCS method, and by increasing $m_H$, $K_H$, and $I$ with LCS method. Meanwhile, the growth of $m_L$, $K_L$, and $J$ with HCS method, and $m_L$, $K_S$, $K_L$, and $J$ with LCS method can help enhancing the system performance of user $L_{\hat{j}}$.*

### 3.4.2 Power and Blocklength Optimization at High SNR

In this subsection, we focus on the problem of blocklength minimization[6] subject to BLER targets and power allocation coefficients to guarantee the reliability requirement and reduce the transmission latency for SPC-based MIMO NOMA systems [9,11,16]. To determine the values of power allocation coefficients (i.e., $\alpha_{H_{\hat{i}}}$ and $\alpha_{L_{\hat{j}}}$) at which a minimum blocklength $N_U\left(U \in \left\{H_{\hat{i}}, L_{\hat{j}}\right\}\right)$ is achieved to guarantee the reliability target $\bar{\varepsilon}_U^{th}$, the following problem needs to be addressed

$$\min_{\alpha_{H_{\hat{i}}}, \alpha_{L_{\hat{j}}}} \quad N_U \tag{3.43a}$$

$$\text{s.t.} \quad \bar{\varepsilon}_U \leq \bar{\varepsilon}_U^{th}, \tag{3.43b}$$

$$\alpha_{H_{\hat{i}}} + \alpha_{L_{\hat{j}}} = 1,\ 0 < \alpha_{L_{\hat{j}}} < 0.5, \tag{3.43c}$$

where the constraint (3.43c) is obtained based on the NOMA principle presented in Section 3.2, in which $0 < \alpha_{L_{\hat{j}}} < \alpha_{H_{\hat{i}}}$ and $\alpha_{L_{\hat{j}}} + \alpha_{H_{\hat{i}}} = 1$, leading to $\alpha_{L_{\hat{j}}} < 0.5$. It is noted that $\alpha_{H_{\hat{i}}} = 1 - \alpha_{L_{\hat{j}}}$ and $\bar{\varepsilon}_U$ is a decreasing function of $N_U$. The problem in (3.43) can be simplified as

$$\min_{\alpha_{L_{\hat{j}}}} \quad N_U \tag{3.44a}$$

$$\text{s.t.} \quad \bar{\varepsilon}_U \leq \bar{\varepsilon}_U^{th}, \tag{3.44b}$$

$$0 < \alpha_{L_{\hat{j}}} < 0.5. \tag{3.44c}$$

By substituting (3.39) into (3.44b) for user $H_{\hat{i}}$ and (3.40) into (3.44b) for user $L_{\hat{j}}$, the blocklengths of users $H_{\hat{i}}$ and $L_{\hat{j}}$ with $s$ ($s \in \{HCS, LCS\}$) method are respectively calculated as

$$N_{H_{\hat{i}},s} = \frac{n_{H_{\hat{i}}}}{\log_2\left(\frac{1+\tau_{H,r}}{1+\alpha_{L_{\hat{j}}}\tau_{H,r}}\right)}, \tag{3.45}$$

and

$$N_{L_{\hat{j}},s} = \frac{n_{L_{\hat{j}}}}{\log_2\left\{1 + \frac{\alpha_{L_{\hat{j}}}\tau_{L,r}}{1+\psi\left(1-\alpha_{L_{\hat{j}}}\right)\tau_{L,r}}\right\}}, \tag{3.46}$$

[6]It is noted that addressing the optimization problems subject to latency requirement (e.g., decoding error probability minimization under the latency constraint) is also an important issue to be tackled in the SPC-based systems to ensure the expected latency [24, 25]. This would be a noteworthy problem to investigate in future work.

where $\eta_{H,r} = \frac{m_H^{b_H a_{H,r}}}{(b_H!)^{a_{H,r}} \lambda_{SH}^{b_H a_{H,r}} \gamma_0^{b_H a_{H,r}}}$, $\eta_{L,r} = \frac{(b_H!)^{\frac{b_L a_{L,r}}{b_H}}}{(b_L!)^{a_{L,r}}} \left(\frac{m_L \lambda_{SH}}{m_H \lambda_{SL}}\right)^{b_L a_{L,r}}$, $\hat{\eta}_{L,r} = \frac{m_L^{b_L a_{L,r}}}{(b_L!)^{a_{L,r}} \lambda_{SL}^{b_L a_{L,r}}}$, $\tau_{H,r} = \left(\bar{\varepsilon}^{th}_{H_{\hat{i}},r}/\eta_{H,r}\right)^{1/b_H a_{H,r}}$, and $\tau_{L,r} = \gamma_0 \left[\frac{\bar{\varepsilon}^{th}_{L_{\hat{j}},r} - \eta_{L,r}\left(\varepsilon^{th}_{H_{\hat{i}},r}\right)^{\frac{b_L a_{L,r}}{b_H a_{H,r}}}}{\hat{\eta}_{L,r}}\right]^{1/b_L a_{L,r}}$.

From (3.45) and (3.46), the derivative of $N_{H_{\hat{i}},s}$ and $N_{L_{\hat{j}},s}$ with respect to $\alpha_{L_{\hat{j}}}$ can be derived as follows:

$$\frac{\partial N_{H_{\hat{i}},s}}{\partial \alpha_{L_{\hat{j}}}} = \frac{n_{H_i} \tau_{H,r}}{\left(1 + \alpha_{L_{\hat{j}}} \tau_{H,r}\right) \left[\log_2\left(\frac{1+\tau_{H,r}}{1+\alpha_{L_{\hat{j}}} \tau_{H,r}}\right)\right]^2 \ln 2} > 0, \tag{3.47}$$

and

$$\frac{\partial N_{L_{\hat{j}},s}}{\partial \alpha_{L_{\hat{j}}}} = -\frac{n_{L_{\hat{j}}} \left(1 + \psi \tau_{L,r}\right) \hat{\tau}_{L,r}}{\left(1 + \alpha_{L_{\hat{j}}} \hat{\tau}_{L,r}\right) \left[\log_2\left(1 + \alpha_{L_{\hat{j}}} \hat{\tau}_{L,r}\right)\right]^2 \ln 2} < 0, \tag{3.48}$$

where $\hat{\tau}_{L,r} = \frac{\tau_{L,r}}{1+\psi\left(1-\alpha_{L_{\hat{j}}}\right)\tau_{L,r}}$. Thus, $N_{H_{\hat{i}},s}$ is an increasing function of $\alpha_{L_{\hat{j}}}$, whereas $N_{L_{\hat{j}},s}$ is a decreasing function of $\alpha_{L_{\hat{j}}}$. Therefore, to guarantee both reliability targets $\bar{\varepsilon}^{th}_{H_{\hat{i}},r}$ and $\bar{\varepsilon}^{th}_{L_{\hat{j}},r}$, the minimum blocklength is obtained by addressing $N_{H_{\hat{i}},s} = N_{L_{\hat{j}},s} = N_{opt,s}$ and the problem of minimizing blocklength in (3.44) is rewritten as

$$\min_{\alpha_{L_{\hat{j}}}} \quad N_{opt,s} \tag{3.49a}$$

$$\text{s.t.} \quad \bar{\varepsilon}^{s}_{H_{\hat{i}}} = \bar{\varepsilon}^{th}_{H_{\hat{i}},r}, \tag{3.49b}$$

$$\bar{\varepsilon}_{L_{\hat{j}},s} = \bar{\varepsilon}^{th}_{L_{\hat{j}},r}, \tag{3.49c}$$

$$0 < \alpha_{L_{\hat{j}}} < 0.5. \tag{3.49d}$$

Given this context, the optimal power allocation coefficient $\alpha_{L_{\hat{j}},opt}$ to minimize $N_{opt,s}$ can be achieved by solving the equation $f\left(\alpha_{L_{\hat{j}}}\right) = N_{L_{\hat{j}},s} - N_{H_{\hat{i}},s} = 0$, which is addressed in Algorithm 1. The minimum blocklength $N_{opt,r}$ is attained by substituting $\alpha_{L_{\hat{j}},opt}$ into (3.45) as follows:

$$N_{opt,s} = \frac{n_{H_{\hat{i}}}}{\log_2\left(\frac{1+\left(\bar{\varepsilon}^{th}_{H_{\hat{i}},r}/\eta_{H,r}\right)^{1/b_H a_{H,r}}}{1+\alpha_{L_{\hat{j}},opt}\left(\bar{\varepsilon}^{th}_{H_{\hat{i}},r}/\eta_{H,r}\right)^{1/b_H a_{H,r}}}\right)}. \tag{3.50}$$

**Algorithm 1** Proposed Power Allocation Algorithm for SPC-Based MIMO NOMA System

Initialize $n_{H_{\hat{i}}}$, $n_{L_{\hat{j}}}$, $\gamma_0$, $\bar{\varepsilon}^{th}_{H_{\hat{i}},r}$, $\bar{\varepsilon}^{th}_{L_{\hat{j}},r}$, $K_S$, $K_H$, $K_L$, $I$, $J$, $\lambda_{SH}$, $\lambda_{SL}$, and tolerance $\mu$.
Initialize $\alpha^{-}_{L_{\hat{j}}} \leftarrow 0$, $\alpha^{+}_{L_{\hat{j}}} \leftarrow 0.5$, and $\hat{\alpha}_{L_{\hat{j}}} \leftarrow \frac{\alpha^{-}_{L_{\hat{j}}}+\alpha^{+}_{L_{\hat{j}}}}{2}$.
**repeat**
    **if** $f\left(\hat{\alpha}_{L_{\hat{j}}}\right) f\left(\alpha^{-}_{L_{\hat{j}}}\right) > 0$ **then**
        Set $\alpha^{-}_{L_{\hat{j}}} \leftarrow \hat{\alpha}_{L_{\hat{j}}}$.
    **else**
        Set $\alpha^{+}_{L_{\hat{j}}} \leftarrow \hat{\alpha}_{L_{\hat{j}}}$.
    **end if**
    Set $\hat{\alpha}_{L_{\hat{j}}} \leftarrow \frac{\alpha^{-}_{L_{\hat{j}}}+\alpha^{+}_{L_{\hat{j}}}}{2}$ and compute $f\left(\hat{\alpha}_{L_{\hat{j}}}\right)$ based on (3.45) and (3.46).
**until** $\left|f\left(\hat{\alpha}_{L_{\hat{j}}}\right)\right| > \mu$.
Set $\alpha_{L_{\hat{j}},opt} \leftarrow \hat{\alpha}_{L_{\hat{j}}}$.
**Return** $\alpha_{L_{\hat{j}},opt}$.

### 3.4.3 Comparison of MIMO NOMA and MIMO OMA Schemes

To perform the comparison between MIMO NOMA and MIMO OMA schemes, we consider a scenario where users $H_{\hat{i}}$ and $L_{\hat{j}}$ are served simultaneously by using NOMA or over different time-slots by utilizing OMA (i.e., time division multiple access). Herein, a MIMO scheme, i.e., TAS, is utilized for both NOMA and OMA scenarios to reduce the complexity of the signal processing and feedback overhead [60,81]. With OMA transmission, the minimum blocklength, $N_{OMA,s}$ ($r \in \{I, II\}$), is the summation of the minimum blocklengths for users $H_{\hat{i}}$ and $L_{\hat{j}}$, $\hat{N}_{H_{\hat{i}}}$ and $\hat{N}_{L_{\hat{j}}}$. Similar to the derivation of blocklengths for users $H_{\hat{i}}$ and $L_{\hat{j}}$ in Section 3.4.2, $N_{OMA,s}$ in the high SNR regime is calculated as

$$\begin{aligned} N_{OMA,s} &= \hat{N}_{H_{\hat{i}}} + \hat{N}_{L_{\hat{j}}} \\ &= \frac{n_{H_{\hat{i}}}}{\log_2\left[1+\left(\bar{\varepsilon}^{th}_{H_{\hat{i}},r}/\eta_{H,r}\right)^{1/b_H a_{H,r}}\right]} + \frac{n_{L_{\hat{j}}}}{\log_2\left[1+\gamma_0\left(\bar{\varepsilon}^{th}_{L_{\hat{j}},r}/\hat{\eta}_{L,r}\right)^{1/b_L a_{L,r}}\right]}. \end{aligned} \tag{3.51}$$

From (3.50) and (3.51), the blocklength gap between NOMA and OMA, i.e., $\Delta N_s$, can be given by

$$\Delta N_r = N_{OMA,r} - N_{opt,r} \approx \hat{N}_{H_{\hat{i}},r} > 0. \tag{3.52}$$

Thus, OMA transmission needs a longer blocklength than NOMA transmission to serve the users $H_{\hat{i}}$ and $L_{\hat{j}}$.

## 3.5 Numerical Results

In this section, we provide numerical results in terms of average BLER and minimum blocklength to characterize the effects of the proposed protocols, i.e., HCS and LCS methods with TAS/SC and TAS/MRC schemes discussed in Section 3.2.1, on the system performance in designing an SPC-based MIMO NOMA network. It is noted that the analysis of these performance metrics have practical significance for the reliability and latency performance evaluation of wireless systems [11–16]. The predetermined simulation parameters are set as follows [11–13]: the number of information bits $n_{H_{\hat{i}}} = n_{L_{\hat{j}}} = 80$ bits; the blocklength $N_{H_{\hat{i}}} = N_{L_{\hat{j}}} = 100$; the path loss exponent $\theta = 2.5$; the distances $d_{SH} = d_{SL} = 5$ (m); the power allocation coefficients $\alpha_{H_{\hat{i}}} = 0.7$, and $\alpha_{H_{\hat{i}}} = 0.3$; the reliability targets $\bar{\varepsilon}^{th}_{H_{\hat{i}}} = 10^{-7}$ and $\bar{\varepsilon}^{th}_{L_{\hat{j}}} = 10^{-6}$.

To evaluate our BLER performance analysis carried out in Section 3.3, we provide the numerical outcomes through the following three types of result: i) Analytical result (Ana.), ii) Asymptotic result (Asymp.), and iii) Simulation result (Sim.). For the simulation results, similar to the method used in [11–14, 16], we create $10^6$ Nakagami-$m$ channel realizations generated randomly through the Nakagami-$m$ distribution given in (3.1) and (3.2) for all the considered schemes. The respective average BLERs are then computed by averaging the instantaneous BLERs according to these generated channel realizations while considering the Gaussian-coded symbols instead of the real symbol constellations. The definition of the instantaneous BLER is provided in (3.19), which has also been used in [11–14, 16]. Furthermore, the different values of Nakagami-$m$ fading parameters, i.e., $m_H$ and $m_L$, are considered in the presented numerical results. For the analytical and asymptotic results, they are obtained by adopting the expressions derived in (3.30), (3.31), (3.32), (3.33) for the analytical results, and (3.39), and (3.40) for the asymptotic results, respectively.

In Fig. 3.2, we perform the rate comparison of SPC and long-packet communications (LPC) for user $H_{\hat{i}}$ to gain more insights on SPC. Note that based on (3.17) and the normal approximation method in [9], the achievable rate of user $H_{\hat{i}}$ with SPC can be approximated as: $R^{SPC}_{H_{\hat{i}}} \approx log_2\left(1+\gamma^{x_{H_{\hat{i}}}}_{H_{\hat{i}}}\right) - \sqrt{v/N}Q^{-1}(\varepsilon) + \log_2 N/2N$. We can observe from this figure that when the blocklength of user $H_{\hat{i}}$ ($N_{H_{\hat{i}}}$) increases, the achievable rate of this user with LPC ($R^{LPC}_{H_{\hat{i}}}$) is unchanged, whereas the rate with the SPC, $R^{SPC}_{H_{\hat{i}}}$ grows up. This can be explained by the fact that the LPC is implemented under the assumption of infinite blocklength to obtain the Shannon's channel rate, i.e., $log_2\left(1+\gamma^{x_{H_{\hat{i}}}}_{H_{\hat{i}}}\right)$, which is not influenced by $N_{H_{\hat{i}}}$. In contrast, finite blocklength is utilized in the SPC scenario. Given the above approximation of $R^{SPC}_{H_{\hat{i}}}$, the increase in $N_{H_{\hat{i}}}$ leads to the higher value of $R^{SPC}_{H_{\hat{i}}}$ in this case, and $R^{SPC}_{H_{\hat{i}}} \to R^{LPC}_{H_{\hat{i}}}$ when $N_{H_{\hat{i}}} \to \infty$. Since (3.17) is applied for both users $H_{\hat{i}}$ and $L_{\hat{j}}$, the same conclusions can also be achieved for user $L_{\hat{j}}$. Furthermore, this figure indicates that HCS method with TAS/MRC provides the best performance for user

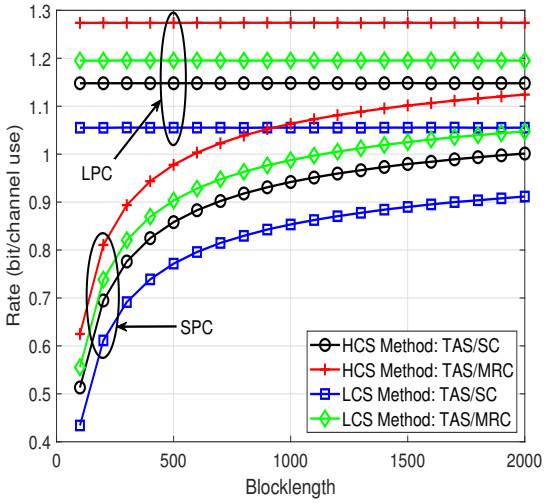



FIGURE 3.2: Rate comparison between SPC and LPC, where $\gamma_0 = 20$ (dB), $m_H = m_L = 2$, and $K_S = K_H = K_L = I = J = 2$.

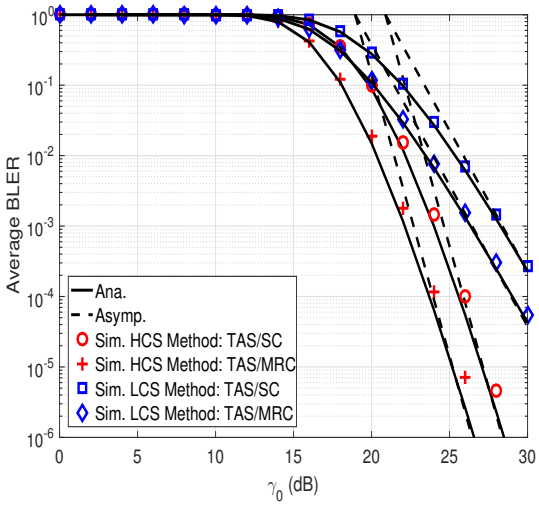



FIGURE 3.3: Average BLER at user $H_{\hat{i}}$ vs. $\gamma_0$ with different methods, where $m_H = m_L = 2$ and $(K_S, K_H, I) = (2, 2, 1)$.

$H_{\hat{i}}$. The effects of the proposed protocols on the performance of users $H_{\hat{i}}$ and $L_{\hat{j}}$ will be clearly analyzed in the following.

Similar to [11–13], we utilize the simulation results to evaluate the correctness of our analysis. Specifically, in Figs. 3.3 and 3.4, we plot the average BLERs at users $H_{\hat{i}}$ and $L_{\hat{j}}$ as a function of $\gamma_0$ with different methods (i.e., HCS method with TAS/SC or TAS/MRC and LCS method with TAS/SC or TAS/MRC). As can be observed from these figures, the BLER gap between the approximated analytical results and the simulation results are very small. Furthermore, the asymptotic curves accurately predict the system performance trend in the higher $\gamma_0$ regime. These observations confirm the correctness of our analysis in Section 3.4. In addition, Figs. 3.3 and 3.4 show that HCS method achieves better performance (i.e., lower value of average BLER is observed) for user $H_{\hat{i}}$ over LCS method, whereas LCS method outperforms HCS method in terms of the system performance for user $L_{\hat{j}}$. This result is achieved based on the fact that HCS and LCS methods are proposed to improve the received signal quality at users $H_{\hat{i}}$ and $L_{\hat{j}}$, respectively, as discussed in Section 3.2.1. Furthermore, these figures indicate that TAS/MRC scheme is better than TAS/SC in improving the system performance.

In Figs. 3.5 and 3.6, we investigate the effects of the number of users at clusters $H$ $(I)$ and $L$ $(J)$, and the number of antennas at BS $S$ $(K_S)$, users $H_{\hat{i}}$ $(K_H)$, and $L_{\hat{j}}$ $(K_L)$, on the system performance. Specifically, Fig. 3.5 shows the variation of average BLER at user $H_{\hat{i}}$ with respect to $\gamma_0$ with different values of $K_S$, $K_H$, and $I$, denoted by $(K_S, K_H, I)$, in

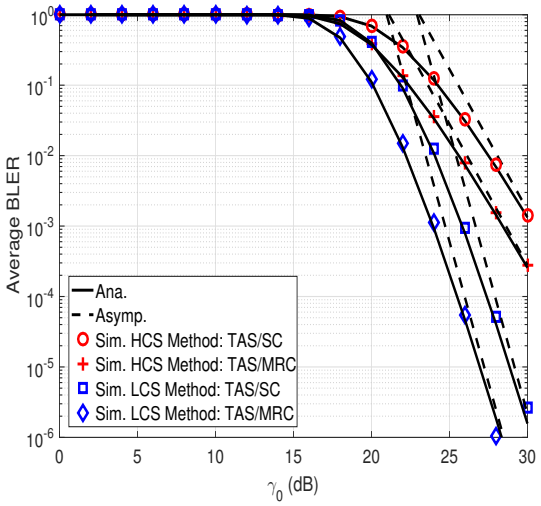



FIGURE 3.4: Average BLER at user $L_{\hat{j}}$ vs. $\gamma_0$ with different methods, where $\psi = 0$, $m_H = m_L = 2$ and $(K_S, K_L, J) = (2, 2, 1)$.

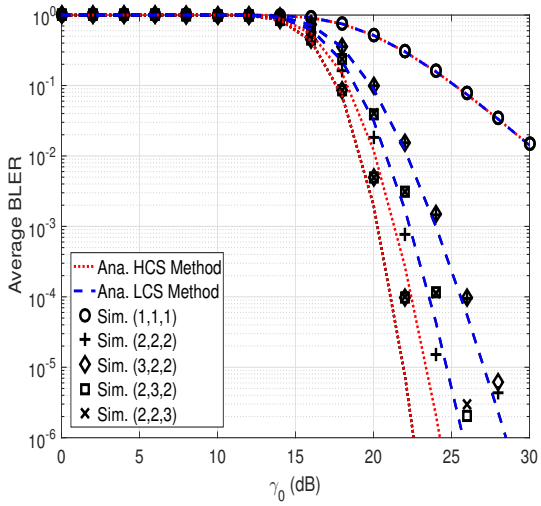



FIGURE 3.5: Average BLER at user $H_{\hat{i}}$ vs. $\gamma_0$ with different values of $(K_S, K_H, I)$, where $m_H = m_L = 2$.

case of utilizing HCS and LCS methods with the TAS/SC scheme. Meanwhile, Fig. 3.6 plots the average BLER at user $L_{\hat{j}}$ versus $\gamma_0$ with different values of $(K_S, K_L, J)$ when using HCS and LCS methods with the TAS/SC scheme. These two figures indicate that as $K_S$, $K_H$, $K_L$, $I$, and $J$ are all equal to one, HCS and LCS methods result in the same curves. Furthermore, the system performance can be significantly improved by increasing $(K_S, K_H, I)$ for user $H_{\hat{i}}$ and $(K_S, K_L, J)$ for user $L_{\hat{j}}$. It is noted that the variation of $K_S$ in LCS method does not impact the system performance at user $H_{\hat{i}}$ (see Fig. 3.5). The same conclusion can be derived for user $L_{\hat{j}}$ when observing the change of $K_S$ in HCS method (see Fig. 3.6). The reason for this is based on the nature of HCS and LCS methods as mentioned in Section 3.2.1 and the discussion part of Figs. 3.3 and 3.4. This phenomenon also confirms our analysis of diversity order for users $H_{\hat{i}}$ and $L_{\hat{j}}$, as shown in Section 3.4.1.

Fig. 3.7 depicts the average BLER at user $L_{\hat{j}}$ as a function of $\gamma_0$ with different values of the residual interference level caused by the ISIC, i.e., $\psi$, in case of using HCS and LCS methods with TAS/SC scheme. In other results presented in this section, we investigate a scenario where the value of $\psi$ is fixed to evaluate the effects of other parameters such as antenna-user selection methods, the number of users, the number of antennas, fading parameters, and power allocation coefficients, on the system performance. In contrast, Fig. 3.7 shows how the variation of $\psi$ affects the BLER performance of SPC in the considered MIMO NOMA system. We can observe from this figure that the increase in $\psi$ leads to the higher interference as in (3.15), making the system performance of user $L_{\hat{j}}$ lower. Thus,

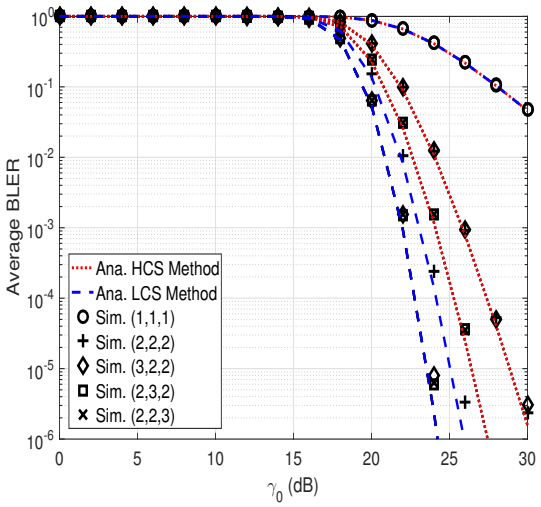



FIGURE 3.6: Average BLER at user $L_{\hat{j}}$ vs. $\gamma_0$ with different values of $(K_S, K_L, J)$, where $\psi = 0$ and $m_H = m_L = 2$.

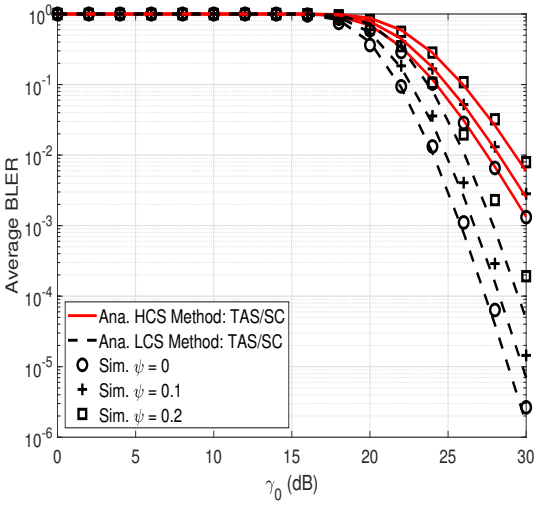



FIGURE 3.7: Average BLER at user $L_{\hat{j}}$ vs. $\gamma_0$ with different values of $\psi$ (residual interference caused due to the ISIC process), where $m_H = m_L = 2$ and $(K_S, K_L, J) = (2, 2, 1)$.

user $L_{\hat{j}}$ can achieve the best performance when $\psi = 0$, where the perfect SIC is observed, which may be difficult to obtain in practical scenarios.

In Fig. 3.8, we consider the change of average BLER at users $H_{\hat{i}}$ and $L_{\hat{j}}$ with respect to the fading parameters, i.e., $m_H$ and $m_L$, in case of using HCS and LCS methods with the TAS/SC scheme. Herein, we set $m_H = m_L = m$. Given the considered systems, it is noteworthy that $m_H$ and $m_L$ are independent and only affect users $H_{\hat{i}}$ and $L_{\hat{j}}$, respectively. We can see from this figure that the performance of users $H_{\hat{i}}$ and $L_{\hat{j}}$ can be improved with the increase in $m_H$ and $m_L$, respectively, due to the better channel quality. More precisely, when $m = 1$, Nakagami-$m$ fading corresponds to Rayleigh fading and the worst performance can be observed. In case of $m = (K+1)^2/(2K+1)$, it approximates the Rician fading with parameter $K$ [86]. This result also verifies the diversity order outcomes obtained in (3.41) and (3.42). Furthermore, similar to Figs. 3.3, 3.4, 3.5, and 3.6, Fig. 3.8 indicates that HCS and LCS methods provide the best performance for users $H_{\hat{i}}$ and $L_{\hat{j}}$, respectively.

Fig. 3.9 depicts the effect of power allocation coefficient $\alpha_{L_{\hat{j}}}$ on the blocklength of users $H_{\hat{i}}$ $\left(N_{H_{\hat{i}}}\right)$ and $L_{\hat{j}}$ $\left(N_{L_{\hat{j}}}\right)$. One can see from this figure that $N_{H_{\hat{i}}}$ and $N_{L_{\hat{j}}}$ are increasing and decreasing functions of $\alpha_{L_{\hat{j}}}$, respectively. Thus, there exists an optimal value of $\alpha_{L_{\hat{j}}}$, at which the minimum blocklength for both users $H_{\hat{i}}$ and $L_{\hat{j}}$ is achieved. The value of optimal $\alpha_{L_{\hat{j}}}$ for different cases (i.e., HCS method with TAS/SC or TAS/MRC; LCS method with

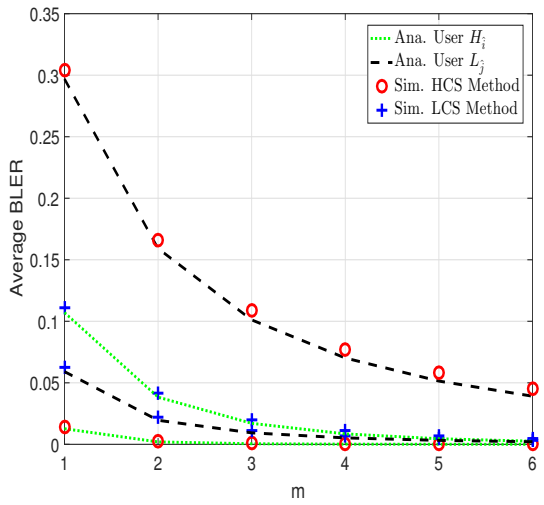



FIGURE 3.8: Average BLER at users $H_{\hat{i}}$ and $L_{\hat{j}}$ vs. $m$ with different methods, where $\gamma_0 = 20$ (dB) and $(K_S, K_H, K_L, I, J) = (2, 2, 2, 1, 1)$.

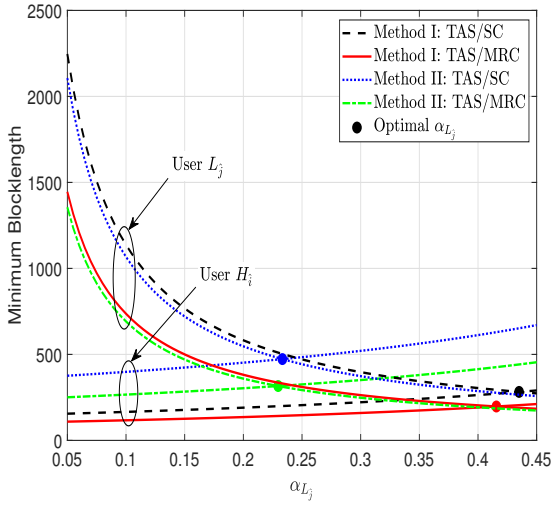



FIGURE 3.9: Minimum blocklength for users $H_{\hat{i}}$ and $L_{\hat{j}}$ vs. $\alpha_{L_{\hat{j}}}$ with different methods, where $\psi = 0$, $m_H = m_L = 2$, $K_S = K_H = K_L = I = J = 2$, and $\gamma_0 = 20$ (dB).

TAS/SC or TAS/MRC) can be found out by using Algorithm 1 and then the minimum blocklength is calculated by using (3.50).

In Fig. 3.10, we perform the minimum blocklength comparison between NOMA and OMA transmissions ($N_{opt}$ and $N_{OMA}$) to clarify the benefits of NOMA over OMA in short-packet transmissions. As can be seen from this figure, the higher blocklength gap between NOMA and OMA, i.e., $\Delta_N$ (calculated from (3.52)), is achieved in case of using HCS method and TAS/SC scheme. This implies that the benefits of MIMO NOMA versus MIMO OMA in terms of minimum blocklength are more pronounced when utilizing HCS method as compared to LCS method. Furthermore, $\Delta_N$ is positive, hence, $N_{opt}$ is always smaller than $N_{OMA}$. In other words, MIMO NOMA can lower the transmission latency of SPC systems as compared to the MIMO OMA case.

From the above achieved results, we provide some useful insights when considering SPC in the considered MIMO NOMA system as follows: i) Compared to LPC, SPC can fulfill more stringent requirements of reliability and latency for MIMO NOMA but achieves lower rate performance; ii) Transmission latency of MIMO NOMA is smaller than that of MIMO OMA in SPC scenario; and iii) Minimum blocklength for MIMO NOMA is achieved at a certain value of power allocation coefficients such that blocklengths of NOMA users are the same.

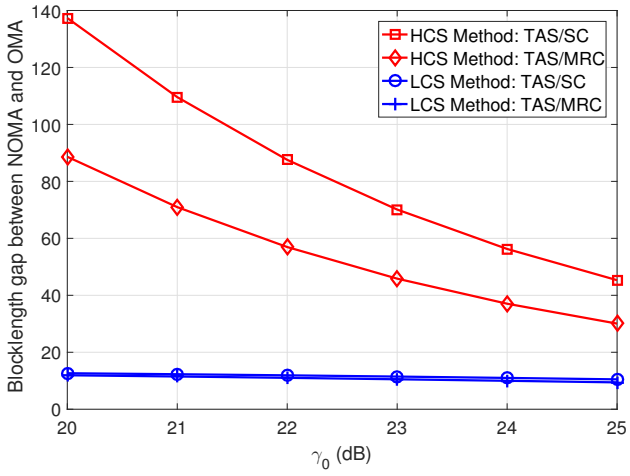



FIGURE 3.10: Blocklength comparison between NOMA and OMA.

## 3.6 Summary

In this chapter, we analyzed the performance of SPC in a QoS-based multiuser downlink MIMO NOMA system over a Nakagami-$m$ fading channel under the ISIC scenario in terms of the average BLER and minimum blocklength. Specifically, we considered the user pairing to perform NOMA, where users are selected from two user clusters having different priority levels. Furthermore, we investigated different MIMO schemes including TAS for BS, SC and MRC for users, and proposed two antenna-user selection methods, i.e., HCS and LCS to design effective communication protocols for the SPC-based MIMO NOMA systems. We characterized the system performance by deriving the approximate and asymptotic (in the high SNR regime) closed-form expressions of the average BLER at the users. From the asymptotic average BLER, we carried out an analysis of diversity order, minimum blocklength, and optimal power allocation. The analytical results verified by simulation results indicated that the system performance decreases with the increase in the value of the residual interference caused by the ISIC process. In addition, among the proposed schemes, the HCS method with TAS/MRC and the LCS method with TAS/MRC provide the best performance with full diversity gains for the users selected from the high-priority and low-priority user clusters, respectively. Moreover, it has been demonstrated that MIMO can significantly improve the performance of NOMA systems with short-packets, and MIMO NOMA outperforms MIMO OMA in ensuring low-latency transmissions.

Chapter 4

# Deep Reinforcement Learning for Resource Allocation Optimization in URLLC Systems

Grant-free non-orthogonal multiple access (GF-NOMA) has emerged as a promising access technology for the fifth generation and beyond wireless networks that enable ultra-reliable and low-latency communications (URLLC) to ensure low access latency and high connectivity density. Furthermore, designing energy-efficient (EE) resource allocation strategies is a crucial aspect of future cellular system development. Taking these goals into account, this chapter proposes an EE sub-channel and power allocation strategy for URLLC-enabled GF-NOMA (URLLC-GF-NOMA) systems based on multi-agent (MA) deep reinforcement learning (MADRL). In particular, the URLLC-GF-NOMA methods using MA dueling double deep Q network (MA3DQN), MA double deep Q network (MA2DQN), and MA deep Q network (MADQN) techniques are designed to enable users to select the most appropriate sub-channel and transmission power for their communications. The aim is to build an efficient MADRL-based solution, ensuring rapid convergence with small signaling overhead, to maximize the network EE while fulfilling the URLLC requirements of all users. Simulation results show that the MADQN and MA2DQN methods, which have lower complexity than MA3DQN, are more appropriate for the URLLC-GF-NOMA systems under consideration. Moreover, our proposed methods exhibit superior convergence characteristics, a reduction in signaling overhead, and enhanced EE performance compared to other benchmark strategies.

The chapter is organized as follows. Introduction to the current state of the art is discussed in Section 4.1. Section 4.2 presents the system model, URLLC method, and the

EE maximization problem. Section 4.3 describes the MADRL-based solution of the EE optimization problem for the considered URLLC-GF-NOMA system. Section 4.4 provides the obtained simulation results and discussions. Finally, Section 4.5 concludes this chapter.

## 4.1 Introduction

URLLC is one of the most critical services of the fifth generation (5G) and beyond wireless networks [6, 7]. It is expected to support mission-critical Internet of Things (IoT) applications, such as smart city, remote surgery, intelligent transportation, and vehicle-to-everything (V2X) communications, with stringent reliability and latency requirements. Specifically, a general URLLC condition for a one-way radio is defined as 99.999% target reliability and 1 ms latency [3, 97]. Due to the unprecedented constraints of high reliability and low latency, the packet lengths of URLLC messages are generally ultra-short. Thus, the channel's blocklength is finite, requiring a thorough analysis of achievable rate and decoding error probability. These considerations can be ignored in traditional wireless communication schemes that mostly focus on the Shannon channel capacity under the assumption of infinite blocklength [3]. Therefore, URLLC-enabled systems require a new transmission method. In this regard, SPC in finite blocklength (FBL) regime could be a promising approach to meet the URLLC requirements [3, 9].

Furthermore, one of the major challenges in 5G and beyond wireless networks is supporting massive access over a limited radio spectrum [64]. To resolve this challenge, non-orthogonal multiple access (NOMA) has been demonstrated as a promising solution [71]. One of the latest NOMA techniques is GF-NOMA, where users can communicate with the base station (BS) simultaneously and quickly on the same time-frequency resource block (RB) without the need for a demand-assigned access from the BS [98]. This access method can improve the spectrum access efficiency and reduce the transmission latency for the system. The application of NOMA to URLLC-enabled systems has also been considered in recent years [99–101] to further enhance the system performance.

GF transmission has been proposed for 5G new radio (NR) as a promising solution to reduce the latency in URLLC and massive access scenarios [71, 102]. In GF URLLC, a user can communicate with the base station in an arrive-and-go manner without the need to schedule the requests and uplink resource grants, thereby reducing the latency. However, the random nature of the GF access might lead to congestion, as multiple users could potentially access the same RB. The GF-NOMA can mitigate this issue by enabling many users to share the same RBs. However, because the GF access is random, a larger number of users can occupy one RB simultaneously, which may lead to severe interference in GF-NOMA systems and degrade the system performance. This demands an intelligent resource allocation approach for GF-NOMA networks to optimize the system performance.

Machine learning (ML), which is recognized as one of the potential technologies for the next generation wireless networks [51], could be an enabling solution to address the above problem. The underlying principle of ML is to learn from the observed data or surrounding environment in order to make optimal decisions in complex, dynamic, and uncertain large-scale environments. ML techniques including supervised learning [103], unsupervised learning, and reinforcement learning (RL) [104, 105], have been recently investigated in order to address various issues in wireless communication schemes such as channel estimation and signal detection, beamforming design, resource allocation, and system security.

Recently, the combination of NOMA and URLLC has been investigated in several works [99–101] to increase connectivity and guarantee the reliability and latency requirements for wireless networks. Specifically, these works considered multiple-input multiple-output (MIMO) and multiple-input single-output (MISO) schemes for URLLC-enabled systems to improve the system performance in terms of reliability and latency. The works proposed user-pairing methods based on the power-domain NOMA principle to enhance connectivity and reduce interference. However, the above works did not examine the GF access method, which can support massive access and reduce the transmission latency for wireless systems requiring high reliability and low latency.

Taking GF transmission into account, the works in [106, 107] studied GF access for OMA. In the GF-OMA scheme, users can select RBs randomly, and each RB is used strictly by a single user for successful reception. This limitation may lead to severe collisions when the number of users is much higher than the number of available RBs. To overcome this challenge, GF-NOMA has emerged as a promising technology for massive access by allowing multiple users to access the same RB based on the power-domain NOMA [71]. In particular, the users occupying the same RB are distinguished by different received power levels, and multi-user data can be decoded at the receivers by utilizing the successive interference cancellation (SIC). The traditional contention-based GF-NOMA schemes are implemented by dividing a cell area into multiple fractions and using the orthogonal resource allocation among those fractions to reduce the inter-fraction collisions [98, 108]. Nevertheless, the spectrum competition among users within the same fraction is still high, resulting in severe interference and reducing system performance. Thus, it is important to find a smart congestion control method to reduce the collisions and improve the long-term system performance.

Intelligent features are an important aspect of future cellular networks, and many current research works have applied RL-based algorithms to address the collisions and severe interference in massive access scenarios [29, 31, 109–118]. Specifically, Sharma *et al.* [109] proposed a collaborative distributed Q-learning algorithm for the frame-based slotted-Aloha (SA) random access (RA) scheme to find the best resource block allocation

strategy for IoT users, in order to avoid collisions in GF-OMA-based IoT systems. The authors in [110–113] investigated the application of Q-learning to different GF-NOMA scenarios with/without SPC to mitigate the congestion and interference in overloaded systems, where the number of users is larger than the number of available RBs. However, RL-based algorithms such as Q-learning are not suitable for large high-dimensional state-action spaces [51], making them inadequate for addressing the network optimization problems in complex and large-scale scenarios of future wireless networks.

To overcome the aforementioned challenges, recent studies have been applying deep RL (DRL) to address the complex resource allocation problems and optimize system performance [29, 31, 114–118]. In particular, the work in [114] proposed a DRL framework to find an optimal resource management strategy for GF-OMA systems and address dynamic spectrum access issues. In [29], a DRL algorithm based on generative adversarial networks was proposed to minimize power consumption while ensuring high reliability and low latency for orthogonal frequency division multiple access (OFDMA) systems. To further improve the spectral access efficiency and enhance the system performance, DRL-based GF-NOMA schemes were investigated in [31, 115–118] under different scenarios. Specifically, the work [115] investigated a pilot sequence-based GF-NOMA system and proposed a centralized training distributed execution multi-agent (MA) DRL (MADRL) solution to maximize the network throughput (number of successfully served users). Additionally, different MADRL-based dynamic resource allocation strategies for power-domain GF-NOMA systems were investigated in [116, 117] to maximize the system throughput [116] and sum rate [117]. In [31,118], DRL-based methods were proposed for GF-NOMA systems enabling massive URLLC (mURLLC) to maximize the long-term average throughput.

Unlike the aforementioned works on GF-NOMA systems, this chapter investigates an MADRL-based resource allocation strategy aimed at maximizing the energy efficiency (EE) while satisfying the users' requirements on reliability and latency for URLLC-enabled GF-NOMA (URLLC-GF-NOMA) systems. Given the stringent requirements of reliability and latency of URLLC users, there is a demand for an efficient and rapid communication protocol. Therefore, our focus is on constructing an effective distributed MADRL-based solution that achieves both EE and rapid convergence with minimal signaling overhead. The approach is designed to reduce the information exchange between the environment and agents, based on which the lower processing latency for URLLC users can be achieved. Indeed, we consider a GF-NOMA scenario where the users compete for the RBs, i.e., subchannels (SCs) and transmission power levels (TPLs), to communicate with the BS by randomly selecting one SC and one TPL for their transmissions. Following the NOMA principle, the users utilizing the same SC are distinguished by their received power at the BS, and their messages are decoded in an orderly manner using SIC [98]. However, with its random access nature, GF-NOMA may cause severe interference since too many users

can select the same SC, leading to the system performance degradation. To overcome this drawback, we utilize DRL techniques to enable the users to find the most suitable SCs and TPLs for their transmissions, optimizing the EE and fulfilling the URLLC requirements of all users. Thus, the main contributions of this chapter are summarized as follows:

1. Given that EE is an important factor due to users' energy limitations, we investigate the problem of maximizing the long-term average EE for URLLC-GF-NOMA systems. The goal must be achieved while also ensuring the strict requirements of users in terms of reliability and latency, which necessitates a rapid and efficient transmission protocol. Building on this EE maximization problem, we further investigate the objectives of maximizing the sum rate and minimizing power consumption to clarify the benefits of the proposed problem in balancing the achievable sum rate against power consumption for energy-limited users.

2. We develop three distributed MADRL-based resource allocation methods to address the considered problem: MA Dueling Double Deep Q Network (MA3DQN), MA Double Deep Q Network (MA2DQN), and MA Deep Q Network (MADQN). Within this context, the MADRL frameworks are designed to provide energy-efficient learning-based solutions which ensure rapid convergence and minimal signaling overhead, ultimately reducing the processing latency for URLLC users.

3. We provide a performance comparison between the proposed mechanisms and other benchmark schemes to clarify the benefits of the former in terms of convergence property and EE performance. Additionally, we evaluate the effects of different state-action spaces, URLLC requirements, and the number of users on the achieved rewards and EE performance. The provided numerical results prove that the proposed solutions outperform other benchmark schemes, achieving higher EE, faster convergence, and reduced signaling overhead.

## 4.2 System Model

We consider an uplink URLLC-GF-NOMA system consisting of one base station (BS) and a set of $M$ URLLC users, denoted by $\mathcal{M}$, allocated uniformly around the BS within a circle-cell radius of $r_c$ (m), as shown in Fig. 5.1. The system bandwidth is equally divided into a set of $K$ orthogonal SCs, denoted by $\mathcal{K}$, to serve the users. Moreover, the GF-NOMA transmission strategy is utilized to improve the spectrum access efficiency and guarantee strict requirements of the URLLC users in overloaded scenarios, i.e., $M > K$. Following this transmission scheme, the users utilize the available SCs to communicate with the BS, and multiple users can share the same SC based on the power-domain NOMA principle [71].

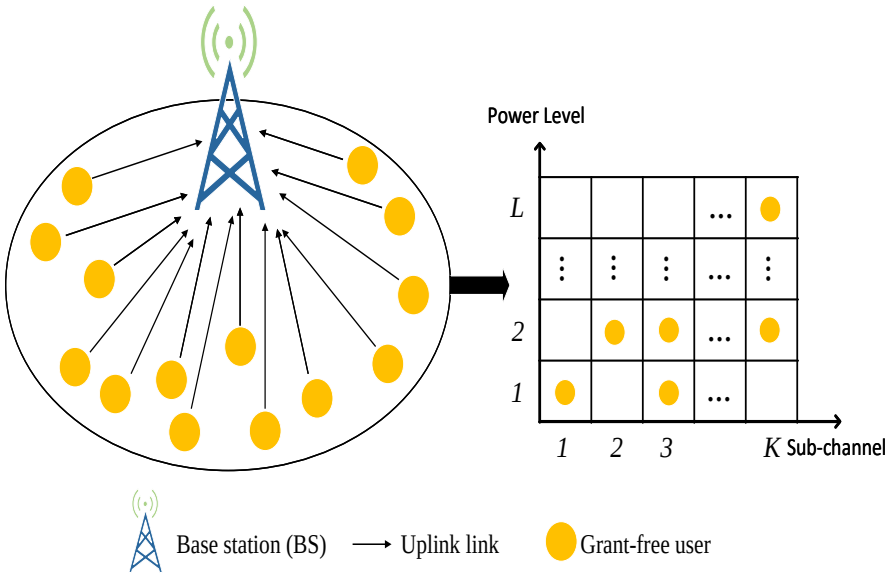



FIGURE 4.1: Illustration of an uplink URLLC-GF-NOMA system.

In 5G new radio (5G-NR) networks, the SC's bandwidth is defined as $2^{\nu}$ times of SC's bandwidth in 4G systems (i.e., 180 kHz), where $\nu \in \{0, 1, 2, 3, 4\}$ denotes the numerology index which stands for the various SC types in order to support different services [119,120]. In particular, the SC with higher bandwidth is used for URLLC service while other services such as eMBB and mMTC can utilize the numerology with smaller SC spacing. Given this context, this chapter considers that the total bandwidth is divided into a set of SCs, i.e., $\mathcal{K}$, serving the URLLC users, and the bandwidth of SCs is defined as $W = 2^{\nu} \times 180$ (kHz).

### 4.2.1 Uplink GF-NOMA Transmission Process

Under the GF strategy, the users are free to choose the SCs for their transmissions without any scheduling instructions from the BS. However, this can lead to severe collision issues as too many users may select the same SCs. To mitigate this drawback, the NOMA technique can be applied, where multiple users can access the same SC. Considering the NOMA transmission process over SC $k$ ($k \in \mathcal{K}$) in time slot (TS) $t$, we denote $x_m^{(k)}(t)$ as a binary SC allocation variable, where $x_m^{(k)}(t) = 1$ if user $m$ occupies SC $k$ and $x_m^{(t)} = 0$ otherwise. The set of users occupying SC $k$ in TS $t$ is described as $\mathcal{M}^{(k)}(t) = \{m | x_m^{(k)}(t) = 1, m \in \mathcal{M}\}$. Let $M_k$ be the number of users using SC $k$ in TS $t$, i.e., $\sum_{k=1}^{K} M_k = M$. Then, the received

signal at the BS over SC $k$ in TS $t$ is given by

$$y^{(k)}(t) = \sum_{m=1}^{M_k} \sqrt{P_m^{(k)}(t)} h_m^{(k)}(t) u_m^{(k)}(t) + n(t), \tag{4.1}$$

where $n(t) \sim \mathcal{CN}\left(0, \sigma^2\right)$ is the additive white Gaussian noise (AWGN), $P_m^{(k)}(t)$ and $u_m^{(k)}(t)$ denote the transmission power and the transmitted message of user $m$ over SC $k$ in TS $t$, respectively. Herein, the transmission power is defined as $P_m^{(k)}(t) = 0$ if $x_m^{(k}(t) = 0$, otherwise, $P_m^{(k)}(t) \neq 0$. Besides, $h_m^{(k)}(t)$ represents the channel coefficient between user $m$ and the BS over SC $k$ in TS $t$.

We assume that the users using SC $k$ are sorted in the descending order of the corresponding received power level at the BS, i.e.,

$$\mathcal{P}_1^{(k)}(t) \geq \cdots \geq \mathcal{P}_{M_k}^{(k)}(t), \tag{4.2}$$

where $\mathcal{P}_m^{(k)}(t) = P_m^{(k)}(t) \left|h_m^{(k)}(t)\right|^2$. Following the NOMA principle, the messages of the users with higher received power level are decoded earlier at the BS. Specifically, the BS decodes the message of a user by treating the messages of users with lower received power level as noise [101,121]. It then reconstructs and removes this component from the received signal to decode the remaining users' messages successively by using the SIC technique. Accordingly, the received signal-to-interference-plus-noise ratio (SINR) of user $m$ over SC $k$ in TS $t$ is expressed as

$$\gamma_m^{(k)}(t) = \frac{\mathcal{P}_m^{(k)}(t)}{\sum\limits_{i=m+1}^{M_k} \mathcal{P}_i^{(k)}(t) + \sigma^2}. \tag{4.3}$$

### 4.2.2 URLLC Communication Model

Due to the stringent low-latency requirement of URLLC communication, very short packets and finite blocklength (FBL) is implemented for data transmission, so-called SPC. Consequently, the Shannon-related capacity formula cannot be applied to the URLLC communication model since it is designed under the assumption of the infinite blocklength (iFBL). According to [9], the achievable rate of user $m$ over SC $k$ in the FBL regime for a quasi-static flat fading channel can be approximated as

$$R_m^{(k)}(t) \approx W \left[\log_2\left(1 + \gamma_m^{(k)}(t)\right) - \sqrt{\frac{v_m^{(k)}(t)}{\tau W}} Q^{-1}\left(\varepsilon_m\right)\right], \tag{4.4}$$

where $v_m^{(k)}(t) = 1 - \frac{1}{\left(1+\gamma_m^{(k)}(t)\right)^2}$ is the channel dispersion, $\tau$ denotes the transmission latency threshold, $\varepsilon_m$ is the decoding error probability, and $Q^{-1}(x)$ represents the inverse of the Gaussian Q-function which is defined as

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \tag{4.5}$$

Based on (4.4), one can define an SNR threshold for user $m$ trying to transmit one packet over one SC $k$ in each transmission TS that satisfies the URLLC requirements (i.e., $\tau$ and $\varepsilon_m$) as [26]

$$\hat{\gamma}_m = 2^{\frac{n_b}{\tau W} + \frac{Q^{-1}(\varepsilon_m)}{\ln 2\sqrt{\tau W}}} - 1, \tag{4.6}$$

where $n_b$ (bits) is the packet size. From (4.6), the target rate for the transmission of user $m$ can be defined as

$$\hat{R}_m \approx W \left[ \log_2 \left(1 + \hat{\gamma}_m\right) - \sqrt{\frac{\hat{v}_m}{\tau W}} Q^{-1}\left(\varepsilon_m\right) \right], \tag{4.7}$$

where $\hat{v}_m = 1 - \frac{1}{(1+\hat{\gamma}_m)^2}$. Similar to [26, 116], we assume that each user $m$ can transmit its packet only once. As the interference over an SC increases, the likelihood of packet drops escalates. Specifically, a successful transmission occurs if $R_m^{(k)}(t) \geq \hat{R}_m$; otherwise, any deviation from this condition results in a failed transmission, i.e., a dropped packet.

### 4.2.3 Energy Efficiency Maximization

Energy efficiency (EE) is considered one of the major goals in 5G and beyond wireless networks [122]. Furthermore, the majority of mobile devices operate on limited battery power [122], resulting in the need to design energy-efficient communication methods. To address this concern, we first define an EE factor with the purpose of ensuring the achievable rate requirement while reducing the power consumption for the system as follows:

$$\mathcal{E}(t) = \frac{\sum_{k=1}^{K} \sum_{m=1}^{M_k} x_m^{(k)}(t) R_m^{(k)}(t)}{MP_c + \sum_{k=1}^{K} \sum_{m=1}^{M_k} P_m^{(k)}(t)}, \tag{4.8}$$

where $P_c$ denotes the circuit power consumption. In what follows, the work focuses on designing an effective distributed power control and SC assignment strategy for URLLC-GF-NOMA systems to maximize the average EE while ensuring the URLLC requirements of all users. This can have a direct impact on the overall sustainability and cost-effectiveness

of the considered networks. The design objective can be cast by the following problem:

$$\max_{\mathbf{x},\mathbf{P}} \quad \mathbb{E}_t\left[\mathcal{E}(t)\right] \tag{4.9a}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} x_m^{(k)}(t) R_m^{(k)}(t) \geq \hat{R}_m, \quad \forall m, \tag{4.9b}$$

$$\mathcal{P}_1^{(k)}(t) \geq \mathcal{P}_2^{(k)}(t) \geq ... \geq \mathcal{P}_{M_k}^{(k)}(t), \quad \forall k, \tag{4.9c}$$

$$\sum_{k=1}^{K} x_m^{(k)}(t) \leq 1, \quad \forall m, \tag{4.9d}$$

$$\sum_{k=1}^{K} P_m^{(k)}(t) \leq P_{\max}, \quad \forall m, \tag{4.9e}$$

where $\mathbb{E}_t[\cdot]$ is the expectation operation over TSs, $\mathbf{x}$ and $\mathbf{P}$ denote the SC assignment and power control strategies, respectively. The constraint (4.9b) represents the rate condition to guarantee the users' URLLC requirements. The constraint (4.9c) ensures the NOMA-based multi-user decoding process. The constraint (4.9d) implies that each user selects at most one SC. The constraint (4.9e) shows the users' power budget.

**Remark 6.** *It is noteworthy that the EE maximization problem defined in* (4.9) *can also include the objectives of maximizing the sum rate and minimizing the power consumption. These objectives can be attained by setting the denominator and numerator as* 1*, respectively. Thus, the considered scenario represents a general case where an efficient solution, striking the trade-off between the achievable sum rate and power consumption, can be achieved. Further evaluation on this matter is provided in Section 4.4.*

## 4.3 MADRL-Based Energy Efficiency Resource Allocation Solution For URLLC-GF-NOMA Systems

The problem described in (4.9) is challenging to solve due to its non-convex nature and NP-hard complexity. Moreover, with the GF access method, the users can select their preferred SC and transmission power independently in each TS without requiring admission approval from the BS. While this feature can reduce the access latency and increase the connectivity density, it also necessitates a decentralized optimization solution. Therefore, to effectively address the problem stated in (4.9), we consider an MADRL-based method, which can be implemented in a distributed manner.

### 4.3.1 MADRL Framework

RL is one of the machine learning methods that enable a learning agent to achieve its specific goal with the best long-term reward by interacting with the environment in a trial-and-error manner [117]. In particular, an agent interacts with the environment by taking an action selected from its action space at the current state. It then receives a respective reward and moves to a new state. These procedures are repeated until convergence is observed, where the learning policy of the agent achieves an optimal value in terms of average reward. This learning process can be formulated as a Markov decision process (MDP) with a tuple of four elements $(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P})$, defined as follows:

- $\mathcal{S}$: The set of states in the environment, where $s(t) \in \mathcal{S}$ denotes the state of an agent at TS $t$.
- $\mathcal{A}$: The set of actions that an agent can take, where $a(t) \in \mathcal{A}$ is the action of an agent at TS $t$.
- $\mathcal{R}$: The reward function, where $r(t)$ represents the immediate reward of the agent at TS $t$ by performing action $a(t)$ in state $s(t)$.
- $\mathcal{P}$: The probability distribution function of the state transition, where $\mathcal{P}(s(t), s(t+1))$ denotes the state transition probability from state $s(t)$ to state $s(t+1)$.

In the considered URLLC-GF-NOMA system, the behavior of all users (i.e., transmission power and SC selection) can be modeled as an MA MDP (MAMDP), which is denoted by $(\{\mathcal{S}\}_{m=1}^{M}, \{\mathcal{A}\}_{m=1}^{M}, \mathcal{R}, \mathcal{P})$. Unlike a single-agent DRL related to the learning process of only one single agent, our proposed MADRL-based model involves a set of agents $\mathcal{M}$, where all agents operate autonomously and concurrently in a sharing environment. In particular, each agent $m$ observes its current state $s_m(t) \in \mathcal{S}_m$ from the environment and performs an action $a_m(t)$ chosen from its own action space $\mathcal{A}_m$. The joint action of all agents can be formulated as $a(t) = \{a_1(t), a_2(t), \ldots, a_M(t)\}$. The agent $m$ then moves from the current state $s_m(t)$ to a new state $s_m(t+1)$. All agents then receive a reward of $r(t+1)$ and perform an update of their current policy according to the feedback from the environment. It is worth noting that each agent having a distinct reward may result in selfish behavior, leading to a reduction of the global network performance [123]. Therefore, we assume that all agents have a common reward to obtain the global optimum. The main elements of the proposed MADRL approach are defined as follows:

- *State:* Due to users' independence and URLLC requirements, the state of agent (user) $m \in \mathcal{M}$ is designed only based on the local information available at this agent to reduce the processing latency and the signaling overhead in information exchange between the agent and environment. Specifically, the state of agent $m$ in TS $t$ can

be defined as the combination of SC index and transmission power value it selected in the previous TS $t-1$, which is expressed as

$$s(t) = \left\{ k_m(t-1), P_m^{k_m(t-1)}(t-1) \right\}, \tag{4.10}$$

where $k_m(t-1)$ and $P_m^{k_m(t-1)}(t-1)$ are the selected SC index and transmission power of agent $m$. Since the users' selection of SC and transmission power will impact the overall EE, it is reasonable to include this information in the defined state. From (4.10), the state of agent $m$ has a cardinality of 2. It is noteworthy that the state definition in (4.10) differs from those in recent related works on GF-NOMA systems, which require a large signaling overhead in information exchange between the environment and the agents during the learning process [116,117]. A performance comparison between different state definitions will be provided in Section 4.4.

- *Action:* At the beginning of TS $t$, agent $m$ selects an SC and transmission power for its transmission. As a feasible solution, the discrete power domain has been widely used for the learning-based GF-NOMA systems in the literature [110, 115, 117]. This approach can ensure stable convergence and reduce the computational complexity of the distributed learning models conducted by the users who have limited computational resources. Given this context, we consider a discrete action space, where the power is quantized into $L$ levels which are determined as $\hat{P}_l = lP_{max}/L$, $l \in \{1, 2, \ldots, L\}$, where $\hat{P}_l$ is the $l$-th TPL. Thus, the action of user $m$ in TS $t$ is defined as

$$a_m(t) \in \mathcal{A}_m = \{1, \ldots, kl, \ldots, KL\}, \tag{4.11}$$

where $a_m(t) = kl$ indicates that agent $m$ selects SC $k$ and TPL $l$ in TS $t$. Thus, the action space size of agent $m$ is $KL$ and the overall action space size of all agents is determined as $(KL)^M$.

- *Reward:* After all agents take their chosen actions, they receive an immediate reward from the environment reflecting if their transmissions are successful or not, i.e., if all constraints in the problem (4.9) are satisfied or not. In the MADRL frameworks, both centralized and decentralized rewards can be considered to build learning models. The centralized-reward mechanism yields a common reward to all agents, whereas in decentralized-reward schemes, each agent receives a distinct reward. However, the decentralized-reward strategy can lead to selfish behavior among agents. They may compete with others to maximize their own rewards, which potentially results in a degradation of overall system performance. To circumvent this issue, a common reward can be implemented to align the agents towards a shared global objective [123]. Since the objective is to maximize the network EE, we use the achieved EE to formulate the reward function (RF). Furthermore, all agents receive the same reward

with the aim of achieving the common objective, i.e., optimizing the network EE and guaranteeing URLLC requirements of all users. Thus, the RF is defined as

$$r(t) = \begin{cases} \mathcal{E}(t), & \text{if all constraints in the} \\ & \text{problem (4.9) are satisfied,} \\ 0, & \text{otherwise.} \end{cases} \tag{4.12}$$

Based on the reward function defined in (4.12), it becomes apparent that inappropriate user actions, such as an excessive number of users choosing the same SC, may degrade the system's EE. Consequently, the users will receive a low reward. Throughout the learning process, users explore the environment to find the best policies that will maximize their reward, ultimately leading to optimal EE performance.

The objective of RL algorithms is to find a policy $\pi$ to maximize the expected reward [56]. Considering the Q-learning algorithm - a popular RL technique, the expected reward achieved by agent $m$ after taking action $a_m$ in state $s_m$ following a policy $\pi$ can be determined based on the action-value function (or Q-value function) as

$$Q_\pi(s_m, a_m) = \mathbb{E}_\pi \left[\hat{r}(t) | s_m(t) = s_m, a_m(t) = a_m\right], \tag{4.13}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator and $\hat{r}(t)$ is the long-term discounted cumulative reward which is given by

$$\hat{r}(t) = \sum_{k=0}^{\infty} \gamma^k r(t+k+1), \tag{4.14}$$

where $\gamma$ is the discount factor that determines the weight of the future reward. Based on (4.13), the optimal Q-function can be calculated as

$$Q^*(s_m, a_m) = \max_\pi Q_\pi(s_m, a_m). \tag{4.15}$$

Through the Q-learning method, the optimal policy can be found based on the available information $(s_m(t), a_m(t), r(t), s_m(t+1))$. The update equation of the Q-value function of agent $m$ can be expressed as [56]

$$Q(s_m(t), a_m(t)) = Q(s_m(t), a_m(t)) + \alpha \left[y_m(t) - Q(s_m(t), a_m(t))\right], \tag{4.16}$$

where $y_m(t) = r(t) + \gamma \max_a Q(s_m(t+1), a)$ and $\alpha \in [0, 1]$ is the learning rate.

Although the Q-learning method has been widely adopted in wireless networks for resource management purposes, it only works well under small state-action spaces, which limits its applicability. Its practicality diminishes as the problem size increases, primarily due to two key factors [117]: (i) the need for a lookup table to store Q-values for every possible state-action pair becomes unmanageable in terms of storage complexity when

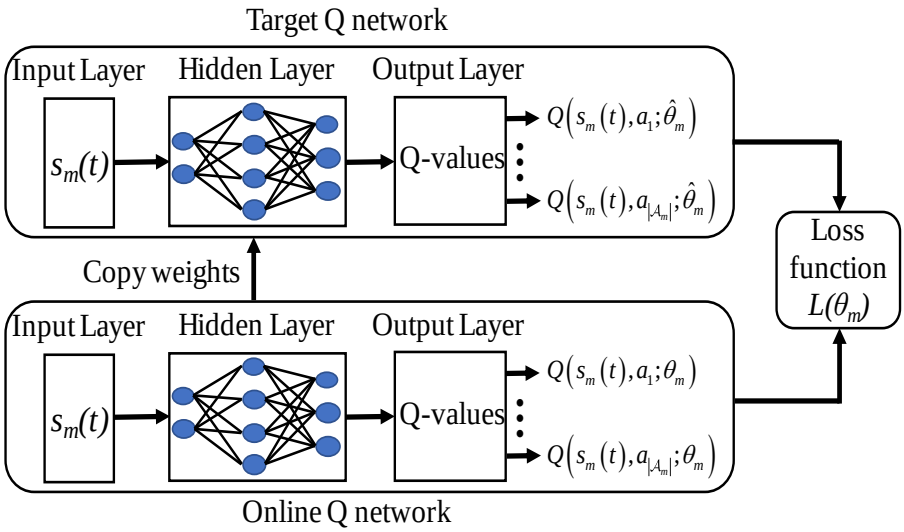



FIGURE 4.2: Illustration of DQN/2DQN model.

dealing with large-scale problems; and (ii) with a larger state space, many states are rarely visited, resulting in decreased performance. To overcome this drawback, we consider DRL techniques to efficiently solve the proposed problem in (4.9). In the DRL method, a deep neural network (DNN) is integrated into the framework of Q-learning to reduce the memory size and computational complexity by calibrating and training the DNN's different layers to define the best action for each state instead of using a large storage space (i.e., Q-table) to store all Q-values [124]. In this chapter, we propose MADRL-based EE URLLC-GF-NOMA methods, where different DRL techniques including deep Q network (DQN), double DQN (2DQN), and dueling 2DQN (3DQN), are investigated[1].

### 4.3.2 Proposed MADRL Algorithms For URLLC-GF-NOMA Systems

#### MADQN-Based Approach

In this section, we consider a MADQN-based URLLC-GF-NOMA approach. With this method, each agent constructs its own DQN model that consists of two different DNNs: the online and target networks, as depicted in Fig. 4.2. Specifically, in each TS $t$, agent $m$ uses the online network for Q-function approximation $Q\left(s_m(t), a_m(t); \theta_m\right)$ to select an action $a_m(t) \in \mathcal{A}_m$ at state $s_m(t) \in \mathcal{S}_m$. Here, $\theta_m$ represents the parameters (weights)

[1]Besides DRL algorithms based on Q-learning and DNN, tile coding and on-policy learning could also be promising methods to achieve an effective solution and analytical convergence. This would be a noteworthy issue to investigate in future work.

of the agent $m$'s online network. Meanwhile, the target network is used to stabilize the learning process, and its parameters $\hat{\theta}_m$ are updated by copying the parameters $\theta_m$ of the online network after a certain number of TSs, which is also known as the parameter update frequency $F$.

Regarding the action selection at each state, one should consider the trade-off between exploration and exploitation during the learning process to achieve the optimal policy. Given this context, the $\epsilon$-greedy policy can be used for action selection to obtain a balance between the exploitation of the best Q-value function and the environmental exploration [56]. In particular, the $\epsilon$-greedy policy selects an action based on two conditions:

$$a_m(t) = \begin{cases} \text{random action}, & \text{with probability } \varepsilon \\ \underset{a \in \mathcal{A}_m}{\arg\max} \{\mathcal{Q}_m(t)\}, & \text{with probability } 1-\varepsilon \end{cases}, \tag{4.17}$$

where $\mathcal{Q}_m(t) = Q(s_m(t), a; \theta_m)$. Herein, the parameter $\epsilon$ determines the level of exploration, and it is usually set to decrease over time to reduce the exploration rate as the learning progresses.

During the learning process, MADQN approach uses the experience replay strategy to achieve learning stability, where the transition in the form of a tuple $(s_m(t), a_m(t), r(t), s_m(t+1))$ is stored in the experience replay memory of each agent $m$. At each iteration, a mini-batch of experiences is sampled uniformly to train the learning model and update the parameters of the online network $\theta_m$ with the purpose of minimizing the loss function defined as

$$L_m(\theta_m) = [y_m(t) - Q(s_m(t), a_m(t); \theta_m)]^2, \tag{4.18}$$

where $y_m(t)$ is the target value calculated from the target network as follows:

$$y_m(t) = r(t) + \gamma \max_{a \in \mathcal{A}_m} Q\left(s_m(t+1), a; \hat{\theta}_m\right). \tag{4.19}$$

Given the DQN model of each agent mentioned above, the proposed MADQN-based URLLC-GF-NOMA approach is summarized in Algorithm 2. In particular, in TS $t$, each agent $m$ observes its current state $s_m(t) \in \mathcal{S}_m$ and takes an independently action $a_m(t) \in \mathcal{A}_m$ selected based on the $\epsilon$-greedy policy in (5.37). After performing the chosen action, agent $m$ receives a common reward $r(t)$ based on the achieved EE and moves to a new state $s_m(t+1)$. It then stores an experience tuple of $(s_m(t), a_m(t), r(t), s_m(t+1))$ into its experience replay memory, and a minibatch of experiences is sampled for training the online network. The parameters of the online network $\theta_m$ are then updated to minimize the loss function in (4.18) by using the stochastic gradient method, where the target value is given by (4.19). After a predetermined number of TSs, the parameters of the target network $\hat{\theta}_m$ are updated by copying $\theta_m$. The above training process continues until reaching a predefined number of episodes guaranteeing the algorithm's convergence.

**Algorithm 2** MADRL-based Energy Efficiency Optimization Algorithm for URLLC-GF-NOMA Systems.

Initialize online Q network with random parameters $\theta_m$, $\forall m \in \mathcal{M}$.
Initialize target Q network with parameters $\hat{\theta}_m = \theta_m$, $\forall m \in \mathcal{M}$.
**for** $e = 1, 2, \ldots, E$ **do**
    Initialize the network state $s_m(t)$, $\forall m$.
    **for** $t = 1, 2, \ldots, T$ **do**
        All agents select their actions $a_m(t) \in \mathcal{A}_m$, $\forall m$, based on the $\epsilon$-greed policy in (5.37).
        All agents take their actions, receive a common reward $r(t)$, and move to the next state $s_m(t+1)$.
        **for** $m = 1, 2, \ldots, M$ **do**
            Store an experience tuple of $(s_m(t), a_m(t), r(t), s_m(t+1))$ to the replay memory of agent $m$.
            Randomly sample a mini-batch of experience from the replay memory for training.
            Determine the loss function $L(\theta_m)$ as follows:
            **- MADQN approach:** Using (4.18) and (4.19).
            **- MA2DQN approach:** Using (4.18) and (4.20).
            **- MA3DQN approach:** Using (4.18) and (4.20), where the Q-value (action-value) functions are calculated by utilizing (4.21).
            Update $\theta_m$ by using stochastic gradient to minimize $L(\theta_m)$.
            Update $\hat{\theta}_m$ as $\hat{\theta}_m = \theta_m$ after every $F$ TSs.
        **end for**
    **end for**
**end for**

### MA2DQN-Based Approach

From (4.19), one can observe that the MADQN approach based on DQN model using the same Q-value function for both tasks, i.e., action selection, $\max_{a \in \mathcal{A}_m} Q(s_m(t+1), a; \hat{\theta}_m)$, and action estimation, $Q(s_m(t+1), a; \hat{\theta}_m)$. This can lead to an unstable learning process since the Q-value function is estimated over-optimistically. To mitigate this issue, we investigate an MA2DQN-based URLLC-GF-NOMA approach, where 2DQN model is considered [125], as shown in Fig. 4.2. In this method, the action selection and evaluation are decoupled to avoid the overestimation issue by replacing the target value in (4.19) with the following one

$$y_m(t) = r(t) + \gamma Q\left(s_m(t+1), \arg\max_{a \in \mathcal{A}_m} \mathcal{Q}_m(t+1); \hat{\theta}_m\right), \tag{4.20}$$

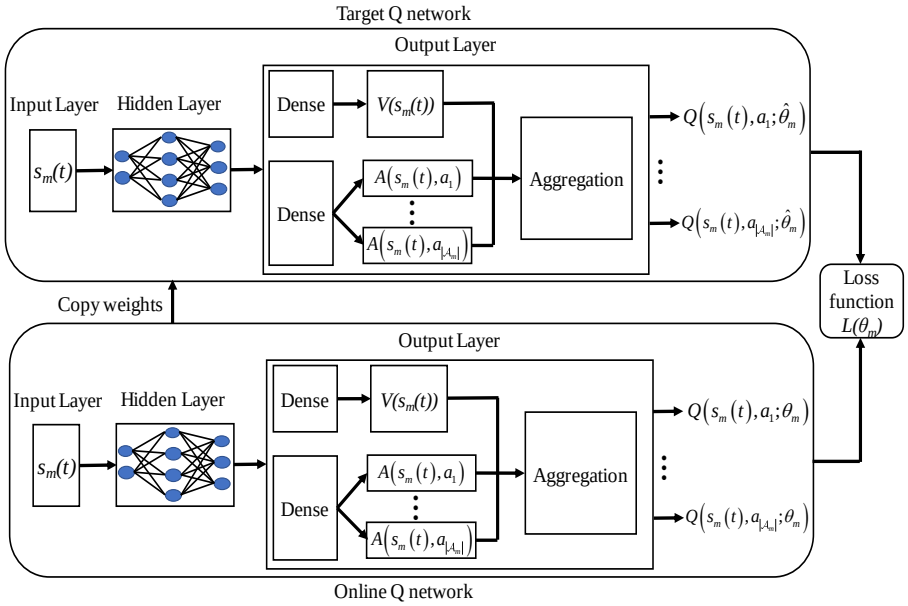



FIGURE 4.3: Illustration of 3DQN model.

where $\mathcal{Q}_m(t+1) = Q\left(s_m\left(t+1\right), a; \theta_m\right)$. As can be seen from (4.20) that the online network $Q\left(s, a; \theta_m\right)$ is used for the action selection, whereas the target network $Q\left(s, a; \hat{\theta}_m\right)$ is applied to estimate the action. The MA2DQN-based URLLC-GF-NOMA algorithm is also summarized in Algorithm 2 with **MA2DQN** remark in **Step 11**.

### MA3DQN-Based Approach

An MA3DQN-based URLLC-GF-NOMA approach is studied in this section. This method uses a 3DQN model whose structure is depicted in Fig. 4.3, to speed up the convergence and improve the learning efficiency [126]. Following MA3DQN approach, each agent $m$ creates its own 3DQN model based on 2DQN, where the last layer of the 2DQN model is split into two parts to evaluate the state value function (SVF) $V(s_m(t))$ and the advantage function (AF) $A(s_m(t), a_m(t))$. Herein, the SVF $V(s_m)$ is used for estimating the quality (goodness or badness) of a given state $s_m(t)$, allowing the agent to evaluate the long-term potential of being in that state. Meanwhile, the AF $A(s_m(t), a_m(t))$ captures how much better or worse a specific action is compared to other actions in state $s_m(t)$. This allows the agent to choose the best action to take in a given state. The two parts are then combined to produce the final action-value function $Q(s_m(t), a_m(t); \theta_m, \theta_m^V, \theta_m^A)$ that is used to select actions in the environment. Here, $\theta_m^V$ and $\theta_m^A$ denote the parameters according to SVF-related and AF-related parts, respectively. Given this context, the action-value function

determined by agent $m$ for a given state $s_m(t)$ and action $a_m(t)$ is calculated as follows:

$$Q(s_m(t), a_m(t); \theta_m, \theta_m^V, \theta_m^A) = V(s_m(t)) + A(s_m(t), a_m(t)) - \frac{1}{|\mathcal{A}_m|} \sum_{a \in \mathcal{A}_m} A(s_m(t), a_m(t)), \tag{4.21}$$

where the last term of the right-hand side of (4.21) is the mean of the AF over all actions. It is subtracted from the AF $A(s_m(t), a_m(t))$ of a specific action to ensure that the AF is centered around zero, making it easier to train the network. This approach improves the convergence and stability of the network and enables the effective separation of the estimation of SVF and AF, resulting in better performance compared to DQN and 2DQN architectures. The MA3DQN-based URLLC-GF-NOMA approach is also cast by Algorithm 2 under the designation **MA3DQN** mentioned in **Step 11**.

### 4.3.3 Analysis of The Proposed Methods

#### Complexity Analysis

Let $H$, $N_h$, and $I_s$ be the number of training layers (input, hidden, and output layers), the number of neurons in layer $h$, and the size of the input layer. For each TS, the computational complexity of URLLC-GF-NOMA algorithms based on MADQN and MA2DQN can be calculated by

$$C_{\mathsf{TS}} = \mathcal{O}(X), \tag{4.22}$$

where $X = I_s N_1 + \sum_{h=1}^{H-1} N_h N_{h+1}$. For the training phase with $M$ agents, $E$ episodes, and $T$ TSs, the computational complexities of the algorithms can be given by

$$C_{\mathsf{MADQN}} = C_{\mathsf{MA2DQN}} = MET \times C_{\mathsf{TS}} = \mathcal{O}(METX). \tag{4.23}$$

Taking the MA3DQN-based URLLC-GF-NOMA algorithm into account, it has higher complexity than MADQN and MA2DQN-based algorithms due to the implementation of the dueling network architecture. Specifically, its complexity can be determined as

$$C_{\mathsf{MA3DQN}} = \mathcal{O}(MET(X + N_{H-1})). \tag{4.24}$$

#### Convergence Analysis

The convergence of a multi-agent system relies on whether the combined strategy of the agents ultimately approaches the optimal state (Nash equilibrium), ensuring the stability of the solution. In this chapter, we propose URLLC-GF-NOMA methods based on

MADQN, MA2DQN, and MA3DQN, which combine the conventional Q-learning and neural networks. To analyze the convergence of these methods, two key aspects need to be addressed [127]: (i) demonstrating the ability of the conventional Q-learning to converge to the optimal state, and (ii) verifying that the neural network approach effectively identifies or approximates the nonlinear Q-values generated by the general Q-learning iteration as depicted in equation (4.16). In particular, it has been shown in [128] that the conventional Q-learning algorithm guarantees the attainment of the optimal state when the learning rate $\alpha_t$ satisfies $0 \leq \alpha_t \leq 1$, $\sum_t \alpha_t = \infty$, and $\sum_t \alpha_t^2 < \infty$. Additionally, based on [129], it is established that the neural network can approximate any nonlinear continuous function when adequately sized and suitably initialized. Thus, the convergence of our proposed methods can be guaranteed. It is noteworthy that as mentioned in [130], the theoretical analysis of the neural network's size and initial conditions for ensuring its convergence before training poses challenges due to the complex quantitative relationship between the network convergence and hyperparameters. Therefore, we utilize simulations to demonstrate the convergence of our proposed methods.

**Solution Analysis**

To clarify the difference between the scenario considered in this chapter and the ones investigated in related works on RL-based GF-NOMA [31, 115–117], this section provides a solution summary examined in these works, as shown in Table 4.1. As can be seen from this table, different DRL frameworks have been proposed to address the unique problems of GF-NOMA systems effectively. In delay-sensitive RL-based systems, signaling overhead is a key performance indicator. It is defined as the number of information bits needed to feed back the channel status data, SC indicators, and the transmission power of a specific user over an SC [131]. Also, the total number of users and SCs, and the exchange of states as well as rewards between the agents and environment can affect the signaling overhead. Higher signaling overhead results in larger processing latency for users.

Following [131], it is assumed that transmitting a continuous value of channel status, data rate, and reward requires 16 bits. Additionally, 1 bit is allocated for acknowledgment (ACK) feedback, 2 bits for decoding status, and 4 bits for the SC indicator, transmission power, and other relevant parameters. The work [115] produces a large signaling overhead because it depends on the decoding status of $\hat{K}$ pilot sequences, users' average throughput, and parameters (weights) of the centralized-training MADRL model transmitted from the BS to users who build local DRL models for distributed execution. These parameters depend on the number of input, hidden, and output layers ($A$) and the number of neurons per layer ($N_a$, $1 \leq a \leq A$). In addition, large signaling overhead can be observed in [116, 117] due to the inclusion of various feedback information. This includes the channel status and ACK information of each user [116], as well as users' data rate [117]. In [31], the BS decides the actions for users (the selection of repetition value and contention transmission

TABLE 4.1: Solution Summary of Related Works

| Refs. | Opt. Problem | Solution | State | Action | Reward | Signalling Overhead |
|---|---|---|---|---|---|---|
| [115] | Throughput | Centralized-training distributed execution MADRL | Decoding states, and average throughput | Pilot sequence | Throughput | $\underbrace{2\hat{K}+M}_{\text{State}} + \underbrace{\sum_{a=1}^{A} 4N_a}_{\text{Parameters}}$ |
| [116] | Throughput | Distributed MADRL | User's action, CSI, and ACK | SC and TPL | Throughput | $\underbrace{16KM+M}_{\text{State}} + \underbrace{4}_{\text{Reward}}$ |
| [117] | Sum rate | Distributed MADRL | Users' achievable rate | SC and TPL | Sum rate | $\underbrace{16KM}_{\text{State}} + \underbrace{16}_{\text{Reward}}$ |
| [31] | Throughput | Centralized MADRL | $V_{cc}$, $V_{ic}$, $V_{sc}$, $V_{sd}$, $V_{ud}$ | Repetition value and CTU | Throughput | $\underbrace{8M}_{\text{Action}}$ |
| Our Paper | Energy efficiency | Distributed MADRL | User's selected SC index and TPL | SC and TPL | Energy efficiency | $\underbrace{16}_{\text{Reward}}$ |

unit (CTU)), hence, the signaling overhead depends on the feedback information from the BS to the users regarding the selected actions for the transmission of each user. Note that $V_{cc}$, $V_{ic}$, $V_{sc}$, $V_{sd}$, and $V_{ud}$ used in Table 4.1 stand for the number of collision CTUs, idle CTUs, singleton CTUs, successfully served users, and failure decoding users, respectively. In our method, only the reward feedback is required to reduce the signaling overhead, but still guarantee an effective learning solution. Consequently, the signaling overhead is determined by the reward feedback.

## 4.4 Simulation Results

In this section, the simulation results are provided to evaluate the performance of the proposed MADRL-based resource allocation methods for the considered URLLC-GF-NOMA system. The simulations were performed on an Intel core i7-8665U CPU with 1.9 GHz frequency, 16 GB of random access memory (RAM), and 64-bit Windows 10 operating system. The learning models were considered with three hidden layers, including 256, 128, and 64 neurons. The experimental parameters are provided in Tables 4.2. Besides the proposed URLLC-GF-NOMA approaches based on MADQN, MA2DQN, and MA3DQN, we further investigate the following methods for comparison purpose.

- *MA Q-learning (MAQL) [110]*: MAQL is applied for GF-NOMA systems in [110]. With this scheme, each agent builds its own Q-table to store Q-values of all possible state-action combinations during learning process.

TABLE 4.2: Simulation Parameters

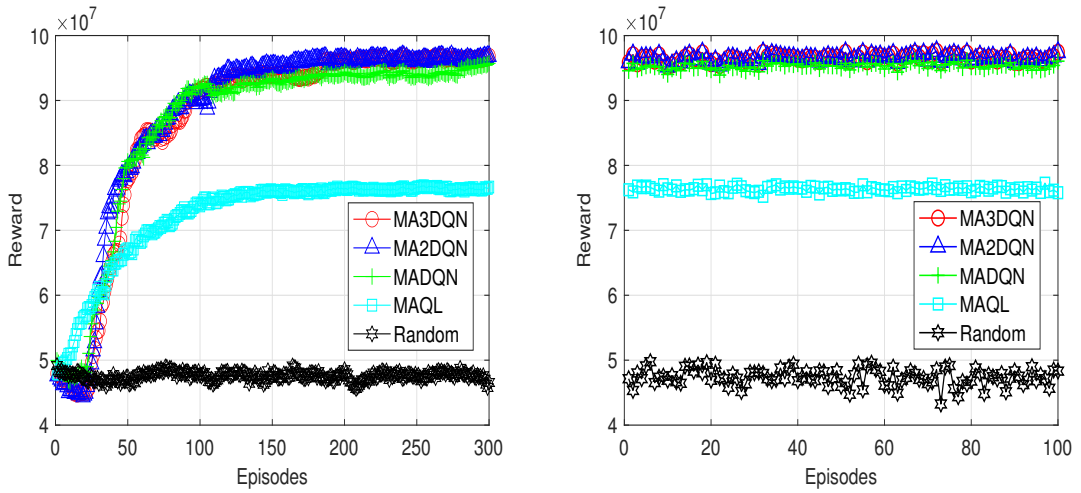
| **Parameters** | **Value** |
|---|---|
| Cell radius ($r_c$) | 500 m |
| Channel model | Rayleigh |
| Number of users ($M$) | $\{2; 4; 6; 8; 10\}$ |
| Number of SCs ($K$) | $\{2;\ 3\}$ |
| Reliability requirement ($\varepsilon_m$) | $\{10^{-1}, 10^{-3}, 10^{-5}, 10^{-7}\}$ |
| Latency threshold ($\tau$) | $\{0.5; 1; 1.5; 2\}$ ms |
| Numerology index ($\nu$) | 2 |
| Number of transmit power levels ($L$) | $\{2; 4; 6; 8; 10\}$ |
| Circuit power consumption ($P_c$) | 0.05 W |
| Noise power ($\sigma^2$) | $-174$ dBm/Hz |
| Packet size ($n_b$) | 256 |
| Number of episodes ($E$) | 500 |
| Number of learning steps ($T$) | 100 |
| Number of hidden layers | 3 |
| Number of neurons per hidden layer | $\{256, 128, 64\}$ |
| $\epsilon$-greedy policy | $\epsilon = 1$ and $\epsilon_{min} = 0.001$ |
| Learning rate ($\alpha$) | 0.001 |
| Discount factor ($\gamma$) | 0.9 |
| Optimizer | Adam |

- *Random approach*: In this scheme, users randomly select SC and TPL for their transmissions without learning.

- *Exhaustive search (ES)*: This method determines the optimal solution through exploration of the entire network space in every TS.

- *GF-OMA method*: This method explores GF-OMA scheme, where the users utilize distinct frequency/time domains for their transmissions [132].

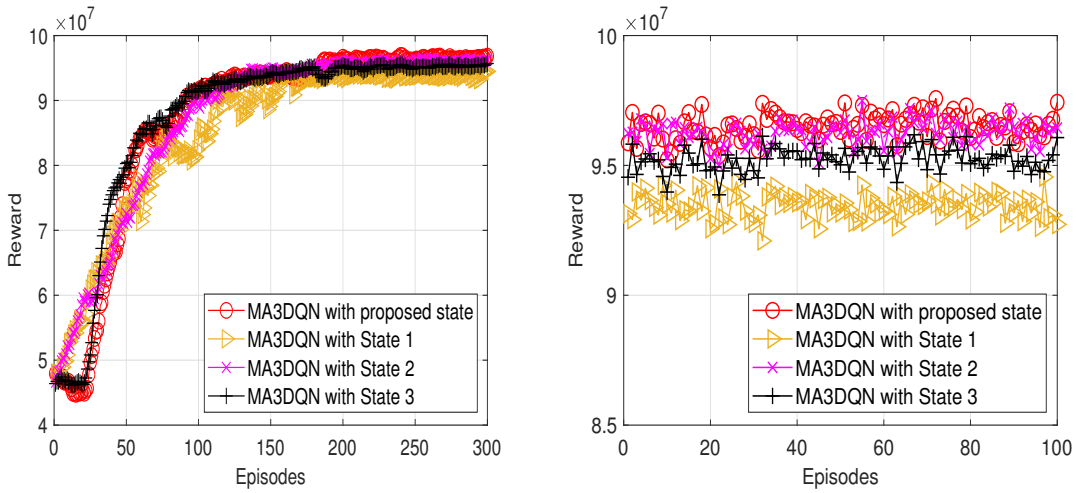
- *Different state spaces [116, 117]*: Various state spaces for MADRL-based GF-NOMA systems introduced in [116, 117] are also considered to assess the proposed methods' efficiency in terms of convergence property and signaling overhead. Specifically, the network state defined for agent $m$ in [116], named State 1, consists of its action, its channel gains over all SCs, and its transmission outcome. Meanwhile, the work [117] defines agent $m$'s state, so-called State 2, as the combination of the achievable rates of all agents.




(a) Training (b) Testing

FIGURE 4.4: Convergence analysis with different approaches, where $M = 4$, $K = 2$, $L = 7$.

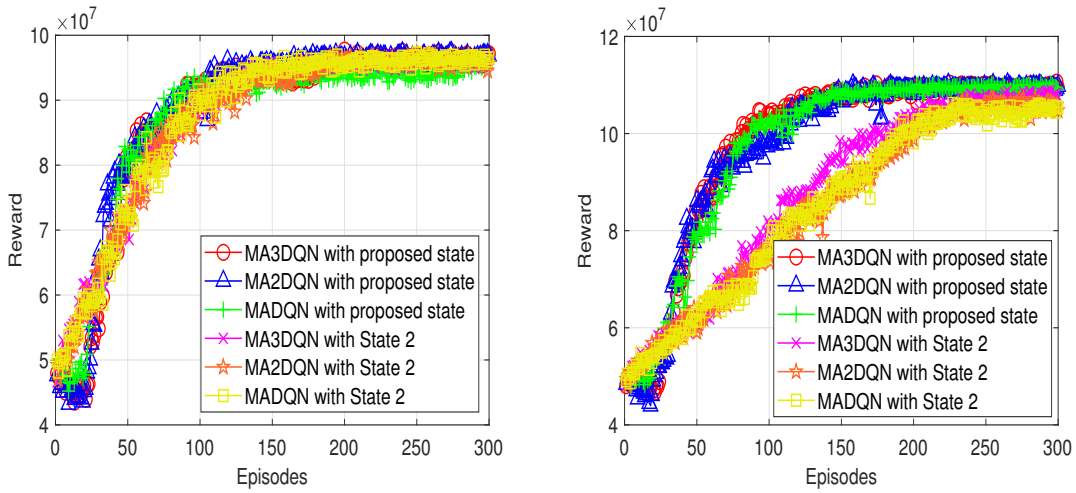
Fig. 4.4(a) shows the convergence behavior during the training phase of the URLLC-GF-NOMA approaches based on MA3DQN, MA2DQN, MADQN, MAQL, and Random schemes by plotting the reward achieved by all agents with respect to the various number of episodes. As can be observed from this figure, the Random method achieves the worst performance (i.e., lowest reward) as compared to other schemes. This is because the users randomly select SC and TPL when using this method. It is, therefore, difficult for them to find the best SC and TPL for their transmissions to optimize the network performance and guarantee URLLC requirements. Among the remaining approaches, the MAQL scheme outperforms the Random method thanks to the application of the Q-learning algorithm, but still achieves worse performance than others. This highlights the constraint of the Q-learning method when applied to a dynamic environment with an extremely large state-action space. Taking our proposed URLLC-GF-NOMA methods (i.e., MA3DQN, MA2DQN, and MADQN) into account, they are superior to the MAQL and Random methods, while achieving the same learning behavior and comparable rewards in this simulation. After the training phase, the testing phase is conducted to evaluate the training results, where the users always select the best action with the highest Q-value based on their learning results under new network conditions (network states and channels). The simulation results for the testing phase are provided in Fig. 4.4(b), where the testing process is performed over 100 episodes. This figure shows that during the testing phase, the learning-based methods (MA3DQN, MA2DQN, MADQN, and MAQL) can guarantee the convergence they achieved in the training phase.




(a) Training (b) Testing

FIGURE 4.5: Convergence analysis with different network states and MA3DQN method, where $M = 4$, $K = 2$, $L = 7$.

In Fig. 4.5, we plot the variation of the achieved reward versus the number of episodes when using the MA3DQN approach with different network state definitions. This is to evaluate the efficiency of our proposed methods in terms of convergence and signaling overhead. Specifically, we investigate two network states used for GF-NOMA systems in [116, 117], namely State 1 and State 2, as mentioned earlier. In addition, a channel-based state definition, so-called State 3, is also investigated, where only the channel state information (CSI) of each user is used to define its state. One can see from Fig. 4.5 that the method utilizing the proposed state in (4.10) attains rewards comparable to the method that uses State 2 and State 3, and larger than the method utilizing State 1. Furthermore, the proposed state demands lower signaling overhead than State 1, State 2, and State 3. In particular, the proposed state only requires the agents to know their own selected SC index and transmission power value, which are available at the agent. Thus, the environment only needs to provide feedback to the agents regarding their transmission outcomes (i.e., reward), which is used for the training process. Meanwhile, State 1 requires the agents to also have knowledge of their own channel quality and incorporate transmission results into their state information. This unnecessarily increases the input data for the agents' learning model. On the other hand, State 2 requires agents to grasp the achievable rates of all users. This necessitates significant information exchange between the environment and the agents, resulting in high signaling overhead. Moreover, State 3 demands for additional information exchange between the agents and the BS to achieve the CSI, increasing the




(a) Small state-action space: $M = 4$, $K = 2$, $L = 7$. (b) Large state-action space: $M = 8$, $K = 3$, $L = 10$.

FIGURE 4.6: Effect of state-action spaces on the achieved reward with different approaches.

signaling overhead but does not contribute to further improving the learning process and the system performance in our considered scenario.

Fig. 4.6(a) and Fig. 4.6(b) illustrate the effect of small and large state-action spaces (i.e., number of users ($M$), SCs ($K$), and TPLs ($L_p$)) on the achieved rewards, respectively. Herein, the MA3DQN, MA2DQN, and MADQN approaches using the proposed state and State 2 are considered. As demonstrated by these figures, the methods using the proposed state and those employing State 2 have similar learning behavior and achieve comparable reward values in the small state-action space. However, in the large state-action space, the methods utilizing the proposed state outperform those using State 2. This is because by utilizing the proposed state, the state-action space of the considered methods is significantly reduced compared to that of the methods employing State 2, resulting in a faster learning process and higher achieved rewards for the methods using the proposed state.

Fig. 4.6(b) also illustrates that the MA3DQN method outperforms the MA2DQN and MADQN methods in the large state-action space generated by State 2. This is due to the MA3DQN approach's ability to rapidly identify optimal actions and important states, leading to better learning outcomes than the MA2DQN and MADQN techniques. The enhanced performance of MA3DQN is achieved by the separation of state and action networks at the last layer of the DNNs model used in these schemes. On the other hand, when the proposed state is employed, it results in a considerably smaller state-action space than State 2, even with an increase in $M$, $K$, and $L_p$, resulting in faster learning. As a result, the MA3DQN, MA2DQN, and MADQN methods employing the proposed state

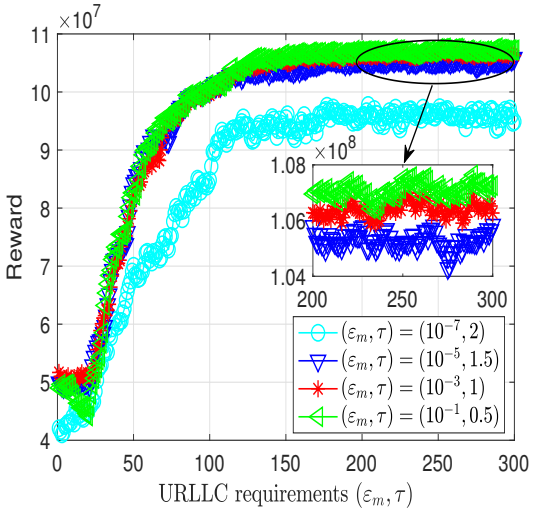



FIGURE 4.7: Effect of URLLC requirements $(\varepsilon_m, \tau)$ on the achieved reward, where $M = 4$, $K = 2$, and $L = 10$.

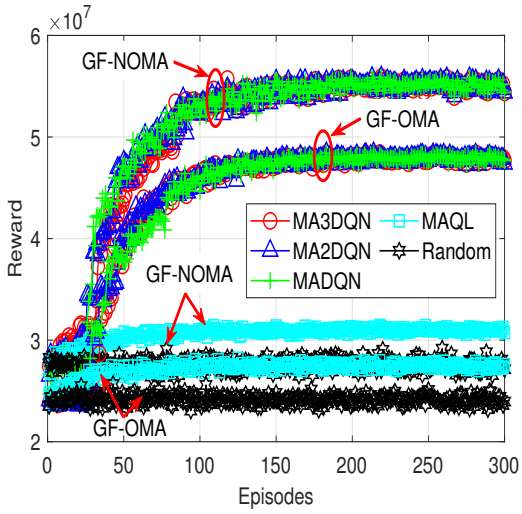



FIGURE 4.8: Performance comparison between the methods using GF-OMA and GF-NOMA, where $M = 4$, $L = 10$.

achieve comparable learning outcomes. Thus, the MA3DQN method is developed for problems with a larger state-action space, whereas the MA2DQN and MADQN methods, with a simpler network design, are suitable for problems with smaller state-action spaces.

To evaluate the effect of the URLLC requirements (i.e., $\varepsilon_m$ and $\tau$) on the system performance, we plot the variation of the achieved reward versus the number of episodes with different value sets of $(\varepsilon_m, \tau)$, while using the MA3DQN method in Fig. 4.7. This figure indicates that the achieved reward can converge to a greater value when the lower URLLC requirements are set; for instance, the reliability decreases (i.e., $\varepsilon_m$ increases from $10^{-7}$ to $10^{-1}$), and the latency threshold is degraded (i.e., $\tau$ increases from 0.5 ms to 2 ms). This can be explained by the fact that the minimum data rate threshold based on (4.7) gets higher with the increase in the URLLC requirements. It is, thus, more difficult to obtain the rate constraint required to fulfill the URLLC conditions in this case, leading to an EE performance degradation.

Fig. 4.8 shows the performance comparison in terms of the achieved reward between the methods using GF-NOMA and GF-OMA. For the GF-OMA scheme, each user occupies a distinct resource block and the system bandwidth $W$ is equally divided among the users [132]. Observing Fig. 4.8 reveals that the methods utilizing GF-NOMA obtain greater reward gains compared to those utilizing GF-OMA. This can be attributed to the performance degradation that occurs in the latter due to the splitting of bandwidth resources among users in the OMA scheme. Moreover, this figure illustrates that in both

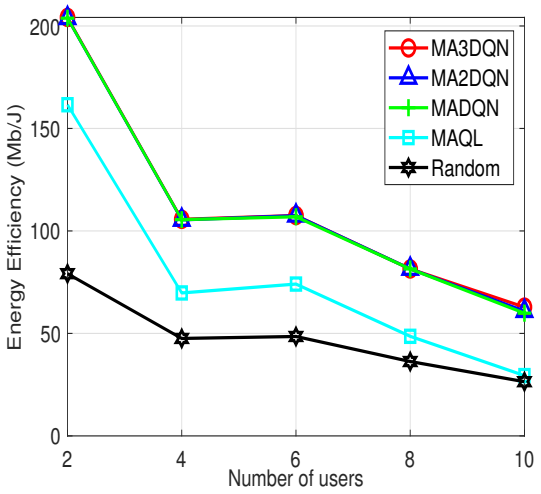



FIGURE 4.9: Effect of number of users on the EE performance with different approaches, where $K = 2$, $L = 10$.

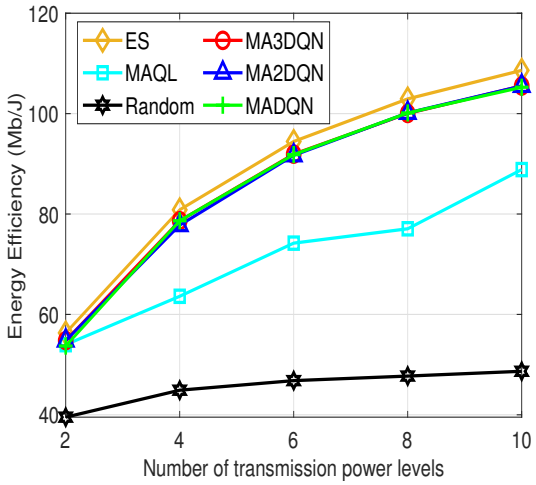



FIGURE 4.10: EE performance comparison between different methods, where $M = 4$, $K = 2$.

GF-NOMA and GF-OMA scenarios, the achieved rewards are comparable for the proposed MA3DQN, MA2DQN, and MADQN methods, and these approaches outperform the MAQL and Random schemes.

Fig. 4.9 depicts the variation of the average EE with respect to the number of users ($M$) for different methods. As observed from this figure, the EE performance decreases as the value of $M$ gets higher since the growth of the number of users sharing the same SCs in this case leads to stronger interference. In addition, the proposed MA3DQN, MA2DQN, and MADQN methods yield better EE performance than the MAQL and Random methods when $M$ increased. Furthermore, they achieve comparable EE gains under the different values of $M$. As mentioned earlier in the previous results, this is because the proposed approaches produce a small state-action space for each agent, accelerating their learning process and leading to equivalent EE performance.

Fig. 4.10 provides an EE performance comparison between the investigated methods (i.e., MA3DQN, MA2DQN, MADQN, MAQL, and Random) and an optimal solution obtained through the ES method by plotting the achieved EE versus the number of TPLs. The ES method finds the largest EE by traversing all possible actions in the network in every TS. As illustrated in Fig. 4.10, the EE values achieved by the MA3DQN, MA2DQN, and MADQN methods are close to those of the ES method and significantly exceed those of the MAQL and Random approaches. It is noteworthy that the ES method is infeasible for large network spaces since it requires exploring the entire network space, leading to

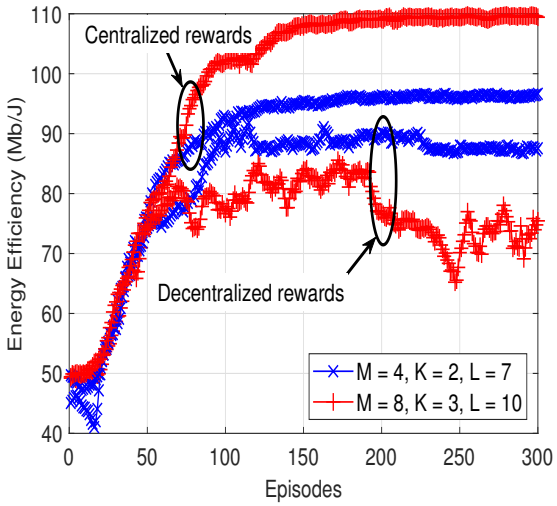
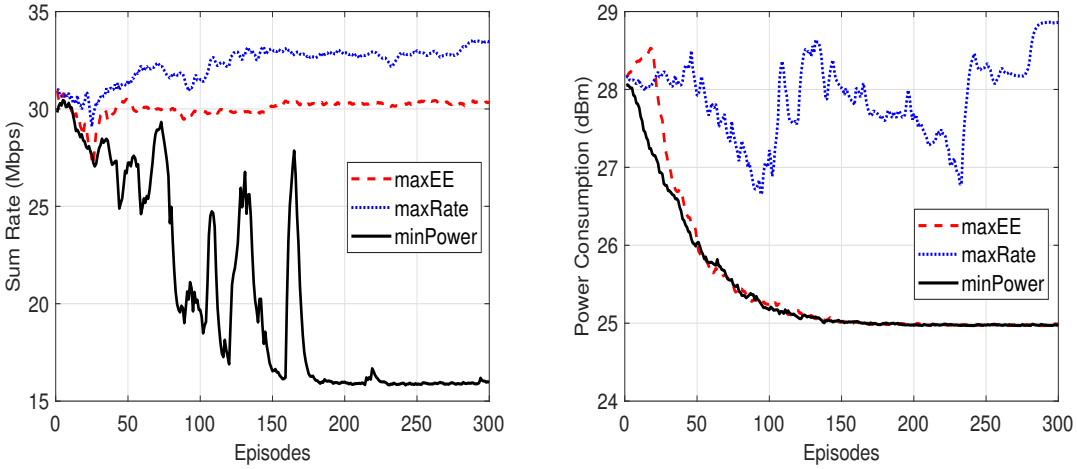



FIGURE 4.11: EE performance of MADQN method with centralized and decentralized rewards.

high computational complexity. To address this issue, the proposed URLLC-GF-NOMA methods based on MA3DQN, MA2DQN, and MADQN enable the users to interact with the wireless environment and learn from their accumulated experiences to rapidly achieve a near-optimal solution without visiting the entire network space.

Fig. 4.11 provides an EE performance comparison between MADQN methods using centralized and decentralized rewards with different values of $M$, $K$, and $L$. Specifically, the centralized reward is defined in (4.12), whereas the decentralized reward implies that each agent can receive a distinct reward depending on its own transmission outcome. In particular, with the objective of maximizing EE, the decentralized reward of each agent $m$ can be defined as $r_m(t) = R_m^{(k)}(t)/P_m^{(k)}(t)$ if its transmission is successful (i.e., $R_m^{(k)}(t) \geq \hat{R}_m$) and $r_m(t) = 0$ otherwise. Herein, $P_m^{(k)}(t)$, $R_m^{(k)}(t)$, and $\hat{R}_m$ are defined in (4.1), (4.4), and (4.7), respectively. As can be seen from Fig. 4.11, the EE performance achieved by using decentralized rewards is much smaller than the cases using centralized rewards. This is due to the fact that employing decentralized rewards can lead to the selfish behavior of agents, where they may compete with each other to maximize their own objective instead of the common one, i.e., maximizing the overall EE while guaranteeing the URLLC requirements of all users. Therefore, a significant global EE performance degradation can be observed as shown in Fig. 4.11.

As mentioned earlier in Section II-C, the problems of maximizing the achievable sum rate, named as maxRate, and minimizing the power consumption, so-called minPower, can also be investigated based on the EE maximization problem, denoted by maxEE,




(a) Achievable sum rate.

(b) Power consumption.

FIGURE 4.12: Achievable sum rate and power consumption of different problems, where $M = 4$, $K = 2$, and $L = 7$.

defined in (4.9). Herein, maxRate and minPower are achieved by setting the denominator and numerator of (4.8) as 1, respectively. Given this context, Figs. 4.12(a) and 4.12(b) depict the achievable sum rate and the power consumption versus learning episodes for different problems, including maxEE, maxRate, and minPower, respectively. These figures demonstrate that maxRate can obtain the highest sum rate but with the largest power consumption since it only focuses on maximizing the sum rate, leading to high power consumption. Meanwhile, minPower can achieve minimum power consumption but results in a poor achievable sum rate due to its power minimization objective. On the other hand, the proposed maxEE problem can achieve a high sum rate close to that obtained by maxRate while minimizing the users' power consumption. Thus, maxEE outperforms maxRate and minPower in guaranteeing the trade-off between the achievable sum rate and power consumption for energy-limited users.

Fig. 4.13 provides the EE performance of different MADRL frameworks proposed for GF-NOMA systems including our proposed solution, throughput-based solution [116], and rate-based solution [117]. As can be seen from this figure, our proposed solution achieves much better EE performance than throughput-based and rate-based solutions. This is because our proposed solution aims to maximize EE with minimum transmission power to save energy for those users with limited energy resources. In contrast, the throughput-based method tries to maximize network throughput, hence, higher transmission power than necessary can be used to ensure the successful decoding of the users' messages. Meanwhile,

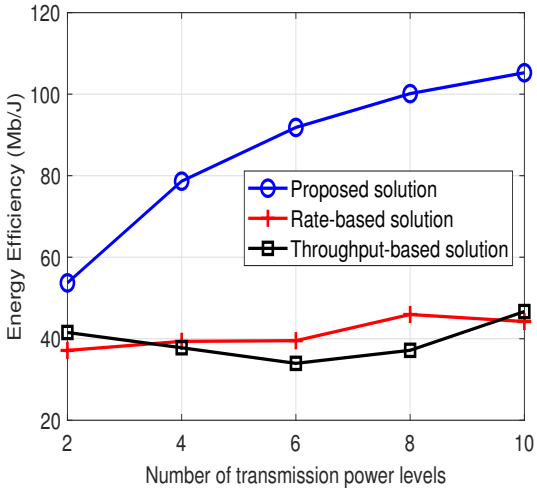



FIGURE 4.13: EE performance of different MADRL solutions for GF-NOMA systems, where $M = 4$ and $K = 2$.

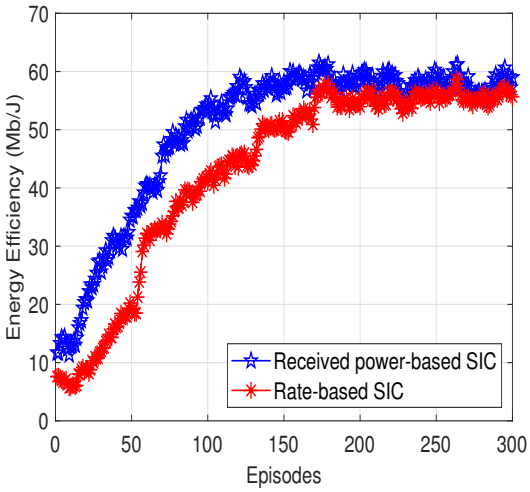



FIGURE 4.14: EE performance of different SIC methods, where $M = 4$, $K = 2$, and $L = 7$.

the rate-based solution focuses on maximizing data rate with large transmission power resulting in EE performance reduction.

To clarify the benefits of received power-based decoding order, Fig. 4.14 shows the EE comparison between received power-based and rate-based SIC methods during the learning process. Here, we consider that the predetermined rate demand of user $m$ ($1 \leq m \leq M$) is set as $m$ bps/Hz. Considering the rate-based SIC method, the message of the user with lower rate demand will be decoded earlier at the BS. This is because the user having its signal decoded earlier would suffer stronger interference and achieve a smaller data rate. As can be observed from Fig. 4.14, the received power-based SIC outperforms the rate-based SIC in terms of EE. The reason behind this result is that the decoding order in the received power-based SIC method is more flexible than that in the rate-based SIC approach, which depends on the users' channel conditions and TPL selection. This can help the users find the most appropriate SC and TPL for their transmissions to optimize the global EE performance and satisfy the different rate demands of all users. In contrast, the decoding order is fixed in the rate-based SIC method due to the predetermined rate demand of the users. It is, therefore, difficult for users to find the best learning policy, especially in time-varying and strong-interference environments, leading to performance degradation.

From the results achieved above, it can be concluded that the proposed URLLC-GF-NOMA methods based on MA3DQN, MA2DQN, and MADQN can obtain similar performance and outperform other benchmark schemes in terms of EE, convergence rate, and signaling overhead. However, the methods based on MA2DQN and MADQN exhibit lower complexity compared to the MA3DQN-based method as indicated in Section 4.3.3, thereby reducing the power consumption and processing latency for the URLLC users. This benefit makes them better suited for the considered URLLC-GF-NOMA system.

## 4.5 Summary

In this chapter, we have investigated a resource allocation problem in an uplink URLLC-GF-NOMA system where the users aim to maximize energy efficiency while satisfying their URLLC requirements. To achieve this, we have proposed three MADRL-based URLLC-GF-NOMA approaches (MA3DQN, MA2DQN, and MADQN) for the users to learn how to select the most suitable sub-channel and transmission power for their transmissions. In particular, we have designed an MADRL framework that guarantees a rapid convergence and small signaling overhead to maximize energy efficiency and satisfy users' URLLC requirements. Our simulation results have shown that the proposed URLLC-GF-NOMA methods based on MA3DQN, MA2DQN, and MADQN can achieve similar performance, but MA2DQN and MADQN are more appropriate for the investigated URLLC-GF-NOMA system due to their lower complexity compared to MA3DQN. Moreover, our proposed methods outperform existing benchmark schemes in terms of energy efficiency performance, convergence property, and signaling overhead to guarantee the URLLC requirements of energy-limited users.

Chapter 5

# Coexistence of eMBB, mMTC, and URLLC Heterogeneous Services

The escalating number of wireless users requiring different services, such as enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable low-latency communications (URLLC), has led to exploring non-orthogonal multiplexing methods like heterogeneous non-orthogonal multiple access (H-NOMA). This method allows users demanding divergent services to share the same resources. However, implementing the H-NOMA scheme faces major resource management challenges due to unpredictable interference caused by the random access mechanism of mMTC users. To address this issue, this paper proposes a joint optimization and cooperative multi-agent (MA) deep reinforcement learning-based resource allocation mechanism, aimed at maximizing the energy efficiency (EE) of H-NOMA-based networks. Specifically, this work initially establishes an optimization framework capable of determining the optimal power allocation for any specific sub-channel assignment (SA) setting for all users. Based on that, a cooperative MA double deep Q network (CMADDQN) scheme is carefully designed at the base station to conduct SA among users. In addition, a distributed full learning-based approach using MADDQN for both SA and power allocation is also designed for comparison purposes. Simulation results show that the proposed joint optimization and machine learning method outperforms the solely-learning-based approach and other benchmark schemes in terms of convergence rate and EE performance.

The rest of the chapter is organized as follows. Introduction to the current state of the art is discussed in Section 5.1. Section 5.2 presents the system model, uplink transmission strategy for H-NOMA-based systems, the achievable rate of users, and the EE maximization problem. Section 5.3 and Section 5.4 describes the proposed learning-based optimization solutions to address the EE optimization problem for the considered system. Section 5.5

provides the obtained simulation results and discussions. Finally, Section 5.6 summarize this chapter.

## 5.1 Introduction

The future wireless networks are anticipated to support a tremendous number of devices requiring heterogeneous services, e.g.,eMBB, mMTC, and URLLC, together with different quality-of-service (QoS) demands [119]. Specifically, the eMBB service aims to bring a significant increase in user data rate; the mMTC service supports the connectivity for a huge number of devices; and the URLLC is expected to provide a service with unprecedented high reliability and low latency [2]. Due to the high data rate demand, eMBB communication is designed under the assumption of infinite blocklength (iFBL) to target Shannon's channel capacity utilizing long data packets. In contrast, the demands of high connectivity density in mMTC, and ultra-reliable and low latency in URLLC require a new transmission method since mMTC and URLLC packets are generally short. In this regard, short-packet communications (SPC) have been applied for mMTC and URLLC transmissions to meet their requirements [133, 134].

To meet the diverse requirements arising from heterogeneous services, non-orthogonal multiple access (NOMA) technology is considered a promising solution [135]. Specifically, numerous studies in the literature have considered employing NOMA to efficiently manage the transmission in systems where heterogeneous services coexist [37, 136]. Thanks to the NOMA mechanism, users of heterogeneous services can simultaneously communicate with the base station (BS) using the same time-frequency resource block (RB). This is achieved through various methods such as power domain [135, 137], rate splitting [138], or codebook/pilot sequences [139]. Recently, NOMA technologies have been empowered by the introduction of the semi-grant-free (semi-GF) strategy, also known as semi-GF NOMA [140]. Following this novel integrated strategy, users having stringent QoS requirements (e.g., eMBB or URLLC users) are scheduled orthogonally by the system controllers (e.g., BS, access point, etc.) using grant-based (GB) access to fulfill their demands. Meanwhile, other users, such as mMTC users, can access the opened RBs freely according to a grant-free (GF) access mechanism. This approach can significantly increase connectivity opportunities in dense networks. On top of the semi-GF scheme, the NOMA transmission becomes particularly advantageous when more than one user accesses a specific RB.

Recently, the applications of NOMA to the systems with multiplexed diverse services have been investigated [37–40, 140–142]. In particular, the authors in [37] explored a heterogeneous NOMA (H-NOMA)-based network slicing scheme for wireless communication systems supporting the eMBB, URLLC, and mMTC services. In this work, H-NOMA is defined as a novel approach to non-orthogonal sharing of the RBs for various services,

distinct from the conventional NOMA which caters to homogeneous demands. In particular, the employment of this H-NOMA network slicing scheme enables users requiring various services to continuously utilize the same specific frequency radio resources over time. This approach leads to a significant improvement in spectrum efficiency. This work also showed that the H-NOMA-based slicing scheme can outperform heterogeneous orthogonal multiple access (H-OMA) in meeting diverse requirements under certain considered scenarios. In [140–142], the advantage of the semi-GF NOMA scheme serving both GB and GF users has been investigated in different circumstances. Furthermore, the closed-form expressions of outage probability and ergodic rate were derived in these works to analyze system performance. In [38], the authors have developed an efficient NOMA-based network slicing solution for eMBB and URLLC coexisting networks to minimize the total power consumption. In [39], the coexistence of eMBB and URLLC in MIMO NOMA systems was investigated by maximizing eMBB users' data rate while fulfilling the latency demands of URLLC users. In [40], a network slicing method for eMBB, URLLC, and mMTC based on a rate-splitting MA (RSMA) scheme was proposed. This work showed that RSMA-based network slicing can achieve better performance in terms of sum-rate in some investigated regions as compared to conventional OMA-based and NOMA-based ones.

Applying the grant-free (GF) strategy can reduce the overhead time associated with setting up transmission links; however, it also introduces complex issues related to interference management. Specifically, in scenarios involving massive access, the random access nature of GF or mMTC users can result in severe interference. This is particularly problematic when a large number of users attempt to access a limited number of RBs. This can make users' heterogeneous QoS requirements unsatisfied, leading to significant performance degradation. Furthermore, in the context of the wireless channel varying unpredictably over time, developing dynamic resource allocation mechanisms addressing the above congestion problem and fulfilling the various QoS requirements from different services becomes more challenging. In recent years, the reinforcement learning (RL) method has been applied to intelligently resolve the resource allocation problem in communications [143]. Its application to the coexistence of heterogeneous services has been investigated in [41–44].

Specifically, the authors in [41] developed an intelligent resource-slicing approach for downlink eMBB-URLLC coexisting orthogonal frequency-division MA (OFDMA) systems by exploiting the well-known deep RL (DRL) tools. Considering uplink transmission, the authors in [42] proposed a multi-agent (MA) DRL (MADRL) resource allocation framework using deep Q network (DQN) and transfer learning for OFDMA-based uplink systems serving multiple users with different QoS requirements, such as high reliability, low latency, and high data rate. In this study, the power quantization (PQ) method is utilized to discretize the continuous range of possible transmission powers into a limited set of transmission power levels (TPLs), thereby facilitating the learning process. However, this approach

may lead to performance loss if the discretized power levels are unable to closely approximate the optimal points. In [43], Fayaz *et al.* developed a DRL-based sub-channel (SC) and power allocation mechanism for semi-GF NOMA-based uplink HetNets, which aims to maximize the sum rate while meeting different rate demands of both GB and GF users. In this work, the PQ method is only employed for GF users' power allocation while the transmission power of the GB users is fixed. In [44], the authors investigated a two-hop NOMA-based uplink HetNet, where each GF user first transmits its message to a selected GB user, the chosen GB users then forward the information to BS. The GF users in this scheme were considered as DQN agents to select SC and transmission power from finite sets of all SCs and pre-discretized TPLs. In addition, a GB user is designated as the head of the GF-user cluster. Meanwhile, the power allocation for GB users is centrally defined at the BS using another DRL algorithm, namely proximal policy optimization (PPO).

Unlike the above-related works, this paper develops a joint optimization and MADRL-based method for H-NOMA-based uplink systems serving eMBB, mMTC, and URLLC users requesting heterogeneous QoS requirements, not only to speed up the learning process but also to achieve the optimal resource allocation solution. Specifically, the main contributions of this paper are summarized as follows:

1. We investigate the coexistence of eMBB, mMTC, and URLLC in a H-NOMA-based uplink system, where eMBB and URLLC users are assigned orthogonally to a number of SCs to fulfill their stringent QoS requirements on high reliability, low latency, and high data rate. Meanwhile, mMTC users can access any SCs freely and quickly without any admission approval from BS to improve the spectrum access efficiency and connectivity density.

2. We formulate an energy efficiency (EE) maximization problem for the considered system. The objective is to maximize the long-term average EE under constraints on various QoS requirements of users.

3. We design a novel learning-based resource allocation strategy to address the proposed problem. In particular, we propose a joint optimization and cooperative MA DDQN (JOCDDQN) method to optimize the resource allocation policy as well as significantly improve learning performance. The JOCDDQN method utilizes a cooperative MA DDQN (CMADDQN) scheme centralized at the BS for SC assignment based on which a dynamic power allocation solution is developed to obtain optimal transmission power for users. In addition, we also design a distributed full learning solution based on MADDQN, namely FDDQN, where all users are considered as learning agents to find the best SC and power level selection policy in order to resolve the investigated problem. It is noteworthy that the PQ method is applied in the proposed FDDQN method similar to the existing works [42, 43].

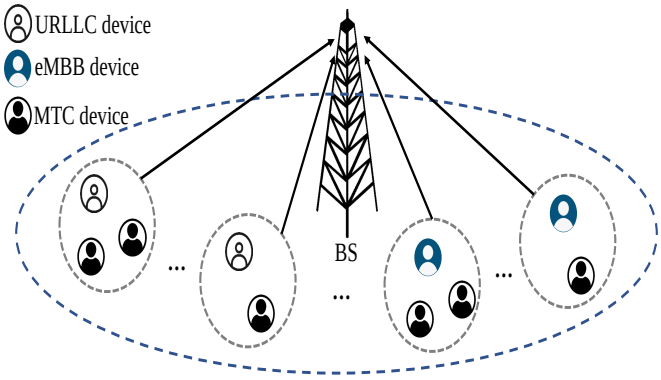



FIGURE 5.1: Illustration of a H-NOMA-based uplink system.

4. We carry out the performance comparison between our proposed methods and other benchmark schemes to evaluate the efficiency of the former in terms of convergence property and EE performance. Additionally, we provide numerical results to analyze the effects of different system parameters, such as the number of SCs, number of transmission power levels (TPLs), number of users, maximum transmission power, and divergent QoS requirements, on the system performance.

## 5.2 System Model and Problem Formulation

As shown in Fig. 5.1, we investigate an H-NOMA-based uplink system that consists of one BS located at the center of the cell with a radius of $r_c$ (m) and a number of users randomly distributed in this cell requiring different services such as eMBB, mMTC, and URLLC. Let $\mathcal{M}_{\mathsf{U}}$, $\mathcal{M}_{\mathsf{E}}$ and $\mathcal{M}_{\mathsf{M}}$ be the sets of URLLC, eMBB, and mMTC users, whose cardinalities are $M_{\mathsf{U}}$, $M_{\mathsf{E}}$ and $M_{\mathsf{M}}$, respectively. For convenience, we also denote the set of all users as $\mathcal{M} = \mathcal{M}_{\mathsf{U}} \cup \mathcal{M}_{\mathsf{E}} \cup \mathcal{M}_{\mathsf{M}}$ and $M = M_{\mathsf{U}} + M_{\mathsf{E}} + M_{\mathsf{M}}$. To serve these users, a total bandwidth of $W$ (Hz) is assumed in the system, which is divided into $K$ SCs. Let $\mathcal{K}$ be the set of all $K$ SCs. Furthermore, due to high requirements of eMBB (data rate) and URLLC (reliability and latency) services, one assumes that each of the eMBB and URLLC users is preassigned several orthogonal SCs for its transmissions. Meanwhile, the mMTC users are assumed to be able to access any available SCs freely to improve the connection density due to the massive access requirement of the mMTC service. Therefore, the mMTC users can use the SCs granted to the eMBB and URLLC users. In this case, when there are more than one user occupy the same SC, the power-domain H-NOMA scheme is applied for multi-user communication. In practice, the number of active mMTC and URLLC users is random which can be described by Poisson distribution [37]. Here, we consider the worst

scenario where all $M_{\mathsf{M}}$ mMTC users and $M_{\mathsf{U}}$ URLLC users have packets to transmit in each time-slot (TS), leading to the highest co-channel interference.

### 5.2.1 Uplink Transmission Strategy for H-NOMA

#### 5G New Radio (NR) Numerologies

5G NR standard introduces various physical-resource-block (PRB) or subchannel (SC) types in order to support different communication requirements and use-cases, which is referred to as *"numerology"*. In particular, the bandwidth of SC in 5G NR schemes is defined as $2^{\nu}$ times of SC's bandwidth in 4G systems (i.e., 180 kHz), where $\nu \in \{0; 1; 2; 3; 4\}$ is the numerology index [119]. PRBs with high SC spacing are arranged for URLLC services while traffic flows from the eMBB service can adopt a numerology with the smaller SC spacing [119]. Therefore, this chapter focuses on an SC setting that the whole bandwidth is divided into two sets of SCs, $\mathcal{K}_{\mathsf{U}}$ and $\mathcal{K}_{\mathsf{E}}$. Particularly, $\mathcal{K}_{\mathsf{U}}$ represents the set of SCs serving URLLC users with numerology $\nu_{\mathsf{U}}$ while $\mathcal{K}_{\mathsf{E}}$ is the set of eMMB-service SCs with numerology $\nu_{\mathsf{E}}$. Herein, $\mathcal{K}_{\mathsf{U}} \cup \mathcal{K}_{\mathsf{E}} = \mathcal{K}$. One assumes that $\nu_{\mathsf{E}} < \nu_{\mathsf{U}}$ and denotes $W_{\mathsf{E}} = 2^{\nu_{\mathsf{E}}} \times 180$ (kHz) and $W_{\mathsf{U}} = 2^{\nu_{\mathsf{U}}} \times 180$ (kHz) as the bandwidth of SCs corresponding to eMBB and URLLC services, respectively.

#### Uplink Transmission Mechanism

Considering the transmission over SC $k$ ($k \in \mathcal{K}$), we denote $c_z^{(k)}(t)$ ($z \in \mathcal{M}$) as a binary SC allocation variable at time-slot (TS) $t$, where $c_z^{(k)}(t) = 1$ if user $z$ occupies SC $k$ and $c_z^{(k)}(t) = 0$ otherwise. As mentioned earlier, each of eMBB and URLLC users is assigned a set of orthogonal SCs to guarantee its strict requirements. In addition, we assume a one-SC freely access strategy for mMTC users where each mMTC user can select only one arbitrary SC for its transmission. These assumptions yield the following conditions

$$(C1): \quad \sum_{z \in \mathcal{M}_{\mathsf{E}} \cup \mathcal{M}_{\mathsf{U}}} c_z^{(k)}(t) \leq 1, \quad \forall k \in \mathcal{K}. \tag{5.1}$$

$$(C2): \quad \sum_{k \in \mathcal{K}} c_z^{(k)}(t) = 1, \quad \forall z \in \mathcal{M}_{\mathsf{M}}. \tag{5.2}$$

Thus, many mMTC users can access the same SC and they can use the SCs granted to the eMBB and URLLC users. To enable a multi-user data stream over the same SC, the power-domain NOMA scheme is employed. Following NOMA principle, many users can occupy the same SC for their transmissions. In this regard, the received signal over SC $k$ at the BS in TS $t$ can be expressed as

$$y_k(t) = \sum_{z \in \mathcal{M}} c_z^{(k)}(t) \sqrt{P_z^{(k)}(t)} h_z^{(k)}(t) x_z^{(k)} + w_k(t), \tag{5.3}$$

where $w_k(t) \sim \mathcal{CN}\left(0, \sigma_k^2\right)$ is the additive white Gaussian noise (AWGN) over SC $k$ at the BS; $P_z^{(k)}(t)$, $h_z^{(k)}(t)$, and $x_z^{(k)}$ denote the transmission power, channel coefficient, and transmitted symbol of user $z$ over SC $k$, respectively. It is worth noting that the transmission power is defined as $P_z^{(k)}(t) = 0$ if $c_z^{(k)}(t) = 0$ and $P_z^{(k)}(t) \neq 0$, otherwise. From (5.3), the BS can decode the received multi-user data systematically through the use of the successive interference cancellation (SIC) technique [144]. In uplink NOMA, the decoding order of the multi-user data stream is affected by various different factors. Specifically, a decoding order can be formulated based on channel gain conditions [141], received power levels [145], or QoS constraints of users [144]. In this chapter, the messages of the users over each SC can be decoded at the BS as follows:

- Due to strict requirements on reliability and latency, the URLLC users' messages need to be decoded first. However, as long as their requirements are guaranteed, the SCs granted to them can still be used by the mMTC users to improve the spectrum efficiency.

- The symbols belonging to eMBB and mMTC users will be decoded in the order of the corresponding channel gains. In particular, the user having the higher channel gain will be decoded earlier at the BS.

- After decoding the message of a user based on the decoding order mentioned above, the BS removes this component from its observation to decode the remaining users' messages by using the successive interference cancellation (SIC) technique.

Without loss of generality, one assumes there are $Z^k$ users accessing SC $k$ in TS $t$, then they are arranged in the decoding order discussed above as $\mathcal{Z}^{(k)}(t) = \left\{z_1^{(k)}, ..., z_{Z^k}^{(k)}\right\}$. Accordingly, the received signal-to-interference-plus-noise ratio (SINR) of user $z_\ell^{(k)}$ ($1 \leq \ell \leq Z^k$) is expressed as

$$\gamma_{z_\ell^{(k)}}^{(k)}(t) = \frac{\mathcal{Y}_{z_l^{(k)}}^{(k)}(t)}{\sum\limits_{j=\ell+1}^{Z^k} \mathcal{Y}_{z_j^{(k)}}^{(k)}(t) + \sigma_k^2}, \tag{5.4}$$

where $\mathcal{Y}_z^{(k)}(t) = P_z^{(k)}(t) g_z^{(k)}(t)$ is the power of signal due to user $z$'s data over SC $k$ in TS $t$, $g_z^{(k)}(t) = \left|h_z^{(k)}(t)\right|^2$ denote the corresponding channel gain, and $\sigma_k^2 = F N_0 W_k$ represents the noise power over SC $k$. Herein, $F$ is the noise figure in dB, $N_0$ is the noise power spectral density (PSD) in dBm/Hz, $W_k$ denotes the bandwidth of SC $k$, $W_k = W_{\mathsf{E}}$ if $k \in \mathcal{K}_{\mathsf{E}}$ and $W_k = W_{\mathsf{U}}$ if $k \in \mathcal{K}_{\mathsf{U}}$.

### 5.2.2 Achievable Rate of Users

#### URLLC Communication

Regarding the transmission of URLLC user $u$ over SC $k$ in $\mathcal{K}_{\mathsf{U}}$, which happens when $c_u^{(k)} = 1$. Based on the NOMA transmission mechanism given in Section 5.2.1, one must have $u \equiv z_1^{(k)}$. Moreover, the SINR of URLLC device $u$ over SC $k$ is expressed as

$$\gamma_u^{(k)}(t) = \frac{\mathcal{Y}_u^{(k)}(t)}{\mathcal{I}_u^{(k)}(t) + \sigma_u^2}, \tag{5.5}$$

where $\mathcal{I}_u^{(k)}(t) = \sum_{j=2}^{Z^k} \mathcal{Y}_{z_j^{(k)}}^{(k)}(t)$ represents the interference caused by mMTC users over SC $k$. Furthermore, bandwidth of SC $k$ in $\mathcal{K}_{\mathsf{U}}$ is $W_{\mathsf{U}}$ and $\sigma_u^2 = FN_0W_{\mathsf{U}}$. In URLLC communication, SPC in FBL regime is implemented to meet the strict URLLC requirements. Consequently, Shannon's capacity formula cannot be applied for URLLC communication model to capture the transmission data rate and decoding error probability effectively since it is designed under the assumption of iFBL. According to [26, 146], the achievable rate of URLLC user $u$ over SC $k$ in FBL regime for a quasi-static flat fading channel can be approximated as

$$R_u^{(k)}(t) = W_{\mathsf{U}}[\log_2(1 + \gamma_u^{(k)}(t)) - \Phi_u^{(k)}(t)], \tag{5.6}$$

where $\Phi_u^{(k)}(t) = \sqrt{\frac{V_u^{(k)}(t)}{D_u W_{\mathsf{U}}}} \frac{Q^{-1}(\varepsilon_u)}{\ln 2}$, $\varepsilon_u$ is the decoding error probability (DEP) which can be used to evaluate the transmission reliability, $D_u$ is the transmission latency threshold, $Q^{-1}(x)$ is the inverse of the Gaussian Q-function, and $V_u^{(k)}(t)$ is the channel dispersion which is given by

$$V_u^{(k)}(t) = 1 - \frac{1}{\left[1 + \gamma_u^{(k)}(t)\right]^2} \approx 1, \tag{5.7}$$

where the approximation in (5.7) is achieved when $\gamma_u^{(k)}(t) \geq 5$ dB and using it for (5.6) in low SNR regime (i.e., $\gamma_u^{(k)}(t) < 5$ dB) can obtain a lower bound of the achievable rate which can guarantee users' QoS requirements [26]. Note that the channel dispersion definition in (5.7) is achieved based on the assumption that each user has its perfect channel state information[1] (CSI), such that the packet error occurs due to the noise instance only [133, 147]. For URLLC service, we assume that each URLLC user tries to upload one packet over one SC in each transmission TS. Thus, the target SNR threshold for URLLC

[1] CSI knowledge requires signaling exchange between BS and users. This can lead to a latency increase. The effect of this scenario can be analyzed by addressing the problem of latency minimization which is beyond the scope of this paper.

user $u$ that satisfies the URLLC requirements (i.e., $D_u$ and $\varepsilon_u$) can be derived based on (5.6) as [26]

$$\gamma_u^{\mathsf{tar}} = 2^{\frac{n_u}{D_u W_{\mathsf{U}}} + \frac{Q^{-1}(\varepsilon_u)}{\ln 2\sqrt{D_u W_{\mathsf{U}}}}} - 1, \tag{5.8}$$

where $n_u$ is the packet size. This demand yields the following constraints,

$$(C3): \quad c_u^{(k)}(t)\gamma_u^{(k)}(t) \geq \gamma_u^{\mathsf{tar}}, \quad \forall k \in \mathcal{K}. \tag{5.9}$$

**eMBB Communication**

Regarding the transmission of eMBB user $e$ over SC $k$ in $\mathcal{K}_{\mathsf{E}}$, which happens when $c_e^{(k)}(t) = 1$. Due to its order in the NOMA-based decoding process, its SINR denoted as $\gamma_e^{(k)}(t)$, can be defined as in (5.4) with noting that $\sigma_k^2 = FN_0W_{\mathsf{E}}$. Then, the achievable rate of eMBB user $e$ is given by

$$R_e^{(k)}(t) = W_{\mathsf{E}} \log_2 \left[1 + \gamma_e^{(k)}(t)\right]. \tag{5.10}$$

Herein, one addresses a predetermined target transmission rate, $R_e^{\mathsf{tar}}$, for each eMBB user $e$ in every TS as

$$(C4): \quad \sum_{k\in\mathcal{K}} c_e^{(k)}(t) R_e^{(k)}(t) \geq R_e^{\mathsf{tar}}, \quad \forall e \in \mathcal{M}_{\mathsf{E}}. \tag{5.11}$$

**mMTC Communication**

Based on the NOMA transmission strategy mentioned earlier in Section 5.2.1, the mMTC users can select a free SC or the one occupied by either eMBB or URLLC user. When $c_m^{(k)}(t) = 1$, mMTC user $m$ utilizes SC $k$ in TS $t$. In such case, the SINR of this user, denoted as $\gamma_m^{(k)}(t)$, can be calculated as in (5.4) with noting that $\sigma_k^2 = FN_0W_k$. In mMTC communication, SPC is utilized to exchange small packets. Consequently, the achievable rate of mMTC user $m$ over SC $k$ in FBL regime for a quasi-static flat fading channel can be approximated as [26, 146]

$$R_m^{(k)}(t) = W_k \left[\log_2(1 + \gamma_m^{(k)}(t)) - \Phi_m^{(k)}(t)\right], \tag{5.12}$$

where the parameters in (5.12) are defined similarly in (5.6). From (5.12), we define a target SINR threshold for mMTC user $m$ to satisfy its predetermined requirements on DEP and latency (i.e., $\varepsilon_m$ and $D_m$) when transmitting one packet over one SC in each TS as [26]

$$\gamma_m^{\mathsf{tar}} = 2^{\frac{n_m}{D_m W_k} + \frac{Q^{-1}(\varepsilon_m)}{\ln 2\sqrt{D_m W_k}}} - 1, \tag{5.13}$$

where $n_m$ is the packet size. This demand yields the following constraint,

$$(C5): \quad c_m^{(k)}(t)\gamma_m^{(k)}(t) \geq \gamma_m^{\mathsf{tar}}, \quad \forall k \in \mathcal{K}. \tag{5.14}$$

### 5.2.3 Energy Efficiency Maximization Problem

In this chapter, we aim to develop an effective SC and power allocation strategy to maximize the long-term average energy efficiency (EE) while ensuring the heterogeneous QoS requirements from users. To guarantee the transmission rate requirement while reducing the power consumption for the system, we define an energy efficiency (EE) factor as follows:

$$\zeta(t) = \frac{R^{\mathsf{tot}}(t)}{P^{\mathsf{Tx}}(t) + MP_{\mathsf{c}}}, \tag{5.15}$$

where $R^{\mathsf{tot}}(t) = \sum_{k\in\mathcal{K}}\sum_{z\in\mathcal{M}} c_z^{(k)}(t)R_z^{(k)}(t)$, $P^{\mathsf{Tx}}(t) = \sum_{k\in\mathcal{K}}\sum_{z\in\mathcal{M}} P_z^{(k)}(t)$, and $P_{\mathsf{c}}$ denotes the circuit power consumption. Based on (5.15), the EE maximization problem can be formulated as

$$\max_{\mathbf{c},\mathbf{P}} \quad \mathbb{E}_t\left[\zeta(t)\right] \tag{5.16a}$$

$$\text{s.t.} \quad \text{constraints } (C1) - (C5), \tag{5.16b}$$

$$(C6): \sum_{k\in\mathcal{K}} P_z^{(k)}(t) \leq P_z^{\mathsf{max}}, \ \forall (z,t), \tag{5.16c}$$

where $\mathbf{c}$ and $\mathbf{P}$ denote the SC assignment and power control strategies, respectively; and constraint (C6) stands for the power budget of users.

**Remark 7.** *Problem* (5.16) *is a mixed-integer non-linear programming (MINLP), well-known as NP-hard, which is difficult to solve. In particular, the challenges of solving this problem include the coupling between binary variables* $\mathbf{c}$ *and continuous ones* $\mathbf{P}$*. Moreover, the complicating NOMA-based SINR formula of users as given in* (5.4) *has raised another extremely critical issue to define the solution of this problem.*

## 5.3 Proposed Joint Optimization and Cooperative MADDQN Method

In this section, we present the proposed joint optimization and MADRL method to solve (5.16), where a cooperative MADDQN (CMADDQN) scheme is built for SC assignment based on which a dynamic power allocation (DPA) for every SC setting is proposed to maximize the EE in (5.16). To do this, one assumes that the perfect CSI is available at the BS.

### 5.3.1 Power Allocation for Given SC Assignment

Before presenting the CMADDQN-based SC assignment method, we first introduce our proposed power allocation solution for a given SC assignment in this section. In particular, for a fixed SC allocation, problem (5.16) in TS $t$ can be expressed as

$$\max_{\mathbf{P}} \quad \zeta(t) \tag{5.17a}$$

$$\text{s.t.} \quad \text{constraints } (C1) - (C6). \tag{5.17b}$$

where $\mathbf{p}(t) = \{p_z^{(k)}(t)\}_{\forall(z,k)}$. To tackle the challenge caused by the fraction form in the objective (5.17), an efficient method well-known as the Dinkelbach algorithm [148,149] can be employed. Following this, an iterative solution approach can be developed to obtain the optimal solution of problem (5.17) by dealing with a sequence of parameterized problems with subtracting-form objective functions. Let us state the parameterized problem for a given value of $\zeta$ as follows:

$$\max_{\mathbf{P}} \quad R^{\mathsf{tot}}(t) - \zeta P^{\mathsf{Tx}}(t) \tag{5.18a}$$

$$\text{s.t.} \quad (C1) - (C6). \tag{5.18b}$$

Then, **Theorem 1** in [149] suggests to iteratively solve problem (5.18) for a certain value of $\zeta$ and adjusting $\zeta$ until an optimal value of $\zeta^\star \geq 0$ satisfying $R^{\mathsf{tot}}(t) = \zeta^\star \left(P^{\mathsf{Tx}}(t) + MP_{\mathsf{c}}\right)$ is found.

To address the problem (5.18), we first provide the following valuable remark based on (5.4) and the uplink NOMA transmission mechanism discussed in Section 5.2.1.

**Remark 8.** *The SINR formula given in* (5.4) *demonstrates that there is no interference suffering the decoding process due to user* $z_{Z^k}^{(k)}$*. Moreover, once the power of all users in set* $\left\{z_{\ell+1}^{(k)}, ..., z_{Z^k}^{(k)}\right\}$ *is defined, the transmission power of user* $z_\ell^{(k)}$*, i.e.,* $P_{z_\ell^{(k)}}^{(k)}$*, can be optimized without coupling to other users. Hence, the power transmission can be determined in the reverse order of the coding sequence.*

Thanks to the observation given in Remark 8, an efficient approach to solving the problem (5.18) will be proposed in the following. The concept of this solution is to decompose problem (5.18) into a number of sub-problems each of which aims to obtain the power transmission of a user separately. Then, the sub-problems are solved in the order suggested in Remark 8. Particularly, the power allocation strategy for all types of users is described as follows:

- For mMTC users, each of them will have its transmission power optimized only when the power of all other users accessing the same SC with weaker channel gains are determined.

- For eMBB users, the transmission power of an eMBB user over all SCs assigned to it will be optimized jointly when all mMTC users using the same SCs with weaker channel gains have their power defined.
- For URLLC users, they will be the last ones having their transmission power optimized over all SCs that they are assigned.

Next, one focuses on presenting the sub-problems as well as their solution corresponding to eMBB, mMTC, and URLLC services.

**Power Allocation for mMTC Users**

Considering the user set accessing SC $k$ $\mathcal{Z}^{(k)}(t)$ defined in Section 5.2.1, one assumes that user $m \equiv z_\ell^{(k)}$ is an mMTC user. As described above, when $\{P^{(k)}_{z_i^{(k)}}\}_{i=\ell+1}^{Z^k}$ are determined, the power of this user, i.e., $P_m^{(k)}$, is optimized by solving the following sub-problem.

$$\max_{p} \quad W_k \left[\log_2\left(1 + A_m^{(k)} p\right) - \Phi_m^{(k)}\right] - \zeta p \tag{5.19a}$$

$$\text{s.t.} \quad P_m^{\mathsf{tar}} \leq p \leq P_m^{\mathsf{max}}, \tag{5.19b}$$

where $P_m^{\mathsf{tar}} = \gamma_m^{\mathsf{tar}}/A_m^{(k)}$, $p$ is the power transmission variable, $A_m^{(k)}$ represents the NOMA-based channel gain over interference and noise ratio (NOMA-CINR) of user $m$, i.e.,

$$A_m^{(k)} = \frac{\left|h_m^{(k)}(t)\right|^2}{\sum_{j=\ell+1}^{Z^k} \left|h^{(k)}_{z_j^{(k)}}(t)\right|^2 P^{(k)}_{z_j^{(k)}} + \sigma_k^2}. \tag{5.20}$$

Note that constraint $P_m^{\mathsf{tar}} \leq p$ is equivalent to $(C5)$ for this user. The solution of the problem (5.19) is determined in the following proposition.

**Proposition 9.** *The transmission power of mMTC user $m$ over SC $k$ is the solution of* (5.19) *which is given as*

$$P_m^{(k)\star} = \min\left(\max\left(\frac{W_{\mathsf{k}}}{\zeta \ln 2} - \frac{1}{A_m^{(k)}}, P_m^{\mathsf{tar}}\right), P_m^{\mathsf{max}}\right). \tag{5.21}$$

*Proof.* The proof can be described simply as follows. Let us define $y_m^{(k)}(p)$ as

$$y_m^{(k)}(p) = W_k \left[\log_2\left(1 + A_m^{(k)} p\right) - \Phi_m(k)\right] - \zeta p. \tag{5.22}$$

As can be seen, problem (5.19) is convex due to that $y_m^{(k)}(p)$ is concave and the feasible set

is convex. If the feasible set is not regarded, we first derive the derivative of $y_m^{(k)}(p)$ with respect to the variable $p$ as

$$\frac{\partial y_m^{(k)}(p)}{\partial p} = \frac{W_k A_m^{(k)}}{\left(1 + A_m^{(k)} p\right) \ln 2} - \zeta. \tag{5.23}$$

From (5.23), the maximum point of the objective function can be defined by resolving the equation $\partial y_m^{(k)}(p)/\partial p = 0$, which yields

$$\hat{p} = \frac{W_k}{\zeta \ln 2} - \frac{1}{A_m^{(k)}}. \tag{5.24}$$

Then, by taking the feasible set into account, the optimal solution of problem (5.19) can be defined as given in (5.21) which has finished the proof. □

**Power Allocation for eMBB Users**

Considering the transmission of eMBB user $e \in \mathcal{M}_{\mathsf{E}}$. Assume that user $e$ is assigned $n$ SCs named as $\{k_1^e, ..., k_n^e\}$, and it is denoted as user $z_\ell^{(k_j^e)} \in \mathcal{Z}^{(k_j^e)}(t)$ over SC $k_j^e$ ($1 \le j \le n$ and $1 \le \ell \le Z^{k_j^e}$), where $\mathcal{Z}^{(k_j^e)}(t) = \left\{ z_1^{(k_j^e)}, ..., z_{Z^{k_j^e}}^{(k_j^e)} \right\}$ denotes the user set accessing SC $k_j^e$ at TS $t$ arranged in the decoding order explained in Section 5.2.1. Then, if all mMTC users with weaker channel gains on $\{k_1^e, ..., k_n^e\}$ have their power defined, the transmission power over all SCs of eMBB user $e$ can be determined as follows. Denote $A_j^e$ be the NOMA-CINR of eMBB user $e$ over SC $k_j^e$ which is defined similarly as in (5.20). Then, the decomposed part of the problem (5.18) according to eMBB user $e$ can be stated as

$$\max_{\mathbf{P}^e} \quad \sum_{j=1}^{n} \left( W_{\mathsf{E}} \log_2 \left(1 + A_j^e p_j^e\right) - \zeta p_j^e \right) \tag{5.25a}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} W_{\mathsf{E}} \log_2 \left(1 + A_j^e p_j^e\right) \ge R_e^{\mathsf{tar}}, \tag{5.25b}$$

$$\sum_{j=1}^{n} p_j^e \le P_e^{\mathsf{max}}, \tag{5.25c}$$

where $\mathbf{P}^e = [p_1^e, ..., p_n^e]$ and $p_j^e$ denotes the transmission power variable corresponding to eMBB user $e$ over SC $k_j^e$. As can be observed, this problem is convex and hence its optimal solution can be obtained by employing the duality method. In particular, the solution

approach can begin with describing the Lagrangian of (5.25) as

$$\mathcal{L}(\mathbf{P}^e, \kappa, \lambda) = \sum_{j=1}^{n} \left[ (1+\kappa) W_{\mathsf{E}} \log_2 \left(1 + A_j^e p_j^e\right) - (\zeta + \lambda) p_j^e \right] - \kappa R_e^{\mathsf{tar}} + \lambda P_e^{\mathsf{max}}, \tag{5.26}$$

where $\kappa$ and $\lambda$ are the Lagrangian multipliers corresponding to constrains (5.25b) and (5.25c), respectively. Then, the dual function can be defined as the maximum of the Lagrangian function as

$$\mathsf{g}(\kappa, \lambda) = \max_{\mathbf{P}^b} \mathcal{L}(\mathbf{P}^b, \kappa, \lambda). \tag{5.27}$$

**Proposition 10.** *The solution of the right-hand-side (RHS) of* (5.27) *is defined as*

$$p_j^b = \max\left( \frac{(1+\kappa) W_{\mathsf{E}}}{(\lambda + \zeta) \ln 2} - \frac{1}{A_j^b}, 0 \right). \tag{5.28}$$

*Proof.* The proof of this proposition can be obtained easily by solving the following equation: $\partial \mathcal{L}(\mathbf{P}^b, \kappa, \lambda)/\partial p_j^b = 0$. □

Then, to determine the values of $\kappa$ and $\lambda$, the dual problem can be written as

$$\max_{\kappa, \lambda} \mathsf{g}(\kappa, \lambda) \quad \text{s.t. } \kappa, \lambda \geq 0. \tag{5.29}$$

Since problem (5.25) is convex, the dual-gap between the primary and dual problem is zero [150]. In the following, one will describe a searching approach to define the optimal solution of the dual problem by using the standard sub-gradient method, where the dual variable $\kappa$ and $\lambda$ can be iteratively updated as follows:

$$\kappa^{(v+1)} = \left[ \kappa^{(v)} - \delta^{(v)} \left( \sum_{j=1}^{n} W_{\mathsf{E}} \log_2 \left(1 + A_j^b p_j^b\right) - R_e^{\mathsf{tar}} \right) \right]^+, \tag{5.30}$$

and

$$\lambda^{(v+1)} = \left[ \lambda^{(v)} + \delta^{(v)} \left( \sum_{j=1}^{n} p_j^e - P_e^{\mathsf{max}} \right) \right]^+, \tag{5.31}$$

where the suffix $(v)$ represents the iteration index, $\delta^{(v)}$ is the step size, and $[x]^+$ is defined as $\max(0, x)$. This sub-gradient method guarantees the convergence if the step-size $\delta^{(v)}$ is chosen appropriately so that $\delta^{(v)} \overset{v \to \infty}{\longrightarrow} 0$ such as $\delta^{(v)} = 1/\sqrt{v}$ [150, 151].

**Power Allocation for URLLC Users**

Similar to the previous section, one assumes that there are $l$ SCs are assigned to URLLC user $u$ which are denoted as $\{k_1^u, ..., k_l^u\}$. Then, if the power of all mMTC users on SCs $\{k_1^u, ..., k_l^u\}$ are determined, the power transmission over all SCs can be determined by solving the following problem

$$\max_{\mathbf{P}^u} \quad \sum_{j=1}^{l} \left( W_{\mathsf{U}} \left( \log_2 \left( 1 + A_j^u p_j^u \right) - \Phi_j^u \right) - \zeta p_j^u \right) \tag{5.32a}$$

$$\text{s.t.} \quad p_j^u \geq \gamma_u^{\mathsf{tar}} / A_j^u, \quad \forall j \tag{5.32b}$$

$$\sum_{j=1}^{l} p_j^u \leq P_u^{\mathsf{max}}, \tag{5.32c}$$

where $\gamma_u^{\mathsf{tar}}$ is defined in (5.8), $\Phi_j^u = \sqrt{\frac{V_j^u}{D_u W_{\mathsf{U}}}} \frac{Q^{-1}\left(\varepsilon_j^u\right)}{\ln 2}$, $V_j^u = 1 - \left(1 + A_j^u p_j^u\right)^{-2} \approx 1$ [152], $\mathbf{P}^u = [p_1^u, ..., p_l^e]$ and $p_j^u$ denotes the transmission power variable corresponding to URLLC user $u$ over SC $k_j^u$.

Similar to the results obtained in solving problem (5.25), the transmission power of URLLC user $u$ over SCs $\{k_1^u, ..., k_l^u\}$, can be defined as

$$p_j^u = \max \left( \frac{W_{\mathsf{U}}}{(\psi + \zeta) \ln 2} - \frac{1}{A_j^u}, \frac{\gamma_u^{\mathsf{tar}}}{A_j^u} \right), \forall j. \tag{5.33}$$

where $\psi$ can be iteratively updated as follows:

$$\psi^{[v+1]} = \left[ \psi^{[v]} + \delta^{[v]} \left( \sum_{j=1}^{l} p_j^u - P_u^{\mathsf{max}} \right) \right]^+. \tag{5.34}$$

Thanks to the Dinkelbach solution approach and the power allocation mechanism given above, one proposes an power control algorithm which is summarized in Algorithm 3.

### 5.3.2 CMADDQN-based SC Assignment Strategy

We assume that each of eMBB and URLLC users is preassigned several SCs for its transmissions to guarantee its high requirements. Meanwhile, mMTC users can use SCs freely to reduce access latency and increase the number of active users [50]. To do this, we investigate a CMADDQN-based DRL method to help mMTC users quickly select the best SCs for their transmissions. By using CMADDQN scheme, each mMTC user is mapped to a learning agent and all $M_{\mathsf{M}}$ agents are centralized at the BS to exploit full information on

**Algorithm 3** Energy-Efficiency Power Allocation Algorithm

Initialize $v = 0$, $\zeta^{(0)} = 0$, step size $\delta$, the Lagrangian multipliers $\kappa$, $\lambda$, and $\psi$, $\chi_i = 1$ $(1 \leq i \leq 4)$, and choose predetermined tolerate $\phi$.
**repeat**
  **for** $k = 1, \ldots, K$ **do**
    Determine the number of users accessing SC $k$, i.e., $Z^k$.
    **for** $z = Z^k, \ldots, 1$ **do**
      **if** $z \in \mathcal{M}_{\mathsf{M}}$ **then**
        Determine $P_z^{(k)}$ as in (5.21).
      **end if**
      **if** $z \in \mathcal{M}_{\mathsf{E}}$ **then**
        Determine $P_z^{(k)}$ as in (5.28).
      **end if**
      **if** $z \in \mathcal{M}_{\mathsf{U}}$ **then**
        Determine $P_z^{(k)}$ as in (5.33).
      **end if**
    **end for**
  **end for**
  Update $\kappa^{(v+1)}$, $\lambda^{(v+1)}$, and $\psi^{(v+1)}$ as in (5.30), (5.31), and (5.34), respectively.
  Update $\zeta^{(v+1)} = \frac{R^{\text{tot}}(t)}{P^{\text{Tx}}(t) + MP_{\text{c}}}$.
  Update $\chi_1 = |\kappa^{(v+1)} - \kappa^{(v)}|$, $\chi_2 = |\lambda^{(v+1)} - \lambda^{(v)}|$, $\chi_3 = |\psi^{(v+1)} - \psi^{(v)}|$, and $\chi_4 = |\zeta^{(v+1)} - \zeta^{(v)}|$.
  Set $v := v + 1$.
**until** $\chi_i \leq \phi$.

users available at the BS [30]. This facilitates the learning process and helps users select the most appropriate SC for their transmissions to maximize the overall network EE.

We denote $\mathcal{S}$, $\mathcal{A}$, and $\mathcal{R}$ as the set of states, actions, and rewards, respectively. At the beginning of each TS $t$, an agent observes the current state $s(t) \in \mathcal{S}$ to take an action $a(t) \in \mathcal{A}$. After performing the action $a(t)$, the agent can receive a reward/penalty from the environment and discover the next state $s(t+1)$. Thanks to the feedback from the environment, the agent can update/strengthen its decision policy. The such process can be operated continuously until optimal policy can be obtained at the agent. In addition, centralized or decentralized rewards can be utilized in MADRL. Specifically, MADRL methods with centralized rewards provide a common reward for all agents, whereas each agent can receive a distinct reward in MADRL methods with decentralized rewards [153]. However, using decentralized rewards can lead to the selfish behavior of agents, where they may compete with each other to maximize their own rewards resulting in a global performance degradation. To avoid this issue, the same reward can be allocated to all agents in order to achieve a common objective [154] (e.g., maximizing the overall network EE while guaranteeing all users' QoS requirements). Given this context, the definitions of state, action, and reward according to each agent $m \in \mathcal{M}_{\mathsf{M}}$ are described as follows.

- State: The state of agent $m$ in TS $t$, $s_m(t) \in \mathcal{S}_m$, is defined as the combination of its current channel gains over $K$ SCs and the SC selection status of all $M_{\mathsf{M}}$ mMTC users in the previous TS, i.e.,

$$s_m(t) = \{\mathbf{g}_m(t), \mathbf{c}(t-1)\}, \tag{5.35}$$

where $\mathbf{g}_m(t) = \left\{g_m^{(1)}(t), \ldots, g_m^{(K)}(t)\right\}$, $\mathbf{c}(t-1) = \{\mathbf{c}_1(t-1), \ldots, \mathbf{c}_{M_{\mathsf{M}}}(t-1)\}$, and $\mathbf{c}_m(t-1) = \left\{c_m^{(1)}(t-1), \ldots, c_m^{(K)}(t-1)\right\}$. Thus, the state of agent $m$ has the cardinality of $K(1+M_{\mathsf{M}})$.

- Action: Since each mMTC user $m$ can use only one SC every TS, the action of agent $m$ in TS $t$, $a_m(t) \in \mathcal{A}_m$, is defined as its current SC selection which is expressed as

$$a_m(t) \in \mathcal{A}_m = \{1, \ldots, K\}. \tag{5.36}$$

As can be observed, the action space size of agent $m$ is determined as $|\mathcal{A}_m| = K$. For the action selection strategy, the $\epsilon$-greedy policy can be exploited, where the random action is taken with the probability of $\epsilon$ and the action with the highest Q-value is employed for the remaining probability. In particular, the action $a_m(t)$ is selected based on the $\epsilon$-greedy policy can be mathematically expressed as

$$a_m(t) = \begin{cases} \text{random action}, & \text{with probability } \epsilon, \\ a_m^{\mathsf{max}}, & \text{with probability } 1-\epsilon, \end{cases} \tag{5.37}$$

where, $a_m^{\mathsf{max}} = \text{argmax}_{a \in \mathcal{A}_m} \{Q(s_m(t), a; \boldsymbol{\theta}_m)\}$, $Q(s_m(t), a_m(t); \boldsymbol{\theta}_m)$ is the Q-value corresponding to action $a_m(t)$ at state $s_m(t)$.

- Reward: this chapter aims to optimize the average network EE while fulfilling the heterogeneous QoS requirements of all users. Therefore, we design a CMADDQN algorithm for SC assignment with centralized rewards to optimize the above common objective. Specifically, we use the achieved EE in (5.15) to define the immediate common reward for all agents. Thus, the reward function in TS $t$, denoted by $r(t)$, is defined as

$$r(t) = \begin{cases} \zeta(t), & \text{if all constraints are satisfied}, \\ 0, & \text{otherwise}. \end{cases} \tag{5.38}$$

Based on the actions and rewards obtained from trials, each agent $m$ builds its own DDQN model consisting of two deep neural networks (DNNs), namely online and target networks corresponding to weight matrices $\boldsymbol{\theta}_m$ and $\boldsymbol{\theta}'_m$, respectively. Herein, the online network is used to select an action. Meanwhile, the target network is applied to evaluate the online-network-based action. Thus, the objective is to reduce the loss function which

is formulated as [155]

$$L(\boldsymbol{\theta}_m) = \left[y_m(t) - Q_m\left(s_m(t), a_m(t); \boldsymbol{\theta}_m\right)\right]^2, \tag{5.39}$$

where $y_m(t)$ denotes the target Q-value determined by the target network as

$$y_m(t) = r(t) + \gamma Q_m\left(s_m(t+1), \underset{a\in\mathcal{A}_m}{\operatorname{argmax}} \mathcal{Q}_m(t+1); \boldsymbol{\theta}'_m\right), \tag{5.40}$$

where $\gamma$ denotes the discount factor, $\mathcal{Q}_m(t+1) = Q_m\left(s_m(t+1), a; \boldsymbol{\theta}_m\right)$. It is noteworthy from (5.40) that in DDQN model, the action selection, i.e., $\underset{a\in\mathcal{A}_m}{\operatorname{argmax}} \mathcal{Q}_m(t+1)$, and action evaluation, i.e., $Q_m(s_m(t+1), \underset{a\in\mathcal{A}_m}{\operatorname{argmax}} \mathcal{Q}_m(t+1); \boldsymbol{\theta}'_m)$, are decoupled using two different Q-value function to avoid the overestimation issue in the conventional DQN model [156].

Given the above discussions, we propose a joint optimization and cooperative MADDQN method, so-called JOCDDQN, to address the problem (5.16). Specifically, in TS $t$, the CMADDQN is applied for SC allocation, where each agent $m$ observes its current state $s_m(t) \in \mathcal{S}_m$ and takes an action $a_m(t) \in \mathcal{A}_m$. The power allocation method in Algorithm 3 is then utilized to achieve optimal transmission power for all users. After that, agent $m$ observes the environment to receive a reward $r(t)$ and moves to a new state $s_m(t+1)$. It then stores an experience tuple of $(s_m(t), a_m(t), r(t), s_m(t+1))$ into its experience replay memory, and a minibatch of experiences is sampled for training the online network. The weight matrix of the online network $\boldsymbol{\theta}_m$ is then updated to minimize the loss function in (5.39) by using the stochastic gradient method. After a predetermined number of TSs ($F$), the weight matrix of the target network $\boldsymbol{\theta}'_m$ is updated by copying $\boldsymbol{\theta}_m$. The above process continues until reaching a predefined number of episodes guaranteeing the algorithm's convergence. The proposed JOCDDQN algorithm is summarized in Algorithm 4.

## 5.4 Proposed Distributed Reinforcement Learning Method

In this section, we present another DRL-based solution to address (5.16) by developing a full MADDQN method, namely FDDQN, in a distributed manner to reduce the information exchange between the BS and users. The FDDQN method is designed to conduct both SC assignment and power allocation.

### 5.4.1 FDDQN Method

Employing FDDQN method, all users are considered as learning agents to find the optimal policies for selecting both SC and transmission power. In addition, the multi-level quantization strategy is exploited to deal with the continuous characteristic of power variables in

**Algorithm 4** JOCDDQN Algorithm for Maximizing Energy Efficiency

Initialize the weight matrices of the online and target networks, i.e., $\boldsymbol{\theta}_m$ and $\boldsymbol{\theta}'_m$, $\forall m \in \mathcal{M}_{\mathsf{M}}$.
**for** $i = 1, \ldots, E_p$ **do**
  Initialize the state $s_m(t)$, $\forall m$.
  **for** $t = 1, \ldots, T$ **do**
    All agents take an action (SC selection) $a_m(t)$ $(\forall m)$ following the $\varepsilon$-greedy policy in (5.37).
    Run Algorithm 3 to achieve the optimal transmission power for all users.
    The BS broadcasts the SC selection and the power allocation to users.
    All agents observe the reward $r(t)$ in (5.38) and move to the next states $s_m(t+1)$ $(\forall m)$.
    **for** $m = 1, \ldots, M_{\mathsf{M}}$ **do**
      Store an experience tuple $(s_m(t), a_m(t), r(t), s_m(t+1))$ to the memory of agent $m$.
      Randomly sample a mini-batch of experiences from the memory to train the online network.
      Update $\boldsymbol{\theta}_m$ by using gradient descent to minimize the loss function in (5.39).
      **if** $t\%F = 0$ **then**
        Set $\boldsymbol{\theta}'_m = \boldsymbol{\theta}_m$.
      **end if**
    **end for**
  **end for**
**end for**

the similar approach introduced in [145, 155]. Specifically, we investigate a scenario where mMTC users build their own DDQN model to learn how to choose the best SC and power level for their transmissions from the available SCs and TPLs sets. Furthermore, eMBB and URLLC users are also learning agents to select suitable TPLs for their communication over preassigned orthogonal SCs.

Given the above context, the states, actions, and rewards of agents according to eMBB, mMTC, and URLLC users are defined as follows:

- State: In TS $t$, the states of agents according to user $z \in \mathcal{M}_{\mathsf{E}} \cup \mathcal{M}_{\mathsf{U}}$ and user $m \in \mathcal{M}_{\mathsf{M}}$ are respectively defined as

$$s_z(t) = \{\mathbf{g}_z(t), \mathbf{p}_z(t-1)\}, \tag{5.41}$$

  and

$$s_m(t) = \{\mathbf{g}_m(t), \mathbf{c}_m(t-1), \mathbf{p}_m(t-1)\}, \tag{5.42}$$

  where $\mathbf{g}_z(t) = \left\{g_z^{(k_1^z)}(t), \ldots, g_z^{(k_b^z)}(t)\right\}$ ($b = n$ if $z \in \mathcal{M}_{\mathsf{E}}$ and $b = l$ if $z \in \mathcal{M}_{\mathsf{U}}$) is the channel gain vector of user $z$ over assigned SCs, $\mathbf{p}_z(t-1) = \left\{P_z^{(k_1^z)}(t-1), \ldots, P_z^{(k_b^z)}(t-1)\right\}$ denotes the TPLs selection of user $z$ over assigned SCs, $\mathbf{g}_m(t)$ and $\mathbf{c}_m(t-1)$ are defined in (5.35), and $\mathbf{p}_m(t-1) = \left\{P_m^{(1)}(t-1), \ldots, P_m^{(K)}(t-1)\right\}$.

**Algorithm 5** FDDQN Algorithm for Maximizing Energy Efficiency

Initialize the weight matrices of the online and target networks, i.e., $\boldsymbol{\theta}_z$ and $\boldsymbol{\theta}'_z$, $\forall z \in \mathcal{M}$.
**for** $i = 1, \ldots, E_p$ **do**
  Initialize the state $s_z(t)$, $\forall z$.
  **for** $t = 1, \ldots, T$ **do**
    All agents take an action $a_z(t)$, $\forall z$, following the $\varepsilon$-greedy policy in (5.37), where $a_z(t)$ is defined in (5.43) if $z \in \mathcal{M}_{\mathsf{E}} \cup \mathcal{M}_{\mathsf{U}}$ and in (5.44) if $z \in \mathcal{M}_{\mathsf{M}}$.
    All agents observe the reward $r(t)$ in (5.38) and move to the next states $s_z(t+1)$ $(\forall z)$.
    **for** $z = 1, \ldots, M$ **do**
      Perform steps $10 - 15$ in Algorithm 4.
    **end for**
  **end for**
**end for**

- Action: In TS $t$, the actions of agents according to user $z \in \mathcal{M}_{\mathsf{E}} \cup \mathcal{M}_{\mathsf{U}}$, $a_z(t)$, and user $m \in \mathcal{M}_{\mathsf{M}}$, $a_m(t)$, are defined as the power selection of user $z$ over $n$ granted SCs, and the SC and power selection of mMTC user $m$, respectively. Thus, $a_z(t)$ and $a_m(t)$ are respectively expressed as

$$a_z(t) = \left\{ P_z^{(k_1^z)}(t), \ldots, P_z^{(k_b^z)}(t) \right\} \in \hat{\mathcal{A}}_z, \tag{5.43}$$

  and

$$a_m(t) \in \hat{\mathcal{A}}_m = \{1, \ldots, kl, \ldots, KL\}, \tag{5.44}$$

  where $P_z^{(k)}(t) \in \mathcal{P}$, $\mathcal{P} = \{P_1, \ldots, P_L\}$ is the available $L$ TPLs set, $a_m(t) = kl$ indicates that mMTC user $m$ selects SC k and TPL $l$ in TS $t$. Thus, the action space size of agents $z$ and $m$ are determined as $\left|\hat{\mathcal{A}}_z\right| = L^b$ and $\left|\hat{\mathcal{A}}_m\right| = KL$.

- Reward: Similar to the JOCDDQN method, the reward function in TS $t$ for the FDDQN approach is defined as in (5.38).

The FDDQN method requires each agent to create its own DDQN model following the same process as described in Section 5.3.2. The details of the proposed FDDQN algorithm are summarized in Algorithm 5.

### 5.4.2 Complexity Analysis

#### FDDQN Algorithm

Let $H$, $N_h$, and $I_s$ be the number of training layers, the number of neurons in layer $h$, and the size of the input layer. For each TS, the computational complexity of FDDQN

algorithm in algorithm 5 can be calculated by

$$C_{\mathsf{TS}} = \mathcal{O}\left(X\right), \tag{5.45}$$

where $X = I_s N_1 + \sum_{h=1}^{H-1} N_h N_{h+1}$. For the training phase with $M$ agents, $E$ episodes, and $T$ TSs, the computational complexity of the algorithm is given by

$$C_{\mathsf{FDDQN}} = \mathcal{O}\left(METX\right). \tag{5.46}$$

**JOCDDQN Algorithm**

The complexity of the JOCDDQN algorithm in Algorithm 4 consists of the complexity of the DPA algorithm in Algorithm 3 and the CMADDQN scheme. For the DPA algorithm, its complexity is given by

$$C_{\mathsf{DPA}} = \mathcal{O}(I(M_{\mathsf{M}} + nM_{\mathsf{E}} + lM_{\mathsf{U}})), \tag{5.47}$$

where $I$ denotes the number of iterations to get convergence. For CMADDQN scheme, its complexity is determined similarly to the one of the FDDQN algorithm as

$$C_{\mathsf{CMADDQN}} = C_{\mathsf{FDDQN}} = \mathcal{O}(METX). \tag{5.48}$$

Thus, the complexity of the JOCDDQN algorithm can be calculated as

$$C_{JOCDDQN} = C_{\mathsf{DPA}} C_{\mathsf{CDDQN}}. \tag{5.49}$$

### 5.4.3 Convergence Discussion

This part provides a discussion regarding the convergence of the proposed learning methods. In particular, the learning mechanisms in the JOCDDQN and FDDQN methods are built based on the DDQN model which combines Q-learning and neural networks (NNs). Therefore, their convergence properties can be described by considering the convergence conditions of the Q-learning and NNs' possibility of effectively approximating the non-linear Q-values [127]. Firstly, the convergence constraints of the Q-learning algorithm are expressed as [157, 158]

$$\lim_{T\to\infty} \sum_{t=1}^{T} \alpha_t = +\infty \quad \& \quad \lim_{T\to\infty} \sum_{t=1}^{T} \alpha_t^2 < +\infty, \tag{5.50}$$

where $0 \leq \alpha \leq 1$ denotes the learning rate. The conditions in (5.50) indicate that it is necessary to progressively reduce the learning rate during the training process to ensure

Table 5.1: Experimental Parameters

| **Parameters** | **Value** |
|---|---|
| Cell radius ($r$) | 500 m |
| Channel model | Rician |
| eMBB and URLLC numerology indices ($\nu_{\mathsf{E}}$ & $\nu_{\mathsf{U}}$) | 1 & 2 |
| Number of SCs ($K$) | 4 |
| Number of URLLC users ($M_{\mathsf{U}}$) | 1 |
| Number of eMBB users ($M_{\mathsf{E}}$) | 1 |
| Number of mMTC users ($M_{\mathsf{M}}$) | 6 |
| Number of SCs assigned to a URLLC/eMBB user | 2 |
| eMBB data-rate demand ($R_e^{\mathsf{tar}}$) | 2 bps/Hz |
| URLLC & mMTC latency threshold ($D_u$ & $D_m$) | 1 & 2 ms |
| URLLC & mMTC reliability threshold ($\varepsilon_u$ & $\varepsilon_m$) | $10^{-5}$ & $10^{-3}$ |
| Maximum transmission power | 23 dBm |
| Circuit power consumption ($P_c$) | 0.05 W |
| Noise figure & PSD ($F$ & $N_0$) | 6 dB & -174 dBm/Hz |
| Packet length ($m_b$) | 32 bytes |
| Number of hidden layers | 3 |
| Number of neurons per hidden layers | {256, 128, 64} |
| Learning rate ($\alpha$) | 0.001 |
| Discount factor ($\gamma$) | 0.9 |
| Optimizer | Adam |

the algorithm's convergence. Secondly, the works in [129,158] showed that NNs are capable of approximating any non-linear continuous functions. Given the above discussions, the proposed methods can achieve the convergence status.

## 5.5 Simulation Results

This section provides the simulation results to evaluate our proposed algorithms' performance. The DDQN model consists of three fully-connected hidden layers including 256, 128, and 64 neurons. The experimental parameters are provided in Table 5.1. Here, we investigate the behavior and the performance of our proposed algorithms (i.e., JOCDDQN and FDDQN) as well as the benchmark methods which are described as follows:

- Fixed transmission power (FTP): In this scheme, each user transmits their messages with a predetermined transmission power [140]. Thus, the MADDQN algorithm is applied only for SC assignment in this case.
- Fixed resource allocation (FRA): This scheme is considered in [43], where SC and power allocation for high-demand users (i.e., URLLC and eMBB users) is fixed, whereas mMTC users can find the best resource allocation policy based on MADDQN scheme.
- Random SC selection (RSS): The SC assignment is carried out randomly in this method. Based on that, Algorithm 3 is implemented for optimal power allocation.

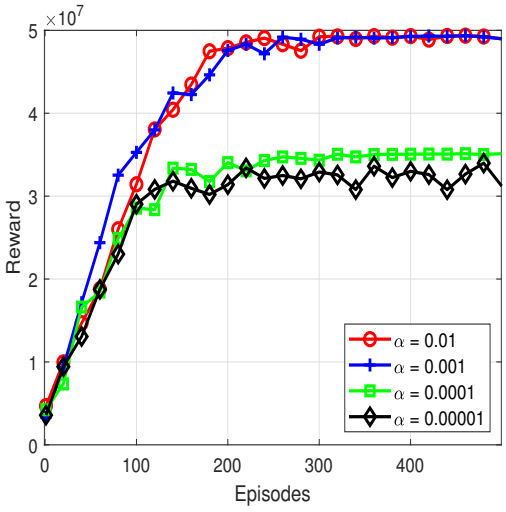



FIGURE 5.2: Convergence performance with different learning rate values.

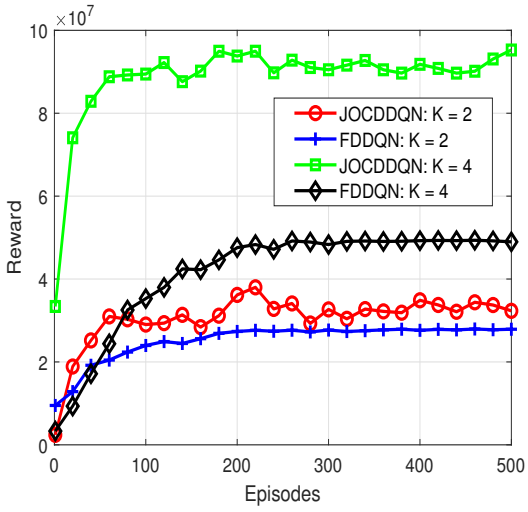



FIGURE 5.3: Convergence performance of learning approaches with different values of $K$.

Since the hyper-parameters significantly affect the learning process in the DRL algorithms, we first evaluate the effect of the learning rate on the convergence performance in Fig. 5.2. In particular, we only investigate the convergence behavior of the FDDQN method versus different values of the learning rate. Fig. 5.2 shows that a lower learning rate increases the learning time and can lead to a poor policy. In contrast, a higher learning rate can make the learning process too fast, leading to a sub-optimal solution. Therefore, the learning rate value should be selected carefully. Based on the results observed from Fig. 5.2, we set the learning rate value as 0.001.

Fig. 5.3 depicts the convergence behavior of the proposed approaches (i.e., JOCDDQN and FDDQN) by plotting the achieved reward versus number of episodes. Furthermore, different values of the number of SCs ($K$) are considered in this figure. One can observe from Fig. 5.3 that both JOCDDQN and FDDQN methods can achieve the convergence status after a number of learning episodes. More specifically, the JOCDDQN method obtains higher reward and converges faster than the FDDQN method. This is because the JOCDDQN method utilizes the CMADDQN scheme to select only SCs for users, resulting in a small action space; and applies the DPA algorithm in Algorithm 3 to attain an optimal power allocation, leading to a performance improvement in terms of the achieved reward and convergence. In contrast, the FDDQN method is designed by using the power quantization (PQ) method [43, 145] to split the transmission power into multiple discrete power levels, and the MADDQN scheme is then applied for SC and power level selection. This significantly increases the action space requiring agents to spend more time (episodes)

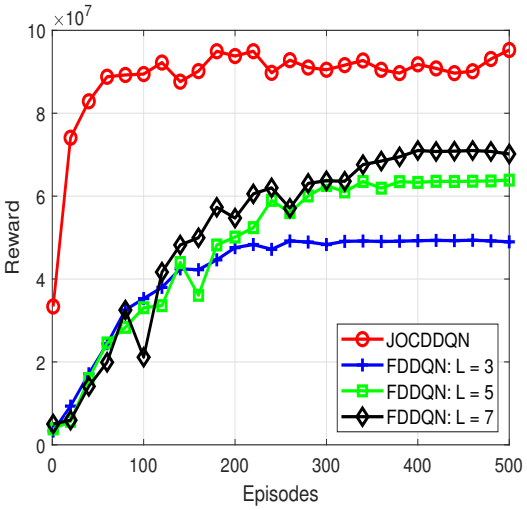



FIGURE 5.4: Convergence performance of learning approaches with different values of $L$.

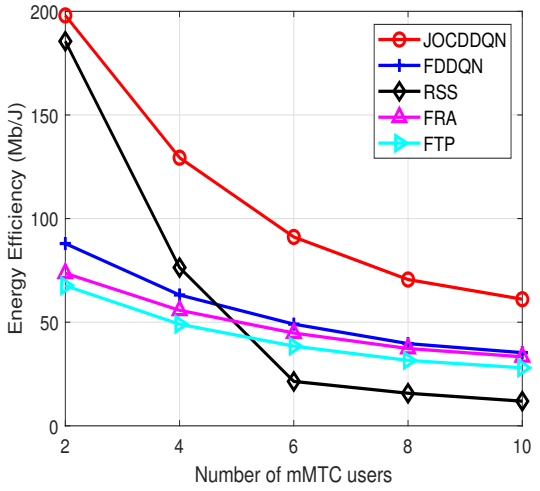



FIGURE 5.5: Energy efficiency of different approaches vesrus number of mMTC users.

exploring the environment. Additionally, the PQ method makes it challenging to obtain an optimal power allocation policy. Therefore, the FDDQN method demonstrates inferior learning performance compared to the JOCDDQN method. Furthermore, Fig. 5.3 shows that better performance in terms of the achieved reward can be observed as $K$ gets higher. This is because when $K$ increases, the number of users using the same SC decreases, reducing the co-channel interference.

Fig. 5.4 shows the effect of the number of TPLs ($L$) on the system performance by plotting the achieved reward of the proposed JOCDDQN and FDDQN methods versus the number of episodes with different values of $L$. It is noteworthy that the JOCDDQN is not influenced by the values of $L$ since it does not employ the PQ method. It can be observed from this figure that the FDDQN method can get a higher reward towards the one achieved by the JOCDDQN method when increasing $L$. However, the value of $L$ should be selected carefully since increasing $L$ makes the action space larger leading to a longer learning process.

Fig. 5.5 shows the variation of EE versus the number of mMTC users ($M_{\mathsf{M}}$) of different approaches including JOCDDQN, FDDQN, RSS, FRA, and FTP. This figure indicates that lower EE performance for all considered methods can be observed with the increase in the value of $M_{\mathsf{M}}$ due to higher interference. Moreover, the proposed JOCDDQN and FDDQN methods outperform other investigated methods in terms of EE performance when $M_{\mathsf{M}}$ increases. In particular, the proposed JOCDDQN method achieves the best EE performance thanks to its joint optimization and DRL scheme, whereas the RSS method gives

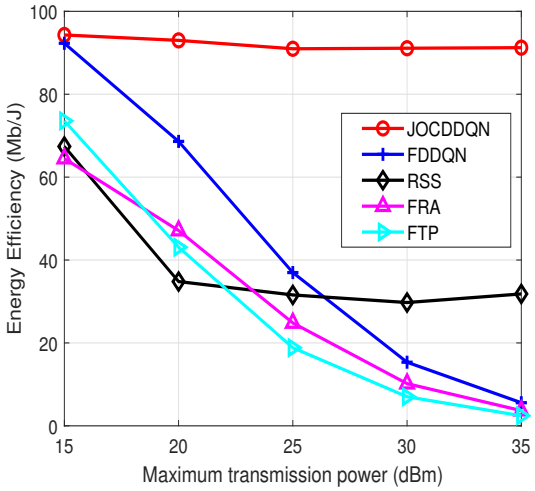



FIGURE 5.6: Energy efficiency of different approaches versus maximum transmission power.

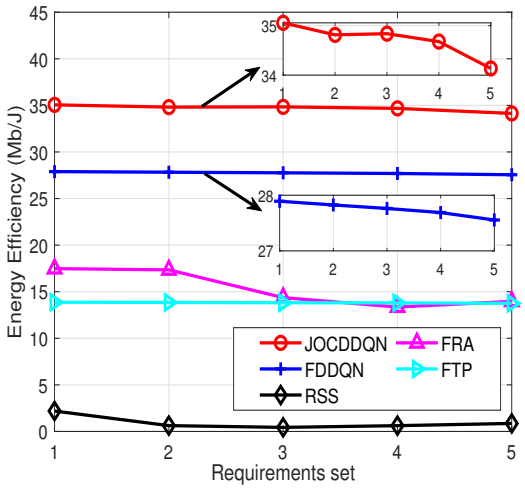



FIGURE 5.7: Energy efficiency of different approaches versus URLLC/eMBB requirement set, where $K = 2$.

the worst performance due to the random SC selection strategy applied in this method. Among the remaining approaches (i.e., FDDQN, FRA, and FTP), the proposed FDDQN method attains higher EE performance as compared to others. This is because the FRA and FTP methods are built based on some ideal assumptions leading to their performance degradation. Specifically, the FRA method fixes the transmission power of high-demand users (i.e., URLLC and eMBB users), whereas the FTP method considers a communication protocol with fixed transmission power for all users.

Fig. 5.6 depicts the effect of the maximum transmission power ($P_{\mathsf{max}}$) on the achieved EE performance of different approaches. This figure demonstrates that the proposed JOCDDQN method can still bring the best EE performance when $P_{\mathsf{max}}$ gets larger. In contrast, the EE performance achieved by the FDDQN, FRA, and FTP methods is significantly reduced with the increase in $P_{\mathsf{max}}$. This phenomenon occurs since the PQ approach utilized in these methods leads to higher transmission power when $P_{\mathsf{max}}$ scales up, causing the EE performance loss. Meanwhile, the RSS method can get higher EE performance than the FDDQN, FRA, and FTP methods when $P_{\mathsf{max}}$ becomes much larger. However, its random SC selection mechanism makes it difficult to obtain an optimal solution.

Fig. 5.7 provides the evaluation regarding the effect of URLLC and eMBB requirements, denoted by $(D_u, \varepsilon_u, R_e^{\mathsf{tar}})$, on the achieved EE performance of different approaches. Specifically, we plot the variation of the EE versus different URLLC and eMBB requirement sets $(D_u(i), \varepsilon_u(i), R_e^{\mathsf{tar}}(i))$, where $1 \leq i \leq 5$, $D_u = \{4, 3, 2, 1, 0.5\}$ (ms), $\varepsilon_u =$

$\{10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}\}$, and $R_e^{\mathsf{tar}} = \{2, 4, 6, 8, 10\}$ (bps/Hz). As can be seen from Fig. 5.7, when the requirements get more stringent, the worse EE performance can be observed. In addition, this figure also indicates that the proposed JOCDDQN method outperforms other schemes, i.e., the FDDQN, RSS, FRA, and FTP methods, in terms of EE performance under different URLLC and eMBB requirements.

Given the above-mentioned results, we can conclude that the proposed JOCDDQN method demonstrates superior performance over other considered approaches in terms of EE and convergence rate thanks to its joint optimization and cooperative MADDQN framework. However, this approach comes with a higher complexity. Meanwhile, although the proposed FDDQN method outperforms other benchmark schemes, it shows lower EE performance and convergence characteristics than the proposed JOCDDQN method. Notably, it allows for a distributed implementation with reduced complexity as compared to the JOCDDQN method.

## 5.6 Summary

In this chapter, we have investigated the H-NOMA method for the coexistence of eMBB, mMTC, and URLLC services. To analyze the system performance, we have formulated the energy efficiency maximization problem subject to divergent QoS constraints of various users. We then have proposed two MADRL-based resource allocation solutions, i.e., JOCDDQN and FDDQN, to address the considered problem. In particular, the JOCDDQN method utilizes a CMADDQN scheme for SC assignment among users while the power corresponding to a given SC setting is optimized effectively using the proposed dynamic power allocation algorithm. Meanwhile, the FDDQN method implements a MADDQN scheme for both SC and power allocation, where the continuous power variable is split into multiple discrete power levels to facilitate the learning process. Simulation results have shown that the JOCDDQN method can achieve higher energy efficiency and converge faster than the FDDQN method and other benchmark schemes.

Chapter 6

# Conclusions and Future Research

In Section 6.1, we present a summary of our works in this thesis along with their key conclusions. Additionally, Section 6.2 explores potential avenues for future research.

## 6.1 Conclusions

The progress made in URLLC has recently motivated the inspiration of researchers, leading to the proposal of numerous exciting ideas. This dissertation centers on the resource allocation design and optimization within various URLLC-enabled systems. To encapsulate the core structure of this thesis, it can be summarized as follows:

- In chapter 1, we offered a foundational understanding of the heterogeneous services in 5G and beyond wireless networks, encompassing eMBB, mMTC, and URLLC. We then presented a general overview of recent advancements and applications of URLLC under different scenarios as discussed in related works. Additionally, we highlighted the limitations and unaddressed technical challenges within the scope of these previous works, setting the stage for this thesis.

- In chapter 2, we provided fundamental knowledge about URLLC and technologies utilized in this dissertation such as short packet communications (SPC), non-orthogonal multiple access (NOMA), grant-free (GF) access, and machine learning, with a particular emphasis on reinforcement learning.

- In chapter 3, we delved into the examination of the short-packet communications (SPC) method, an auspicious transmission approach tailored for URLLC scenarios. Our focus was on the thorough investigation of SPC within downlink MIMO NOMA systems, where we derived a closed-form expression for the average block error rate

(BLER) and conducted an asymptotic analysis of its behavior in the high signal-to-noise ratio (SNR) regime. These findings served as the basis for the formulation of a problem aimed at minimizing blocklength while adhering to BLER constraint, with the goal of reducing transmission latency.

- In chapter 4, our focus was on the integration of GF-NOMA and deep reinforcement learning (DRL) in uplink URLLC scenarios. More precisely, we developed an efficient distributed resource allocation strategy for GF-NOMA systems catering to URLLC needs, employing a multi-agent DRL (MADRL) approach. This strategy's primary objective was to jointly optimize the assignment of sub-channels (SCs) and the transmission power, with the aim of maximizing network energy efficiency (EE) while also expediting algorithm convergence and reducing signaling overhead, thereby decreasing latency. Our findings demonstrated the superiority of our proposed method over benchmark schemes, particularly in terms of its convergence properties and EE performance, effectively addressing URLLC requirements.

- In chapter 5, we emphasized the coexistence scenario between eMBB, mMTC, and URLLC services within H-NOMA-based uplink systems. Our primary focus in this chapter was to introduce a joint optimization and DRL approach for devising an optimal solution for SC assignment and power allocation. This solution was designed with the goal of maximizing network energy efficiency (EE) while meeting the diverse quality-of-service (QoS) requirements of various services. To achieve this, we devised a cooperative MADRL framework centralized at the base station (BS) to intelligently manage SC assignment, based on which we developed a dynamic power allocation scheme to optimize users' transmission power. Furthermore, we also formulated a full MADRL-based resource allocation method for comparative analysis, wherein both SC and power allocation were driven by the developed MADRL modeling. The achieved results exhibited the superior performance of our proposed approach when compared to other benchmark schemes, ensuring not only optimal EE performance but also the harmonious coexistence of different services.

## 6.2 Potential Avenues for Future Research

Within this dissertation, we have delved into targeted challenges and offered efficient resolutions for various URLLC-related scenarios. Nevertheless, the evolution of the forthcoming generation of wireless communications, namely 6G, is motivated by a growing set of ambitious requirements that surpass the current network capabilities, necessitating more advanced technologies. This motivates the need for further investigation and research on URLLC and its related aspects. Based on our works conducted in this thesis, the promising research directions for future research can be outlined in the following.

- **More detailed examinations of SPC for URLLC-related systems:** Building upon the findings outlined in chapter 3, we can explore promising research avenues. In chapter 3, the focus was on assessing performance through deriving the closed-form expression of average block error rate (BLER) and minimizing blocklength while adhering to BLER constraints. A fascinating extension of this research involves delving into the challenge of maximizing throughput while simultaneously upholding reliability (BLER) and latency (blocklength) constraints. This broader investigation is essential for gaining comprehensive insights into SPC in scenarios related to URLLC. Furthermore, a compelling direction for extension lies in real-time predictive modeling for optimal throughput and blocklength. Leveraging deep learning, an optimization framework based on an efficient deep convolutional neural network (CNN) can be devised to accurately and expeditiously estimate optimal performance in terms of throughput and blocklength, in a real-time operational setting.

- **A more comprehensive analysis of scenarios enabled by MADRL for URLLC in the context of GF-NOMA:** In Chapter 4, we highlighted the potential of the MADRL-based GF-NOMA transmission method in meeting the stringent URLLC requirements within uplink massive access scenarios, one of the key objectives in the 6G landscape. This involved our consideration of a GF-NOMA-based uplink URLLC model and the development of a MADRL-driven resource allocation strategy aimed at maximizing energy efficiency while guaranteeing the users' initial URLLC demands. To build upon this research, we can explore a more practical scenario focused on multiple configured-grants (MCG)-based GF-NOMA for URLLC. MCG represents a mechanism proposed by 3GPP to enhance the efficiency of resource allocation in wireless networks, especially in situations characterized by high traffic demands and a need for low-latency communication. The fundamental concept behind MCG lies in the capacity for a base station to grant multiple transmission opportunities to a user equipment (UE) in a concurrent fashion. This simultaneous allocation of resources has the potential to boost the utilization of network resources while mitigating latency. It allows the UE to transmit multiple data packets consecutively without requesting resources individually for each packet. This is particularly useful in applications with stringent latency requirements, such as URLLC. Moreover, the prospect of exploring a distributed cooperative learning strategy among users, based on transfer learning, can be examined to address dynamic scenarios involving active users, which has seen limited comprehensive research to date. In this approach, when a new user joins the network, it can interact with neighboring users to gain insights from their experiences. This not only reduces the learning time for the network but also ensures an effective solution for low-latency ultra-dense systems in dynamic and ever-evolving environments.

- **A rate-splitting-based NOMA for multiplexing diverse services in heterogeneous networks:** In chapter 5, we conducted an investigation into the utilization of power-domain NOMA for multiplexing eMBB, mMTC, and URLLC services. We introduced a novel approach involving joint optimization and cooperative MADRL to achieve optimal SC and power allocation. Our objective was to maximize network EE while accommodating the diverse requirements of users. Notably, recent developments in the field have introduced rate-splitting-based NOMA (RS-NOMA), which promises superior performance compared to conventional NOMA techniques. Given this context, our work in chapter 5 lays the foundation for an extension that considers RS-NOMA in Heterogeneous Networks (HetNets). In RS-NOMA, the basic idea is to divide the data destined for a user into two parts: a common part and a private part. Each part is then allocated specific data rates and assigned to users based on their unique channel conditions and QoS demands. Consequently, the exploration of RS-NOMA-based resource allocation optimization in HetNets presents new challenges that necessitate advanced optimization methods. This aspect represents an open avenue for future research endeavors.
- **Reconfigurable intelligent surfaces (RIS) for URLLC systems:** Reconfigurable Intelligent Surfaces (RIS) have significant potential in enhancing the performance of URLLC systems. RIS represents a transformative technology that leverages programmable metasurfaces to control and manipulate electromagnetic waves, offering unprecedented flexibility and adaptability in wireless communication environments. By intelligently altering the propagation characteristics of signals, RIS enables network operators to mitigate issues like signal blockage, interference, and latency, which are critical concerns in URLLC scenarios. Thus, the combination of RIS and URLLC represents a highly promising research avenue for exploration in the coming years. While significant research efforts have already been conducted in this field, there is a growing need for more comprehensive studies to meet the evolving and diverse requirements, particularly in the context of 6G technology.
- **Machine learning (ML)-based resource management for grant-free massive access in non-terrestrial networks (NTNs):** NTNs are increasingly recognized as a promising technology for future communication networks beyond 5G. Their potential lies in the ability to provide global connectivity, effectively closing the digital divide and facilitating communication in remote or underserved regions. Furthermore, grant-free access could be beneficial for NTNs to reduce their long propagation delay, facilitating their global coverage abilities. Additionally, the integration of RL techniques has emerged as a viable means to intelligently address the complicated resource management challenges in wireless communications. There has been a growing interest in the application of ML-based resource management for NTNs in recent years. However, with the exponential proliferation of wireless devices together

with their diverse QoS requirements in ultra-dense network environments, there is an urgent need for further exploration and the development of advanced optimization methods in this field. Consequently, this research avenue is expected to continue being a hot topic in the coming years.

- **ML for multi-layer ground-air-space networks (MLGASNs) supporting heterogeneous services:** Another potential future research is the investigation of applying ML, such as DRL, for resource and control management in MLGASNs serving heterogeneous services including eMBB, mMTC, and URLLC. MLGASNs represent a cutting-edge approach to wireless communication that integrates terrestrial, aerial, and satellite components into a seamless and interconnected system. These networks exploit the collective capabilities of ground-based infrastructure, drones, and satellite technology to provide robust, ubiquitous, and high-capacity connectivity. In this context, terrestrial networks and unmanned aerial vehicles (UAV) can be applied for delivering various services, especially low-latency applications like URLLC, whereas satellite technology can excel in providing services that demand high data rate and connectivity density, such as eMBB and mMTC. In addition, one of the key enablers for the optimization and efficient operation of MLGASNs is the application of ML. ML techniques play an important role in dynamically managing network resources, mitigating interference, optimizing routing, and ensuring QoS. These technologies enable the network to adapt to changing environmental conditions and traffic patterns, ensuring that each layer operates efficiently and cooperatively to optimize the overall system performance.

Appendix A

# Appendices for Chapter 3

## A.1 Proof for Proposition 1 in Chapter 3

Using (3.3) and (3.5), the CDF of $g_{SH}$ in this case is given by [82]

$$F_{g_{SH}}^{HCS}(x) = \left(1 - \sum_{p=0}^{b_H-1} \frac{m_H^p}{p!\lambda_{SH}^p} x^p e^{-\frac{m_H x}{\lambda_{SH}}}\right)^{a_{H,I}}. \tag{A.1}$$

Applying binomial expansion in [96, Eq. (1.111)], (A.1) can be rewritten as

$$F_{g_{SH}}^{HCS}(x) = 1 + \sum_{p=1}^{a_{H,I}} \phi \underbrace{\left(\sum_{q=0}^{b_H-1} \frac{m_H^q x^q}{q!\lambda_{SH}^q}\right)^p}_{\Psi}, \tag{A.2}$$

where $\phi = \begin{pmatrix} a_{H,I} \\ p \end{pmatrix} (-1)^p e^{-\frac{p m_H x}{\lambda_{SH}}}$.

To derive (A.2), we first resolve $\Psi$ in (A.2) by utilizing the multinomial theorem as follows:

$$\Psi = \sum_{\Delta_H = p} \psi \left[\prod_{q=0}^{b_H-1} \left(\frac{m_H^q}{q!\lambda_{SH}^q}\right)^{\delta_{H,q}}\right] x^{\varphi_H}, \tag{A.3}$$

where $\psi = \begin{pmatrix} p \\ \delta_{H,0}, \ldots, \delta_{H,b_H-1} \end{pmatrix}$.

The final expression of $F_{g_{SH}}^{HCS}(x)$ is achieved in (3.25) by substituting (A.3) into (A.2).

## A.2 Proof of Theorem 1 in Chapter 3

From (3.12), the CDF of $\gamma_{H_{\hat{i}}}^{x_{H_{\hat{i}}}}$ is given by

$$\begin{aligned} F_{\gamma_{H_{\hat{i}}}^{x_{H_{\hat{i}}}}}(x) &= \Pr\left\{\frac{\alpha_{H_{\hat{i}}}\gamma_0 g_{SH}}{\alpha_{L_{\hat{j}}}\gamma_0 g_{SH}+1} < x\right\} \\ &= F_{g_{SH}}(B_x), \end{aligned} \tag{A.4}$$

where (A.4) is obtained under the condition $\alpha_{H_{\hat{i}}} - \alpha_{L_{\hat{j}}} x > 0$ and $B_x = \frac{x}{\gamma_0\left(\alpha_{H_{\hat{i}}} - \alpha_{L_{\hat{j}}} x\right)}$ as defined in (3.30).

By substituting (A.4) into (3.23) and using (3.25), the average BLER at user $H_{\hat{i}}$ in HCS method with TAS/SC or TAS/MRC is expressed as

$$\bar{\varepsilon}_{H_{\hat{i}}}^{HCS} \approx 1 + \chi_{H_{\hat{i}}}\sqrt{N_{H_{\hat{i}}}} \sum_{p=1}^{a_{H,I}} \sum_{\Delta_H = p} \Phi_H c_{H,I} \int_{v_{H_{\hat{i}}}}^{\mu_{H_{\hat{i}}}} B_x^{\varphi_H} e^{-\frac{p m_H B_x}{\lambda_{SH}}} dx, \tag{A.5}$$

To derive the integral in (A.5), we carry out the change of variable by letting $t = B_x$ and (A.5) can be rewritten as

$$\bar{\varepsilon}_{H_{\hat{i}}}^{HCS} \approx 1 + \mathcal{A}_{H,1} \sum_{p=1}^{a_{H,I}} \sum_{\Delta_H = p} \Phi_H c_{H,I} \int_{B_{v_{H_{\hat{i}}}}}^{B_{\mu_{H_{\hat{i}}}}} \frac{t^{\varphi_H} e^{-\frac{p m_H t}{\lambda_{SH}}}}{\left(\frac{1}{\gamma_0 \alpha_{L_{\hat{j}}}} + t\right)^2} dt. \tag{A.6}$$

By letting $u = \frac{1}{\gamma_0 \alpha_{L_{\hat{j}}}} + t$ and using binomial expansion [96, Eq. (1.111)], (A.6) has the following form

$$\bar{\varepsilon}_{H_{\hat{i}}}^{HCS} \approx 1 + \mathcal{A}_{H,1} \widetilde{\sum}_{H,I} c_{H,I} \mathcal{A}_{H,2} \underbrace{\int_{\phi_{H_{\hat{i}}}}^{\kappa_{H_{\hat{i}}}} u^{\hat{\varphi}_H} e^{-\omega_H u} du}_{\mathcal{A}_{H,3}}. \tag{A.7}$$

We derive $\mathcal{A}_{H,3}$ in (A.7) with the aid of [96, Eqs. (3.351.4), (3.352.2), and (3.351.2)] and the final expression of $\bar{\varepsilon}_{H_{\hat{i}}}^{HCS}$ is achieved as in (3.30).

## A.3 Proof of Theorem 2 in Chapter 3

From (3.14) and (3.15), the CDF of $\gamma_{L_{\hat{j}}}^{x_{H_{\hat{i}}}}$ and $\gamma_{L_{\hat{j}}}^{x_{L_{\hat{j}}}}$ are, respectively, given by

$$\begin{aligned} F_{\gamma_{L_{\hat{j}}}^{x_{H_{\hat{i}}}}}(x) &= \Pr\left\{\frac{\alpha_{H_{\hat{i}}}\gamma_0 g_{SL}}{\alpha_{L_{\hat{j}}}\gamma_0 g_{SL}+1} < x\right\} \\ &= F_{g_{SL}}(B_x), \end{aligned} \tag{A.8}$$

and

$$\begin{aligned} F_{\gamma_{L_{\hat{j}}}^{x_{L_{\hat{j}}}}}(x) &= \Pr\left\{\frac{\alpha_{L_{\hat{j}}}\gamma_0 g_{SL}}{\psi\alpha_{H_{\hat{i}}}\gamma_0 g_{SL}+1} < x\right\} \\ &= \begin{cases} F_{g_{SL}}\left(\frac{x}{\alpha_{L_{\hat{j}}}\gamma_0}\right), & \psi = 0 \\ F_{g_{SL}}\left(\hat{B}_x\right), & 0 < \psi \le 1 \end{cases}. \end{aligned} \tag{A.9}$$

To derive $\bar{\varepsilon}_{L_{\hat{j}}}^{HCS}$ in (3.31), we need to resolve $\bar{\varepsilon}_{L_{\hat{j}}}^{x_{H_{\hat{i}}},HCS}$ and $\bar{\varepsilon}_{L_{\hat{j}}}^{x_{L_{\hat{j}}},HCS}$. For $\bar{\varepsilon}_{L_{\hat{j}}}^{x_{H_{\hat{i}}},HCS}$, from (3.26), (3.24), and (A.8), it can be expressed as

$$\bar{\varepsilon}_{L_{\hat{j}}}^{x_{H_{\hat{i}}},HCS} \approx 1 + \chi_{H_{\hat{i}}}\sqrt{N_{H_{\hat{i}}}}\sum_{p=1}^{a_{L,I}}\sum_{\Delta_L=p}\Phi_L c_{L,I}\int_{v_{H_{\hat{i}}}}^{\mu_{H_{\hat{i}}}} B_x^{\varphi_L} e^{-\frac{pm_L B_x}{\lambda_{SL}}} dx. \tag{A.10}$$

After some algebraic manipulations similar to the proof of Theorem 1 in Appendix B, the final expression for $\bar{\varepsilon}_{L_{\hat{j}}}^{x_{H_{\hat{i}}},HCS}$ can be obtained as in (3.31).

For $\bar{\varepsilon}_{L_{\hat{j}}}^{x_{L_{\hat{j}}},HCS}$ in (3.31), we need to derive $\bar{\varepsilon}_{L_{\hat{j}},1}^{x_{L_{\hat{j}}},HCS}$ and $\bar{\varepsilon}_{L_{\hat{j}},2}^{x_{L_{\hat{j}}},HCS}$ to obtain its final expression. Specifically, with the aid of (3.24), (3.26), and (A.9) for $\psi = 0$, $\bar{\varepsilon}_{L_{\hat{j}},1}^{x_{L_{\hat{j}}},HCS}$ can be expressed as

$$\bar{\varepsilon}_{L_{\hat{j}},1}^{x_{L_{\hat{j}}},HCS} \approx 1 + \chi_{L_{\hat{j}}}\sqrt{N_{L_{\hat{j}}}}\sum_{p=1}^{a_{L,I}}\sum_{\Delta_L=p}\frac{\Phi_L c_{L,I}}{\left(\alpha_{L_{\hat{j}}}\gamma_0\right)^{\varphi_L}}\int_{v_{L_{\hat{j}}}}^{\mu_{L_{\hat{j}}}} x^{\varphi_L} e^{-\hat{\omega}_L x} dx. \tag{A.11}$$

By using [96, Eq. (3.351.2)], the integral in (A.11) can be represented as

$$\int_{v_{L_{\hat{j}}}}^{\mu_{L_{\hat{j}}}} x^{\varphi_L} e^{-\hat{\omega}_L x} dx = \hat{\omega}_L^{-\varphi_L - 1} \Xi_{L,4}, \tag{A.12}$$

where $\Xi_{L,4}$ is defined in (3.31). By substituting (A.12) into (A.11), we obtain the final expression for $\bar{\varepsilon}_{L_{\hat{j}},1}^{x_{L_{\hat{j}}},HCS}$ as in (3.31).

By utilizing (3.24), (3.26), and (A.9) for the case $0 < \psi \leq 1$, $\bar{\varepsilon}_{L_{\hat{j}},2}^{x_{L_{\hat{j}}},HCS}$ is given by

$$\bar{\varepsilon}_{L_{\hat{j}},2}^{x_{L_{\hat{j}}},HCS} \approx 1 + \chi_{L_{\hat{j}}} \sqrt{N_{L_{\hat{j}}}} \sum_{p=1}^{a_{L,I}} \sum_{\Delta_L = p} \Phi_L c_{L,I} \int_{v_{L_{\hat{j}}}}^{\mu_{L_{\hat{j}}}} \hat{B}_x^{\varphi_L} e^{-\frac{p m_L \hat{B}_x}{\lambda_{SL}}} dx. \tag{A.13}$$

After some algebraic manipulations similar to the proof of Theorem 1 in Appendix B, the final expression for $\bar{\varepsilon}_{L_{\hat{j}},2}^{x_{L_{\hat{j}}},HCS}$ can be achieved as in (3.31).

# Bibliography


[1] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 10–21, 2011.

[2] G. J. Sutton, J. Zeng, R. P. Liu, W. Ni, D. N. Nguyen, B. A. Jayawickrama, X. Huang, M. Abolhasan, Z. Zhang, E. Dutkiewicz, and T. Lv, "Enabling technologies for ultra-reliable and low latency communications: From PHY and MAC layer perspectives," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2488–2524, 2019.

[3] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultra-reliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[4] 3rd Generation Partnership Project (3GPP), "Study on scenarios and requirements for next generation access technologies," 3GPP, Tech. Rep. TR 38.913, v16.0.0, 2020.

[5] ——, "5G; service requirements for the 5g system," 3GPP, Tech. Rep. TS 22.261, v17.2.0, 2020.

[6] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 124–130, Jun. 2018.

[7] P. Popovski, J. J. Nielsen, C. Stefanovic, E. de Carvalho, E. Strom, K. F. Trillingsgaard, A.-S. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sorensen, "Wireless access for ultra-reliable low-latency communication: Principles and building blocks," *IEEE Netw.*, vol. 32, no. 2, pp. 16–23, Mar. 2018.

[8] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[9] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[10] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, Jun. 2014.

[11] Y. Yu, H. Chen, Y. Li, Z. Ding, and B. Vucetic, "On the performance of non-orthogonal multiple access in short-packet communications," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 590–593, Mar. 2018.

[12] J. Zheng, Q. Zhang, and J. Qin, "Average block error rate of downlink NOMA short-packet communication systems in nakagami-m fading channels," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1712–1716, Oct. 2019.

[13] X. Lai, Q. Zhang, and J. Qin, "Cooperative NOMA short-packet communications in flat Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6182–6186, Jun. 2019.

[14] Z. Wang, T. Lv, Z. Lin, J. Zeng, and P. T. Mathiopoulos, "Outage performance of URLLC NOMA systems with wireless power transfer," *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 380–384, Mar. 2020.

[15] C. Xiao, J. Zeng, W. Ni, X. Su, R. P. Liu, T. Lv, and J. Wang, "Downlink MIMO-NOMA for ultra-reliable low-latency communications," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 780–794, Apr. 2019.

[16] X. Huang and N. Yang, "On the block error performance of short-packet non-orthogonal multiple access systems," in *IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019.

[17] R. Hashemi, S. Ali, N. H. Mahmood, and M. Latva-aho, *IEEE Trans. Veh. Technol.*

[18] T.-H. Vu, T.-V. Nguyen, D. B. d. Costa, and S. Kim, "Intelligent reflecting surface-aided short-packet non-orthogonal multiple access systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4500–4505, 2022.

[19] Y. Xu, C. Shen, D. Cai, and G. Zhu, "Latency constrained non-orthogonal packets scheduling with finite blocklength codes," *IEEE Trans. Veh. Technol.*, pp. 1–5, Jul. 2020, Early Access.

[20] Y. Xu, C. Shen, T. Chang, S. Lin, Y. Zhao, and G. Zhu, "Transmission energy minimization for heterogeneous low-latency NOMA downlink," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1054–1069, Feb. 2020.

[21] C. Li, N. Yang, and S. Yan, "Optimal transmission of short-packet communications in multiple-input single-output systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 7199–7203, 2019.

[22] T.-V. Nguyen, V.-D. Nguyen, D. B. da Costa, T. Huynh-The, R. Q. Hu, and B. An, "Short-packet communications in multihop networks with wet: Performance analysis and deep learning-aided optimization," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 439–456, 2023.

[23] D. Van Huynh, V.-D. Nguyen, S. Chatzinotas, S. R. Khosravirad, H. V. Poor, and T. Q. Duong, "Joint communication and computation offloading for ultra-reliable and low-latency with multi-tier computing," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 2, pp. 521–537, 2023.

[24] C. Pan, H. Ren, Y. Deng, M. Elkashlan, and A. Nallanathan, "Joint blocklength and location optimization for URLLC-enabled UAV relay systems," *IEEE Commun. Lett.*, vol. 23, no. 3, pp. 498–501, Mar. 2019.

[25] H. Ren, C. Pan, Y. Deng, M. Elkashlan, and A. Nallanathan, "Joint power and blocklength optimization for URLLC in a factory automation scenario," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1786–1801, Mar. 2020.

[26] C. Sun, C. She, C. Yang, T. Q. S. Quek, Y. Li, and B. Vucetic, "Optimizing resource allocation in the short blocklength regime for ultra-reliable and low-latency communications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 402–415, Jan. 2019.

[27] N. B. Khalifa, M. Assaad, and M. Debbah, "Risk-sensitive reinforcement learning for URLLC traffic in wireless networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2019, pp. 1–7.

[28] H. Yang, X. Xie, and M. Kadoch, "Intelligent resource management based on reinforcement learning for ultra-reliable and low-latency IoV communication networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4157–4169, 2019.

[29] A. T. Z. Kasgari, W. Saad, M. Mozaffari, and H. V. Poor, "Experienced deep reinforcement learning with generative adversarial networks (GANs) for model-free ultra reliable low latency communication," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 884–899, Feb. 2021.

[30] Y. Liu, Y. Deng, M. Elkashlan, A. Nallanathan, and G. K. Karagiannidis, "Optimization of grant-free noma with multiple configured-grants for murllc," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 4, pp. 1222–1236, 2022.

[31] Y. Liu, Y. Deng, H. Zhou, M. Elkashlan, and A. Nallanathan, "Deep reinforcement learning-based grant-free NOMA optimization for mURLLC," *IEEE Trans. Commun.*, pp. 1–16, 2023, Early Access.

[32] N. B. Khalifa, M. Assaad, and M. Debbah, "Risk-sensitive reinforcement learning for URLLC traffic in wireless networks," in *IEEE WCNC*, Marrakech, Morocco, Apr. 2019, pp. 1–7.

[33] M. Alsenwi, N. H. Tran, M. Bennis, A. Kumar Bairagi, and C. S. Hong, "eMBB-URLLC resource slicing: A risk-sensitive approach," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 740–743, 2019.

[34] P. Korrai, E. Lagunas, S. K. Sharma, S. Chatzinotas, A. Bandi, and B. Ottersten, "A RAN resource slicing mechanism for multiplexing of eMBB and URLLC services in OFDMA based 5G wireless networks," *IEEE Access*, vol. 8, pp. 45 674–45 688, 2020.

[35] A. K. Bairagi, M. S. Munir, M. Alsenwi, N. H. Tran, S. S. Alshamrani, M. Masud, Z. Han, and C. S. Hong, "Coexistence mechanism between eMBB and uRLLC in 5G wireless networks," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1736–1749, 2021.

[36] P. K. Korrai, E. Lagunas, A. Bandi, S. K. Sharma, and S. Chatzinotas, "Joint power and resource block allocation for mixed-numerology-based 5G downlink under imperfect CSI," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1583–1601, 2020.

[37] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55 765–55 779, 2018.

[38] F. Saggese, M. Moretti, and P. Popovski, "NOMA power minimization of downlink spectrum slicing for eMBB and URLLC users," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2022, pp. 1725–1730.

[39] Q. Chen, J. Wu, J. Wang, and H. Jiang, "Coexistence of URLLC and eMBB services in MIMO-NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 839–851, 2023.

[40] Y. Liu, B. Clerckx, and P. Popovski, "Network slicing for eMBB, URLLC, and mMTC: An uplink rate-splitting multiple access approach," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2023.

[41] M. Alsenwi, N. H. Tran, M. Bennis, S. R. Pandey, A. K. Bairagi, and C. S. Hong, "Intelligent resource slicing for eMBB and URLLC coexistence in 5G and beyond: A deep reinforcement learning based approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4585–4600, 2021.

[42] H. Yang, Z. Xiong, J. Zhao, D. Niyato, C. Yuen, and R. Deng, "Deep reinforcement learning based massive access management for ultra-reliable low-latency communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 5, pp. 2977–2990, 2021.

[43] M. Fayaz, W. Yi, Y. Liu, and A. Nallanathan, "A power-pool-based power control in semi-grant-free NOMA transmission," *arXiv preprint arXiv:2106.11190v2*, pp. 1–14, 2022.

[44] A. Alajmi, M. Fayaz, W. Ahsan, and A. Nallanathan, "Semi-centralized optimization for energy efficiency in IoT networks with NOMA," *IEEE Wireless Commun. Lett.*, vol. 12, no. 2, pp. 366–370, 2023.

[45] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proc. IEEE*, vol. 106, no. 10, pp. 1834–1853, 2018.

[46] R. Ali, Y. B. Zikria, A. K. Bashir, S. Garg, and H. S. Kim, "URLLC for 5G and beyond: Requirements, enabling incumbent technologies and network intelligence," *IEEE Access*, vol. 9, pp. 67 064–67 095, 2021.

[47] B. Makki, T. Svensson, and M. Zorzi, "Finite block-length analysis of the incremental redundancy HARQ," *IEEE Wireless Commun. Lett.*, vol. 3, no. 5, pp. 529–532, Oct. 2014.

[48] S. Doğan, A. Tusha, and H. Arslan, "NOMA with index modulation for uplink URLLC through grant-free access," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 6, pp. 1249–1257, 2019.

[49] Y. Liu, Y. Deng, M. Elkashlan, A. Nallanathan, and G. K. Karagiannidis, "Analyzing grant-free access for URLLC service," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 741–755, 2021.

[50] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805–1838, 2020.

[51] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[52] J. Yuan, H. Q. Ngo, and M. Matthaiou, "Machine learning-based channel prediction in massive MIMO with channel aging," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 2960–2973, 2020.

[53] X. Shi, Y. D. Wong, C. Chai, M. Z.-F. Li, T. Chen, and Z. Zeng, "Automatic clustering for unsupervised risk diagnosis of vehicle driving for smart road," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 17 451–17 465, 2022.

[54] H. Liang, X. Zhang, X. Hong, Z. Zhang, M. Li, G. Hu, and F. Hou, "Reinforcement learning enabled dynamic resource allocation in the internet of vehicles," *IEEE Trans. Ind. Inform.*, vol. 17, no. 7, pp. 4957–4967, 2021.

[55] P. Zhang, L. Pan, T. Laohapensaeng, and M. Chongcheawchamnan, "Hybrid beamforming based on an unsupervised deep learning network for downlink channels with imperfect CSI," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1543–1547, 2022.

[56] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction.* The MIT Press, 2018.

[57] F. AlMahamid and K. Grolinger, "Reinforcement learning algorithms: An overview and classification," in *IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, 2021, pp. 1–7.

[58] Y. Liu, W. Yi, Z. Ding, X. Liu, O. A. Dobre, and N. Al-Dhahir, "Developing noma to next generation multiple access: Future vision and research opportunities," *IEEE Wireless Commun.*, vol. 29, no. 6, pp. 120–127, 2022.

[59] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[60] Y. Yu, H. Chen, Y. Li, Z. Ding, L. Song, and B. Vucetic, "Antenna selection for MIMO nonorthogonal multiple access systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3158–3171, Apr. 2018.

[61] 3rd Generation Partnership Project (3GPP), "Study on downlink multiuser superposition transmission (MUST) for LTE," 3GPP, Tech. Rep. 36.859, Jan. 2016.

[62] ——, "Evolved universal terrestrial radio access (E-UTRA); Physical channels and modulation," 3GPP, Tech. Rep. TS 36.211, Apr. 2020.

[63] ——, "Study on non-orthogonal multiple access (NOMA) for NR," 3GPP, Tech. Rep. 38.812, Dec. 2018.

[64] S. K. Sharma and X. Wang, "Toward massive machine type communications in ultra-dense cellular IoT networks: Current issues and machine learning-assisted solutions," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 426–471, Firstquarter 2020.

[65] H. Ren, C. Pan, Y. Deng, M. Elkashlan, and A. Nallanathan, "Joint pilot and payload power allocation for massive-MIMO-enabled URLLC IIoT networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 5, pp. 816–830, May 2020.

[66] ——, "Resource allocation for secure URLLC in mission-critical IoT scenarios," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5793–5807, Sep. 2020.

[67] M. Mousaei and B. Smida, "Optimizing pilot overhead for ultra-reliable short-packet transmission," in *IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017.

[68] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Nonorthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," vol. 53, no. 9, pp. 74–81, Sep. 2015.

[69] L. Lei, L. You, Y. Yang, D. Yuan, S. Chatzinotas, and B. Ottersten, "Load coupling and energy optimization in multi-cell and multi-carrier NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11 323–11 337, Nov. 2019.

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

[71] A. C. Cirik, N. M. Balasubramanya, L. Lampe, G. Vos, and S. Bennett, "Toward the standardization of grant-free operation and the associated NOMA strategies in 3GPP," *IEEE Commun. Stand. Mag.*, vol. 3, no. 4, pp. 60–66, Dec. 2019.

[72] X. Sun, S. Yan, N. Yang, Z. Ding, C. Shen, and Z. Zhong, "Short-packet downlink transmission with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4550–4564, Jul. 2018.

[73] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Open J. Commun. Soc.*, no. 1, pp. 179–189, Jan. 2020.

[74] I. Budhiraja, N. Kumar, and S. Tyagi, "Cross-layer interference management scheme for D2D mobile users using NOMA," *IEEE Syst. J.*, pp. 1–12, Jun. 2020, Early Access.

[75] Z. Chang, L. Lei, H. Zhang, T. Ristaniemi, S. Chatzinotas, B. Ottersten, and Z. Han, "Secure and energy-efficient resource allocation for multiple-antenna NOMA with

wireless power transfer," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 4, pp. 1059–1071, Dec. 2018.

[76] Q. Yu, C. Han, L. Bai, J. Wang, J. Choi, and X. Shen, "Multiuser selection criteria for MIMO-NOMA systems with different detectors," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1777–1791, Feb. 2020.

[77] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "Capacity comparison between MIMO-NOMA and MIMO-OMA with multiple users in a cluster," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2413–2424, Oct. 2017.

[78] Z. Wei, L. Zhao, J. Guo, D. W. K. Ng, and J. Yuan, "Multi-beam NOMA for hybrid mmWave systems," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1705–1719, Feb. 2019.

[79] M. Qiu, Y. Huang, and J. Yuan, "Downlink non-orthogonal multiple access without SIC for block fading channels: An algebraic rotation approach," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 3903–3918, Aug. 2019.

[80] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, Oct. 2004.

[81] Y. Yu, H. Chen, Y. Li, Z. Ding, and L. Zhuo, "Antenna selection in MIMO cognitive radio-inspired NOMA systems," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2658–2661, Dec. 2017.

[82] N. Yang, P. L. Yeoh, M. Elkashlan, R. Schober, and I. B. Collings, "Transmit antenna selection for security enhancement in MIMO wiretap channels," *IEEE Trans. Commun.*, vol. 61, no. 1, pp. 144–154, Jan. 2013.

[83] Z. Ding, L. Dai, and H. V. Poor, "MIMO-NOMA design for small packet transmission in the internet of things," *IEEE Access*, vol. 4, pp. 1393–1405, Apr. 2016.

[84] Z. Ding, H. Dai, and H. V. Poor, "Relay selection for cooperative NOMA," *IEEE Wireless Commun. Lett.*, vol. 5, no. 4, pp. 416–419, Jun. 2016.

[85] D.-D. Tran, D.-B. Ha, V. N. Vo, C. So-In, H. Tran, T. G. Nguyen, Z. A. Baig, and S. Sanguanpong, "Performance analysis of DF/AF cooperative MISO wireless sensor networks with NOMA and SWIPT over Nakagami-m fading," *IEEE Access*, vol. 6, pp. 56 142–56 161, Oct. 2018.

[86] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[87] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K.Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and

future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[88] M. Zeng, A. Yadav, O. A. Dobre, and H. V. Poor, "Energy-efficient joint user-RB association and power allocation for uplink hybrid NOMA-OMA," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5119–5131, Jun. 2019.

[89] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative nonorthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[90] J. Wang, B. Xia, K. Xiao, and Z. Chen, "Performance analysis and power allocation strategy for downlink NOMA systems in large-scale cellular networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3459–3464, Mar. 2020.

[91] N. T. Do, D. B. D. Costa, T. Q. Duong, and B. An, "A BNBF user selection scheme for NOMA-based cooperative relaying systems with SWIPT," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 664–667, Mar. 2017.

[92] D. C. González, D. B. da Costa, and J. C. S. S. Filho, "Distributed TAS/MRC and TAS/SC schemes for fixed-gain AF systems with multiantenna relay: Outage performance," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4380–4392, Jun. 2016.

[93] I. Abu Mahady, E. Bedeer, S. Ikki, and H. Yanikomeroglu, "Sum-rate maximization of NOMA systems under imperfect successive interference cancellation," *IEEE Commun. Lett.*, vol. 23, no. 3, pp. 474–477, 2019.

[94] H. Ren, C. Pan, K. Wang, Y. Deng, M. Elkashlan, and A. Nallanathan, "Achievable data rate for URLLC-enabled UAV systems with 3-D channel model," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1587–1590, Dec. 2019.

[95] X. Zhang, X. Zhou, and M. R. McKay, "Enhancing secrecy with multi-antenna transmission in wireless ad hoc networks," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 11, pp. 1802–1814, Nov. 2013.

[96] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Academic Press, Mar. 2007.

[97] T. T. Nguyen, V. N. Ha, and L. B. Le, "Wireless scheduling for heterogeneous services with mixed numerology in 5G wireless networks," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 410–413, 2020.

[98] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "A novel analytical framework for massive grant-free NOMA," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2436–2449, Nov. 2018.

[99] C. Xiao, J. Zeng, W. Ni, X. Su, R. P. Liu, T. Lv, and J. Wang, "Downlink MIMO-NOMA for ultra-reliable low-latency communications," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 780–794, Apr. 2019.

[100] Z. Wang, T. Lv, Z. Lin, J. Zeng, and P. T. Mathiopoulos, "Outage performance of URLLC NOMA systems with wireless power transfer," *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 380–384, Mar. 2020.

[101] D.-D. Tran, S. K. Sharma, S. Chatzinotas, I. Woungang, and B. Ottersten, "Short-packet communications for MIMO NOMA systems over Nakagami-m fading: BLER and minimum blocklength analysis," *IEEE Trans. Veh. Technol.*, pp. 1–16, Mar. 2021.

[102] 3rd Generation Partnership Project (3GPP), "5G NR, physical layer procedures for data," 3GPP, Tech. Rep. TS 38.214, v15.9.0, Mar. 2020.

[103] H. Ye, G. Y. Li, and B. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[104] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, and X. Zhu, "Unsupervised learning-based fast beamforming design for downlink MIMO," *IEEE Access*, vol. 7, pp. 7599–7605, 2019.

[105] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y. Liang, and D. I. Kim, "Applications of deep reinforcementlearning in communications and networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3133–3174, Fourth quarter 2019.

[106] J. Yu and L. Chen, "Stability analysis of frame slotted aloha protocol," *IEEE Trans. Mobile Comput.*, vol. 16, no. 5, pp. 1462–1474, Jul. 2016.

[107] H. Cao and J. Cai, "Distributed opportunistic spectrum access in an unknown and dynamic environment: A stochastic learning approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4454–4465, Jan. 2018.

[108] M. Shirvanimoghaddam, M. Condoluci, M. Dohler, and S. J. Johnson, "On the fundamental limits of random non-orthogonal multiple access in cellular massive IoT," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2238–2252, Jul. 2017.

[109] S. K. Sharma and X. Wang, "Collaborative distributed Q-learning for RACH congestion minimization in cellular IoT networks," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 600–603, Apr. 2019.

[110] M. V. da Silva, R. D. Souza, H. Alves, and T. Abrão, "A NOMA-based Q-learning random access method for machine type communications," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1720 –1724, Oct. 2020.

[111] D.-D. Tran, S. K. Sharma, and S. Chatzinotas, "BLER-based adaptive Q-learning for efficient random access in NOMA-based mMTC networks," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Helsinki, Finland, Apr. 2021, pp. 1–5.

[112] D.-D. Tran, S. K. Sharma, S. Chatzinotas, and I. Woungang, "Learning-based multiplexing of grant-based and grant-free heterogeneous services with short packets," in *IEEE Global Commun. Conf. (GLOBECOM)*, Madrid, Spain, Dec. 2021.

[113] D.-D. Tran, V. N. Ha, and S. Chatzinotas, "Novel reinforcement learning based power control and subchannel selection mechanism for grant-free NOMA URLLC-enabled systems," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, 2022, pp. 1–5.

[114] Y. Hua, R. Li, Z. Zhao, X. Chen, and H. Zhang, "GAN-powered deep distributional reinforcement learning for resource management in network slicing," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 2, pp. 334–349, Feb. 2020.

[115] R. Huang, V. W. S. Wong, and R. Schober, "Throughput optimization for grant-free multiple access with multiagent deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 228–242, Jan. 2021.

[116] J. Zhang, X. Tao, H. Wu, N. Zhang, and X. Zhang, "Deep reinforcement learning for throughput improvement of the uplink grant-free NOMA system," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6369–6379, Jul. 2020.

[117] M. Fayaz, W. Yi, Y. Liu, and A. Nallanathan, "Transmit power pool design for grant-free NOMA-IoT networks via deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7626–7641, Nov. 2021.

[118] Y. Liu, Y. Deng, M. Elkashlan, and A. Nallanathan, "Cooperative deep reinforcement learning based grant-free NOMA optimization for mURLLC," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2022, pp. 1–6.

[119] S.-Y. Lien, S.-L. Shieh, Y. Huang, B. Su, Y.-L. Hsu, and H.-Y. Wei, "5G new radio: Waveform, frame structure, multiple access, and initial access," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 64–71, 2017.

[120] V. N. Ha, T. T. Nguyen, L. B. Le, and J.-F. Frigon, "Admission control and network slicing for multi-numerology 5G wireless networks," *IEEE Netw. Lett.*, vol. 2, no. 1, pp. 5–9, 2020.

[121] H. Liu, N. I. Miridakis, T. A. Tsiftsis, K. J. Kim, and K. S. Kwak, "Coordinated uplink transmission for cooperative NOMA systems," in *IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[122] Z. Sheng, D. Tian, and V. C. M. Leung, "Toward an energy and resource efficient internet of things: A design principle combining computation, communications, and protocols," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 89–95, 2018.

[123] L. Liang, H. Ye, and G. Y. Li, "Spectrum sharing in vehicular networks based on multi-agent reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2282–2292, 2019.

[124] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G.Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, and G. Ostrovski, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[125] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," *arXiv preprint arXiv:1509.06461*, 2015.

[126] M. H. H. M. L. Z. Wang, T. Schaul and N. Freitas, "Dueling network architectures for deep reinforcement learning," *arXiv preprint arXiv:1511.06581*, 2016.

[127] X. Liu, Y. Liu, Y. Chen, and L. Hanzo, "Enhancing the fuel-economy of V2I-assisted autonomous driving: A reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8329–8342, 2020.

[128] X. Liu, Y. Liu, and Y. Chen, "Machine learning empowered trajectory and passive beamforming design in UAV-RIS wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2042–2055, 2021.

[129] A. Sannai, Y. Takai, and M. Cordonnier, "Universal approximations of permutation invariant/equivariant functions by deep neural networks," *arXiv:1903.01939*, pp. 1–17, 2019.

[130] F. Wu, H. Zhang, J. Wu, and L. Song, "Cellular UAV-to-device communications: Trajectory design and mode selection by multi-agent deep reinforcement learning," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4175–4189, 2020.

[131] A. Nouruzi, A. Rezaei, A. Khalili, N. Mokari, M. R. Javan, E. A. Jorswieck, and H. Yanikomeroglu, "Toward a smart resource allocation policy via artificial intelligence in 6G networks: Centralized or decentralized?" *arXiv preprint arXiv:2202.09093*, pp. 1–8, 2022.

[132] M. Zeng, X. Li, G. Li, W. Hao, and O. A. Dobre, "Sum rate maximization for IRS-assisted uplink NOMA," *IEEE Communications Letters*, vol. 25, no. 1, pp. 234–238, 2021.

[133] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultra-reliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[134] S. K. Sharma and X. Wang, "Toward massive machine type communications in ultra-dense cellular IoT networks: Current issues and machine learning-assisted solutions," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 426–471, Firstquarter 2020.

[135] A. C. Cirik, N. M. Balasubramanya, L. Lampe, G. Vos, and S. Bennett, "Toward the standardization of grant-free operation and the associated NOMA strategies in 3GPP," *IEEE Commun. Stand. Mag.*, vol. 3, no. 4, pp. 60–66, Dec. 2019.

[136] E. N. Tominaga, H. Alves, R. D. Souza, J. Luiz Rebelatto, and M. Latva-aho, "Non-orthogonal multiple access and network slicing: Scalable coexistence of eMBB and URLLC," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, 2021, pp. 1–6.

[137] D.-D. Tran, S. K. Sharma, S. Chatzinotas, I. Woungang, and B. Ottersten, "Short-packet communications for MIMO NOMA systems over Nakagami-m fading: BLER and minimum blocklength analysis," *IEEE Trans. Veh. Technol.*, pp. 1–16, Mar. 2021.

[138] Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski, and H. V. Poor, "Rate-splitting multiple access: Fundamentals, survey, and future research trends," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 2073–2126, 2022.

[139] J. Zhang, L. Lu, Y. Sun, Y. Chen, J. Liang, J. Liu, H. Yang, S. Xing, Y. Wu, J. Ma, I. B. F. Murias, and F. J. L. Hernando, "PoC of SCMA-based uplink grant-free transmission in UCNC for 5G," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1353–1362, 2017.

[140] Z. Ding, R. Schober, P. Fan, and H. V. Poor, "Simple semi-grant-free transmission strategies assisted by non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4464–4478, 2019.

[141] Z. Yang, P. Xu, J. Ahmed Hussein, Y. Wu, Z. Ding, and P. Fan, "Adaptive power allocation for uplink non-orthogonal multiple access with semi-grant-free transmission," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1725–1729, 2020.

[142] C. Zhang, Y. Liu, W. Yi, Z. Qin, and Z. Ding, "Semi-grant-free NOMA: Ergodic rates analysis with random deployed users," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 692–695, 2021.

[143] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for iot: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805–1838, 2020.

[144] Z. Ding, R. Schober, and H. V. Poor, "Unveiling the importance of SIC in NOMA systems-part 1: State of the art and recent findings," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2373–2377, 2020.

[145] M. V. da Silva, R. D. Souza, H. Alves, and T. Abrão, "A NOMA-based Q-learning random access method for machine type communications," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1720–1724, 2020.

[146] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, Jul. 2014.

[147] S. Schiessl, J. Gross, and H. Al-Zubaidy, "Delay analysis for wireless fading channels with finite blocklength channel coding," in *18th ACM Int. Conf. Model., Anal. Simul. Wireless Mobile Syst.*, 2015, p. 13–22.

[148] W. Dinkelbach, "On Nonlinear Fractional Programming," *Management Science*, vol. 13, no. 7, pp. 492–498, March 1967. [Online]. Available: https://ideas.repec.org/a/inm/ormnsc/v13y1967i7p492-498.html

[149] V. N. Ha, D. H. N. Nguyen, and J.-F. Frigon, "System energy-efficient hybrid beamforming for mmwave multi-user systems," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 4, pp. 1010–1023, 2020.

[150] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[151] V. N. Ha and L. B. Le, "End-to-end network slicing in virtualized OFDMA-based cloud radio access networks," *IEEE Access*, vol. 5, pp. 18 675–18 691, 2017.

[152] X. Xi, X. Cao, P. Yang, J. Chen, T. Q. S. Quek, and D. Wu, "Network resource allocation for eMBB payload and URLLC control information communication multiplexing in a multi-UAV relay network," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1802–1817, 2021.

[153] D. Lee, N. He, P. Kamalaruban, and V. Cevher, "Optimization for reinforcement learning: From a single agent to cooperative agents," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 123–135, 2020.

[154] L. Liang, H. Ye, and G. Y. Li, "Spectrum sharing in vehicular networks based on multi-agent reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2282–2292, 2019.

[155] M. Fayaz, W. Yi, Y. Liu, and A. Nallanathan, "Transmit power pool design for grant-free NOMA-IoT networks via deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7626–7641, 2021.

[156] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," *arXiv preprint arXiv:1509.06461*, 2015.

[157] F. S. Melo, "Convergence of Q-learning: A simple proof," Inst. Syst. Robot., Tech. Rep., 2001.

[158] X. Liu, Y. Liu, Y. Chen, and L. Hanzo, "Trajectory design and power control for multi-UAV assisted wireless networks: A machine learning approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7957–7969, 2019.