# Medium Access Design for 5G and Beyond

## Multi-Service Coexistence and Collaboration Strategies

Farnaz Khodakhah


Department of Computer and Electrical Engineering
Mid Sweden University Doctoral Thesis No. 424
Sundsvall, Sweden
2025

Mittuniversitetet
Informationsteknologi och medier
ISBN 978-91-90017-14-2 SE-851 70 Sundsvall
ISNN 1652-893X SWEDEN

Akademisk avhandling som med tillstånd av Mittuniversitetet framlägges till offentlig granskning för avläggande av teknologie doktorsexamen torsdagen den 24 april 2025 i M108, Mittuniversitetet, Holmgatan 10, Sundsvall.



Tryck: Tryckeriet Mittuniversitetet

*To My Parents*

# Abstract

The transition from 5G to beyond 5G presents fundamental challenges in medium access (MA) design, requiring efficient coexistence and collaboration across heterogeneous services. As networks evolve, MA strategies must become spectrally efficient to support diverse applications, including enhanced mobile broadband (eMBB), ultra-reliable low-latency communication (URLLC), and multi-user augmented reality (AR). To address these demands, this thesis proposes three non-orthogonal MA solutions: cooperative relaying for heterogeneous coexistence, collaborative transmissions for multi-user AR, and MA switching mechanisms for balanced multi-service integration.

The first contribution introduces a cooperative relaying framework using non-orthogonal multiple access (NOMA) for primary-secondary coexistence. A novel piecewise-forward NOMA (PF-NOMA) protocol integrates decode-and-forward (DF) and amplify-and-forward (AF) relaying, dynamically adjusting power and time allocation based on real-time conditions.

The second contribution focuses on latency-sensitive AR applications, where high data rates, low latency, and energy efficiency are critical. A NOMA-based rate-distortion optimization framework is proposed, integrating adaptive power allocation to minimize video distortion while optimizing resource utilization.

The third contribution advances radio access network (RAN) slicing by introducing a hybrid MA framework that integrates NOMA, rate-splitting multiple access (RSMA), and puncturing for efficient resource allocation across slices with competing quality of service (QoS) requirements. By balancing age of information (AoI) and throughput, the proposed solution ensures scalable and adaptive resource management for URLLC and eMBB services.

Extensive mathematical modeling and simulation-based evaluations validate the proposed frameworks, demonstrating significant gains in spectral efficiency, reliability, and energy efficiency across diverse deployment scenarios. This work establishes a unified and adaptive MA paradigm, enabling seamless multi-service coexistence, fostering collaboration, and ensuring scalable resource allocation in future 5G and beyond networks.

# Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Mikael Gidlund, for his guidance and support throughout this research. A special thanks to my co-supervisor, Dr. Aamir Mahmood, whose invaluable insights, patience, and continuous encouragement have been instrumental in shaping this work. His dedication and willingness to assist at every stage have made a significant impact on my academic journey. I also extend my appreciation to my co-supervisor, Professor Patrik Österberg, for his constructive feedback and support.

I am especially grateful to Professor Čedomir Stefanović, who supervised me during my eight-month research exchange at Aalborg University. His expertise, guidance, and generous support significantly contributed to my research and professional growth. I truly appreciate the time and effort he dedicated to helping me during this period. I would also like to thank Dr. Hossam Farag for being part of the collaborative and supportive research environment at Aalborg and for co-authoring related publications with me during my stay.

I am immensely grateful to my family for their unwavering support and encouragement. My heartfelt thanks go to my parents, who have always believed in me, my brother, who has been a constant source of motivation, and my uncle, whose wisdom and support have been invaluable. Their love and encouragement have given me the strength to persevere.

I would also like to extend my gratitude to my colleagues Sujash, Fateme, Neda, Mehdi, Duc and Soheib for their collaboration, discussions, and support throughout this journey. Their companionship has made this experience more enriching and enjoyable.

A heartfelt thank you to my wonderful friends from Aalborg — Chiara, Evi, Alejandra, and Malena — for all the laughter, coffee breaks, and support. The moments we shared made my exchange experience truly special, and I am so grateful for the meaningful friendships we built along the way.

Finally, I am deeply grateful to my dear friends Anka and Sabrina for their constant encouragement, patience, and emotional support. Their friendship has been a source of positivity throughout this journey. A special thanks to Joaquin, whose support and presence have meant more than words can express.

To everyone who has supported me in one way or another—thank you.

# Contents

**Abstract** **v**

**Acknowledgements** **vii**

**List of Papers** **xiii**

**Terminology** **xix**

**1 Introduction** **1**

1.1 Problem Formulation . . . . . 4

1.2 Thesis Objective . . . . . 5

1.3 Research Questions . . . . . 6

1.4 Theis Methodology . . . . . 9

1.5 Contributions . . . . . 11

1.6 Outline . . . . . 13

**2 Multiple Access Techniques and RAN Slicing for 5G and Beyond** **15**

2.1 Multiple Access Techniques . . . . . 16

2.2 Non-Orthogonal Multiple Access (NOMA) . . . . . 18

2.2.1 Basic Principles of NOMA . . . . . 18

2.2.2 Variants of NOMA . . . . . 19

2.2.3 Downlink and Uplink NOMA . . . . . 19

2.2.4 Cooperative NOMA (C-NOMA) . . . . . 21

2.2.5 Advancements and Standardization of NOMA . . . . . 23

2.3 Rate Splitting Multiple Access (RSMA) . . . . . 23

2.3.1 Basic Principle of RSMA . . . . . 24

2.3.2 Variants of RSMA . . . . . 24
2.3.3 Downlink and Uplink RSMA . . . . . 25
2.3.4 Advancements and Standardization of RSMA . . . . . 27
2.4 Radio Access Network (RAN) Slicing . . . . . 28
2.4.1 Overview of RAN Slicing . . . . . 29
2.4.2 Integration with Multiple Access Techniques . . . . . 29
2.4.3 Advancements and Standardization of RAN Slicing . . . . . 30
2.5 Age of Information (AoI) . . . . . 31
2.5.1 Advancements and Standardization of AoI . . . . . 33
2.6 Short Summary of Chapter . . . . . 33

**3 Multiple Access Strategies for Cooperative Relaying and AR Applications 35**
3.1 Multiple Access Schemes for Efficient Cooperative Relaying in Shared Spectrum . . . . . 35
3.1.1 Framework and Challenges . . . . . 36
3.1.2 Relaying Protocols . . . . . 37
3.1.3 Optimizing Power and Time Allocation . . . . . 39
3.2 Collaborative Resource Management for AR Applications . . . . . 39
3.2.1 Efficient Collaborative Transmission Strategies with NOMA . . 41
3.2.2 Optimizing Resources with Rate-Distortion Models in AR . . . 42
3.2.3 Energy-Efficiency and Quality Trade-Offs . . . . . 42
3.3 Short Summery of Chapter . . . . . 44

**4 Balancing QoS in RAN Slicing with Advanced Access Techniques 47**
4.1 Hybrid Access Strategies for QoS Optimization in RAN Slicing . . . . 47
4.1.1 Threshold-Based Strategy Selection . . . . . 48
4.1.2 Threshold-Guided Access for Mixed QoS Demands . . . . . 51
4.2 Adaptive User Pairing with advanced MA Choices . . . . . 52
4.2.1 Adaptive Pairing and Access Scheme Selection . . . . . 52
4.2.2 System-Level Analysis of Adaptive Pairing for Heterogeneous QoS in RAN Slicing . . . . . 53
4.3 Short Summary of Chapter . . . . . 55

**5 Summary of Publications 57**
5.1 Paper 1 : NOMA Relaying for Optimal Access and Reliability . . . . . 57

5.1.1 Motivation and Contributions . . . 57
5.1.2 Novelty . . . 57
5.1.3 Limitation . . . 58
5.1.4 Author Contribution . . . 58
5.2 Paper 2: Efficient NOMA Transmission for Industrial AR Systems . . . 58
5.2.1 Motivation and Contributions . . . 58
5.2.2 Novelty . . . 59
5.2.3 Limitation . . . 59
5.2.4 Author Contribution . . . 59
5.3 Paper 3: NOMA or Puncturing for Uplink eMBB-URLLC from an AoI View . . . 59
5.3.1 Motivation and Contributions . . . 59
5.3.2 Novelty . . . 60
5.3.3 Limitation . . . 60
5.3.4 Author Contribution . . . 60
5.4 Paper 4: AoI-Rate Tradeoff for Mission-Critical and eMBB Uplink . . . 60
5.4.1 Motivation and Contribution . . . 61
5.4.2 Novelty . . . 61
5.4.3 Limitation . . . 61
5.4.4 Author Contribution . . . 61
5.5 Paper 5: Adaptive Pairing and Access for Cellular IoT Services . . . 62
5.5.1 Motivation and Contributions . . . 62
5.5.2 Novelty . . . 62
5.5.3 Limitation . . . 63
5.5.4 Author Contributions . . . 63

**6 Conclusions** **65**
6.1 Ethical and Societal Considerations . . . 66
6.2 Future Directions . . . 67

**Bibliography** **69**

# List of Papers

This thesis is mainly based on the following papers, herein referred by their Roman numerals

I **Khodakhah, F.**, Mahmood, A., Österberg, P., & Gidlund, M. (2021). Multiple access-enabled relaying with piece-wise and forward NOMA: Rate optimization under reliability constraints. *Sensors*, 21(14), 4783.

II **Khodakhah, F.**, Mahmood, A., Abedin, S. F., Thar, K., Österberg, P., & Gidlund, M. (2022, December). Design and Resource Allocation of NOMA-based Transmission Scheme for Industrial Collaborative AR. In *2022 IEEE Globecom Workshops (GC Wkshps)* (pp. 1604-1609). IEEE.

III **Khodakhah, F.**, Stefanović, Č., Mahmood, A., Farag, H., Österberg, P., & Gidlund, M. (2023, December). NOMA or Puncturing for Uplink eMBB-URLLC Coexistence from an AoI Perspective?. In *GLOBECOM 2023-2023 IEEE Global Communications Conference* (pp. 4301-4306). IEEE.

IV **Khodakhah, F.**, Mahmood, A., Stefanović, Č., Farag, H., Österberg, P., & Gidlund, M. (2024). Balancing AoI and Rate for Mission-Critical and eMBB Coexistence with Puncturing, NOMA, and RSMA in Cellular Uplink in *IEEE Transactions on Vehicular Technology, vol. 74, no. 1, pp. 1475-1488, Jan. 2025.*

V **Khodakhah, F.**, Mahmood, A., Österberg, P., & Gidlund, M. (2025). Adaptive User Pairing with Non-orthogonal Medium Access Choices for Balanced Coexistence of Mission-Critical and eMBB Services in Cellular IoT. *IEEE Open Journal of the Communications Society* (Under Review)

# List of Figures


1.1 Evolution of mobile communication from 1G to 6G [1]. . . . . 2

1.2 Comparison of key performance requirements between 5G and 6G [2]. 3

1.3 Methodology employed in the research work . . . . . 10

2.1 Downlink and uplink NOMA network [3]. . . . . 20

2.2 Illustration of non-cooperative and cooperative NOMA with a two-user case [4]. . . . . 21

2.3 Performance of Cooperative NOMA transmission [4]. . . . . 22

2.4 Adaptive change in information to be decoded and treated as noise with respect to change in interference levels. . . . . 25

2.5 Downlink RSMA network. . . . . 26

2.6 Two-user Gaussian multiple-access capacity region. $x_1$ and $x_2$ are the information symbols of user-1 and user-2, respectively. $y$ is the received signal. $I(x_2; y)$ is the mutual information between $x_2$ and $y$. $I(x_2; y|x_1)$ is the conditional mutual information between $x_2$ and $y$ given $x_1$ [5, 6]. . . . . 27

2.7 Uplink RSMA network. . . . . 28

2.8 Illustration of the slicing of the wireless resources in a time-frequency frame for supporting the three generic services with: (a) H-OMA (b) H-NOMA. The idle time-frequency blocks are not used for transmission due to absence of traffic. With H-OMA, some of the frequency channels are reserved to URLLC traffic, whereas with H-NOMA the same channels are allocated to both URLLC and eMBB [7]. . . . . 30

2.9 The DTMC model representing the AoI evolution of MC traffic. . . . . 31

2.10 An example of the AoI evolution for the MC user. . . . . 32

3.1 The cooperative network consisted of a primary transmitter (PT) and secondary transmitter (ST). The ST used PF as a relaying protocol. . . . 36

3.2 Time block of cooperative NOMA. . . . . 36
3.3 A comparison of the input output relationship between different relaying protocols. . . . . 37
3.4 A comparison of the locations that PF-NOMA, DF-NOMA and AF-NOMA can cover for the rate greater than or equal to 0.5 bit/s/Hz at the second receiver (SR). . . . . 38
3.5 Maximizing rate at the second receiver (SR) by finding the optimum value for time fraction ($\beta$) and power fraction ($\omega$) in PF-NOMA protocol. 40
3.6 Maximizing rate at the second receiver (SR) by finding the optimum value for time fraction ($\beta$) and power fraction ($\omega$) in AF-NOMA protocol. 40
3.7 Maximizing rate at the second receiver (SR) by finding the optimum value for time fraction ($\beta$) and power fraction ($\omega$) in DF-NOMA protocol. 40
3.8 Two-step NOMA-enabled collaborative AR transmission system. . . . 41
3.9 Relation between distortion (D) and rate (R) for hyperbolic model. . . 42
3.10 Achievable distortion on the common information in C-NOMA and S-NOMA versus allowable maximum distortion on the common information scheme by proposed method for N = 5. . . . . 43
3.11 The energy allocation for proposed method (N = 5). . . . . 44

4.1 System model of eMBB-MC coexistence: (a) Single-cell uplink transmission scenario of eMBB and MC users, (b) Time-frequency resource grid with prescheduled eMBB user and generic overlapping access scheme for a "generate-at-will" MC user, (c) Puncturing-based access, (d) NOMA-based access, and (e) RSMA-based access scheme. . . . . 48
4.2 Evaluation of instantaneous AoI for MC versus minislots index. . . . . 49
4.3 Achievable rate ($R_e$) for eMBB user under NOMA and puncturing as a function of $d_m$ for distance based. . . . . 50
4.4 Probability of successful transmission ($s_m$), average AoI ($\bar{\Delta}$), and PAoI violation probability ($p_v$) for the MC user, as functions of the SNR gap ($\Gamma_d$). . . . . 51
4.5 System model of eMBB-MC coexistence: (a) Single-cell uplink transmission scenario of eMBB and MC users, and (b) Time-frequency resource grid with prescheduled eMBB users ($e_1, e_2, e_3, e_4$) and generic overlapping access scheme for "generate-at-will" MC users ($m_1, m_2, m_3, m_4$). 53
4.6 Performance results of the proposed pairing strategy: *top row*–user distribution scenarios, *middle row*–MC user outage probability as a function of $p_m$, *last row*–eMBB user data rate as a function of SNR gap ($\Gamma_d$). . . . . 55

# List of Tables

1.1 Mapping of Papers to Research Questions . . . . . 12

2.1 Evolution of MA Schemes from 1G to 6G . . . . . 18

# Terminology

## Abbreviations and Acronyms

| | |
|---|---|
| 1G | first-generation |
| 2G | second-generation |
| 3G | third-generation |
| 3GPP | third-generation partnership project |
| 4G | fourth-generation |
| 5G | fifth-generation |
| 6G | sixth-generation |
| AF | amplify and forward |
| AI | artificial intelligence |
| AoI | age of information |
| AR | augmented reality |
| BCs | broadcast channels |
| BPSK | binary phase shift keying |
| BS | base station |
| C-NOMA | cooperative/collaborative non orthogonal multiple access |
| CoMP | coordinated multipoint |
| CR | cognitive radio |
| CSI | channel state information |
| DF | decode and forward |
| D-NOMA | downlink non orthogonal multiple access |
| D-RSMA | downlink rate splitting multiple access |
| D2D | device to device |
| DTMC | discrete time markov chains |
| EF | estimate and forward |
| eMBB | enhanced mobile broadband |
| FDMA | frequency division multiple access |
| GB | gigabytes |
| Gbps | gigabit-per-second |
| GSM | global system for mobile communications |

| | |
|---|---|
| H-NOMA | heterogeneous non-orthogonal multiple access |
| H-OMA | heterogeneous orthogonal multiple access |
| H-RSMA | heterogeneous rate splitting multiple access |
| IoT | Internet of things |
| ITU | international communication union |
| IDMA | interleave division multiple access |
| KPI | key performance indicator |
| LDC | low density signature |
| MA | multiple access |
| MAC | medium access channel |
| MARL | multi-agent reinforcement learning |
| MC | mission critical |
| MIMO | multiple input multiple output |
| ML | machine learning |
| mMTC | massive machine type communication |
| MUST | multi user superposition transmission |
| MC-CDMA | multi carrier code division multiple access |
| NFV | network function virtualization |
| NGMA | next generation multiple access |
| NOMA | non orthogonal multiple access |
| NR | new radio |
| OFDMA | orthogonal frequency division multiple access |
| OMA | orthogonal multiple access |
| O-RAN | open radio access network |
| PAoI | peak age of information |
| PDMA | pattern division multiple access |
| PF | piecewise and forward |
| PHY | physical layer |
| PR | primary receiver |
| QoS | quality of service |
| QoE | quality of experience |
| RAN | radio access network |
| RB | resource block |
| RIS | reconfigurable intelligent surfaces |
| RL | reinforcement learning |
| RSMA | rate splitting multiple access |
| S-NOMA | single non-orthogonal multiple access |
| SC | suer position coding |
| SCA | successive convex approximation |
| SCMA | sparse code multiple access |
| SDMA | spatial division multiple access |
| SDN | software-defined networking |
| SIC | successive interference cancellation |

| | |
|---|---|
| SINR | signal to interference and noise ratio |
| SNR | signal to noise ratio |
| TDMA | time division multiple access |
| THz | terahertz |
| URLLC | ultra reliable low latency communications |
| UAV | unmanned aerial vehicle |
| V2X | vehicle to everything |
| VLC | visible light communication |

## Mathematical Notation

| | |
|---|---|
| $s_i$ | information signal for user $i$ |
| $P_i$ | transmission power allocated to user $i$ |
| $\alpha_i$ | power coefficient for user $i$ |
| $h_i$ | channel coefficient for user $i$ |
| $n_i$ | complex additive gaussian noise |
| $s_c$ | common stream in RSMA |
| $s_P$ | private stream in RSMA |
| $\Gamma_d$ | SNR gap threshold |
| $\bar{\Delta}$ | average age of Information (AoI) |
| $p_v$ | peak age of information (PAoI) violation probability |
| $R$ | data rate |
| $p_m$ | activation probability of MC users |
| $\nu_m$ | transmission rate of MC packets |
| $\zeta$ | fixed packet size in bits |
| $D$ | distortion |
| $K$ | distortion threshold |
| $\tau_{\text{ms}}$ | minislot duration |
| $\mathbb{I}_A$ | indicator function |

# Chapter 1

# Introduction

Wireless systems have been one of the most transformative advancements in modern communication technology. Over the past few decades, wireless communication systems have evolved through five distinct generations, each approximately a decade apart, delivering substantial improvements in speed, reliability, and responsiveness. Today, fifth-generation (5G) networks are revolutionizing connectivity, while ongoing research is shaping the vision of sixth-generation (6G) networks to address emerging challenges and opportunities for digitalization in various industry verticals [8].

The evolution of mobile communication has progressed through multiple generations, each introducing groundbreaking innovations that have transformed wireless connectivity. As illustrated in Fig. 1.1, mobile networks began with analog-based first-generation (1G) technology in the 1980s, supporting basic voice communication with limited capacity. The 1990s marked the transition to digital with second-generation (2G) networks, enhancing voice reliability and introducing SMS messaging. The early 2000s saw the emergence of third-generation (3G) networks, enabling internet access and web applications but facing capacity constraints. The advent of fourth-generation (4G) networks brought high-speed data connectivity, fueling the rise of video streaming, online gaming, cloud computing, and real-time applications [9].

By 2020, global smartphone subscriptions exceeded 7.75 billion, with individual users consuming an average of 10 gigabytes (GB) of data per month. This unprecedented growth, driven by video streaming, Internet of things (IoT) devices, and emerging technologies like virtual and augmented reality (VR/AR), demanded a new generation of wireless networks [8]. The completion of the 5G communication standardization process marked a significant milestone in wireless networks, enabling broad implementation worldwide. To address the diverse and growing demands of modern networks, the international telecommunication union (ITU) defined three core services for 5G, each tied to specific key performance indicators (KPIs): ultra-reliable low-latency communication (URLLC), enhanced mobile broadband (eMBB), and massive machine-type communication (mMTC) [10]. Among

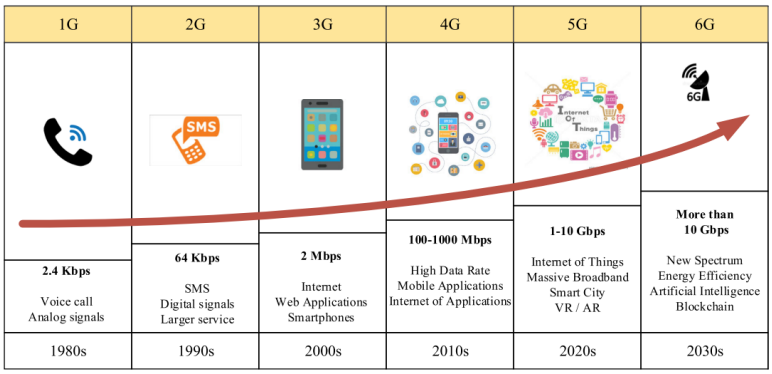



Figure 1.1: Evolution of mobile communication from 1G to 6G [1].

these, URLLC is critical for supporting time-sensitive applications such as autonomous vehicles, remote surgery, and industrial automation, where even delay variation or outdated information can lead to system failures. URLLC achieves low latency with air interface delays reduced to the millisecond level, ensuring real-time data delivery for these critical services. In contrast, eMBB supports ultra-high data rates—reaching up to 10–20 gigabit-per-second (Gbps) in ideal conditions—enabling bandwidth-intensive applications such as 8K video streaming, immersive VR/AR, and cloud gaming,while mMTC focuses on connecting massive numbers of devices within a small area, prioritizing energy efficiency and connection density, as seen in large-scale IoT deployments like smart city sensor networks and industrial automation [11]. Together, these scenarios reflect the diverse performance requirements of modern wireless communication, with URLLC standing out as the cornerstone for enabling reliable, low-latency operation of mission-critical services under dynamic conditions.

Despite the advancements in 5G, spectrum scarcity remains a fundamental challenge. As networks operate at higher frequencies, such as millimeter-wave and terahertz bands, they offer increased bandwidth but suffer from severe propagation limitations, necessitating efficient spectrum utilization strategies [12]. Traditional orthogonal access schemes struggle to meet these demands, making non-orthogonal approaches crucial for maximizing spectral efficiency and enabling seamless multi-service coexistence. Furthermore, energy efficiency is a pressing concern, particularly in ultra-dense networks and massive IoT deployments, where minimizing power consumption while maintaining high spectral efficiency is essential for sustainable operation [13]. Finally, latency and reliability remain critical for mission-critical applications, requiring advanced multiple access techniques that optimize resource allocation while ensuring ultra-low latency and consistent performance under stringent quality of service (QoS) constraints.

6G is expected to bring significant advancements over 5G across multiple key performance indicators, as illustrated in Fig. 1.2. To support advanced applications such as immersive extended reality (XR), mobile holograms, and digital replicas, 6G aims to achieve a peak data rate of 1000 Gbps and a user-experienced rate of 1 Gbps,

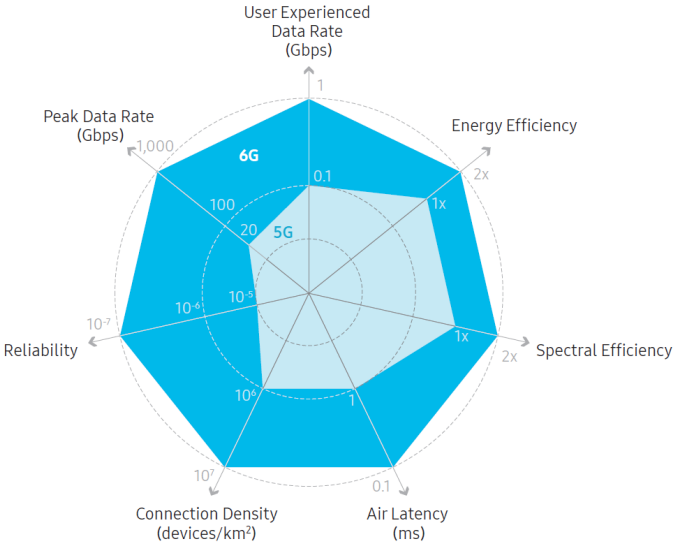



Figure 1.2: Comparison of key performance requirements between 5G and 6G [2].

far surpassing 5G's 20 Gbps peak rate. Spectral efficiency is projected to double, optimizing network resource utilization [12]. Latency improvements are also critical—while 5G offers higher latency, 6G is designed to achieve air latency below 100 $\mu s$, end-to-end (E2E) latency under 1 ms, and user-experienced latency below 10 ms, ensuring seamless responsiveness for real-time applications such as tactile internet and autonomous systems. Additionally, 6G reliability is expected to improve by a factor of 100 over 5G, achieving an error rate of $10^{-7}$ , crucial for applications like industrial automation, emergency response, and remote surgery [12]. The demand for massive machine-type communication will also increase, with 6G supporting $10^7$ devices per square kilometer, ten times higher than 5G, enabling large-scale IoT, smart cities, and automated industries. Moreover, 6G aims to extend network coverage beyond 5G's limitations, accommodating mobile speeds exceeding 500 km/h to support future high-speed transportation [14]. Energy efficiency is another major focus, with 6G targeting at least a twofold improvement over 5G, ensuring lower power consumption while maintaining high performance [13]. Fig. 1.2 provides a comparative visualization of these key enhancements, demonstrating how 6G will surpass 5G in data rate, reliability, latency, efficiency, and connectivity, defining the next era of wireless communication [15].

Central to addressing the challenges of next-generation wireless networks are multiple access (MA) techniques, which enable multiple users or devices to efficiently share limited spectrum resources. Over time, MA methods have evolved to enhance spectrum utilization, energy efficiency, and QoS [16]. As wireless systems grow increasingly complex, advanced MA schemes are essential to meeting the diverse performance requirements of 5G and 6G networks. In this thesis, QoS is defined by ensuring minimum data rate guarantees for users while maintaining a

minimum success probability tailored to the specific performance demands of their applications, including optimizing data freshness, which is quantified through the age of information (AoI) metric. Reliability, particularly in critical applications, is intricately tied to AoI, a key performance indicator that reflects the freshness of received data. Real-time and reliable information delivery is vital for applications such as industrial automation, autonomous vehicles, and remote robotics, where operational safety and performance hinge on the timely arrival of accurate data. Metrics like peak AoI (PAoI) provide insights into worst-case scenarios, serving as a foundation for performance guarantees in time-sensitive systems [17, 18].

## 1.1 Problem Formulation

The rapid evolution of wireless communication technologies, particularly with the advent of 5G and the anticipated deployment of 6G networks, introduces unprecedented opportunities and challenges in multi-service coexistence and collaboration. These next-generation networks must support a diverse range of applications, including IoT, AR, industrial automation, and URLLC, while maintaining high spectral efficiency, energy efficiency, and network reliability [19, 20].

MA design plays a pivotal role in managing the dynamic resource demands of heterogeneous services. Efficient MA strategies must account for the unique requirements of each application while ensuring seamless coexistence and collaboration among various network entities. Conventional orthogonal multiple access (OMA) techniques struggle to cope with the explosive growth in connected devices and data traffic, necessitating the adoption of advanced MA schemes such as non-orthogonal multiple access (NOMA) and rate-splitting multiple access (RSMA) [21, 22]. These techniques enhance spectral efficiency and optimize resource allocation, making them suitable for complex scenarios such as relaying, AR, and radio access network (RAN) slicing.

Relaying remains a fundamental mechanism for extending coverage and improving energy efficiency in 5G and beyond, particularly in dense IoT deployments, remote areas, and environments with poor coverage. In multi-service networks, efficient relay-based communication is essential for ensuring seamless coexistence between diverse applications with varying QoS demands. In smart cities, relays alleviate congestion at base stations (BSs) and facilitate reliable communication between IoT devices, enabling spectrum sharing among latency-sensitive and high-throughput services while reducing energy consumption [23, 24]. Similarly, in industrial automation and remote healthcare, cooperative relaying enhances link reliability and ensures uninterrupted operation of mission-critical services while coexisting with less stringent traffic. However, achieving efficient coexistence in relay-assisted networks requires intelligent relay selection, power control, and scheduling. Advanced MA techniques, such as NOMA, are crucial for optimizing spectrum utilization while maintaining fairness and reliability across heterogeneous service requirements.

AR applications demand ultra-reliable, low-latency, and high-speed communica-

tion, posing significant challenges for existing transmission schemes [25, 26, 27]. In domains such as AR-assisted surgery, industrial maintenance, and immersive education, these challenges are amplified in collaborative AR environments, where multiple users interact in real time and require synchronized, high-fidelity data exchange. Traditional MA strategies struggle to dynamically adapt to the stringent QoS requirements of AR, particularly in resource-constrained environments, leading to increased latency and reduced quality of experience. Collaborative AR further intensifies these demands by requiring consistent synchronization, shared processing, and seamless communication across multiple devices. Advanced MA schemes, particularly NOMA, provide potential solutions by enabling efficient spectrum sharing and interference mitigation, allowing AR applications to maintain real-time responsiveness while ensuring smooth multi-user interaction and immersive experiences [28, 6].

RAN slicing, a cornerstone of 5G and 6G networks, enables virtualized and application specific allocation of network resources [29, 30]. Key use cases include autonomous driving, precision agriculture, and live streaming, each with distinct latency, bandwidth, and reliability requirements. As networks evolve, the demand for multi-service scenarios—where heterogeneous applications such as IoT, AR, URLLC, and eMBB coexist within the same infrastructure—necessitates enhanced RAN slicing frameworks that dynamically allocate resources based on real-time traffic conditions. Managing RAN slices efficiently across access, transport, and core network domains requires intelligent resource allocation strategies that balance performance, scalability, and energy efficiency while ensuring seamless multi-service integration. Existing MA techniques must evolve beyond traditional orthogonal access methods to support the diverse traffic profiles and stringent QoS requirements of multi-service environments, enabling efficient coexistence and resource sharing across different network slices [7, 31, 32].

To address these challenges, this thesis investigates advanced MA design for 5G and beyond, focusing on multi-service coexistence and collaboration strategies. The study explores the integration of NOMA, RSMA, and OMA within relaying, AR, and RAN slicing frameworks to optimize spectral efficiency, energy utilization, and network reliability. By developing a comprehensive MA framework, this research aims to improve resource allocation strategies and foster a more resilient wireless ecosystem capable of supporting next-generation applications. The proposed solutions will contribute to the design of robust, energy-efficient, and highly adaptive wireless systems that drive innovation and performance in future networks.

## 1.2 Thesis Objective

The overarching objective of this research is to enhance the performance, efficiency, and reliability of 5G-and-beyond wireless networks by optimizing MA techniques and addressing key challenges in cooperative relaying, AR, and RAN slicing. Each of these areas presents unique constraints and requirements, necessitating advanced MA strategies to optimize spectral efficiency, energy utilization, and QoS across di-

verse applications.

In relaying-based networks, primary-secondary coexistence introduces spectrum sharing challenges, where a primary system (e.g., licensed users or priority services) must coordinate efficiently with a secondary system (e.g., unlicensed users or non-priority services) to maintain fairness and performance. This research develops a cooperative relaying framework leveraging NOMA to enhance spectral efficiency and reliability. A novel piece-wise and forward NOMA (PF-NOMA) protocol is introduced, integrating DF and AF relaying with dynamic power and time allocation. By adapting to real-time channel conditions, this approach ensures efficient resource utilization, mitigates interference, and improves the coexistence of primary and secondary users in shared-spectrum environments.

For AR and other high-data-rate, latency-sensitive applications, maintaining high throughput, low latency, and energy efficiency is crucial. Traditional orthogonal access schemes struggle to meet these demands in congested wireless environments, necessitating the adoption of non-orthogonal MA strategies. This research develops a NOMA-based rate-distortion optimization framework that dynamically allocates power and resources to minimize video distortion while maximizing spectral efficiency. By intelligently balancing throughput, latency, and energy constraints, this approach ensures seamless AR experiences in bandwidth limited and high interference scenarios, improving QoS for multi-user immersive applications.

As next-generation networks accommodate a broad spectrum of services, from URLLC to mMTC and eMBB, RAN slicing plays a critical role in ensuring efficient and scalable resource management. This work introduces a hybrid MA framework integrating NOMA, RSMA, and puncturing to optimize resource allocation and user pairing across heterogeneous slices. By employing threshold-based selection mechanisms, such as distance-based and SNR gap-based thresholds, the framework dynamically adjusts access strategies to balance AoI and throughput, ensuring fair coexistence between latency-critical and high-data-rate applications. This approach supports diverse use cases, from autonomous vehicle communication to massive IoT deployments, offering scalable and adaptive network slicing solutions.

By integrating advanced multiple access techniques into cooperative relaying, AR transmission, and RAN slicing, this research aims to develop scalable, robust, and adaptable solutions for next-generation wireless networks. The proposed frameworks enhance spectral efficiency, optimize energy utilization, and improve QoS management, enabling 5G/6G networks to support the diverse and dynamic needs of a highly connected, data-driven world.

## 1.3 Research Questions

This thesis seeks to address several critical challenges in optimizing multiple access techniques and resource allocation for 5G and beyond wireless networks, with a specific focus on applications such as cooperative relaying, collaborative AR, and RAN slicing. These challenges are particularly relevant for applications requiring high

data throughput, low latency, and efficient resource utilization, while also ensuring fairness and reliability across diverse services. The research is guided by the following key research questions (RQs) :

## Research Question 1 (RQ1)

*How can advanced MA schemes enhance the spectral efficiency and rate guarantees of cooperative relaying systems in shared-spectrum networks?*

Cooperative relaying plays a vital role in improving the spectral efficiency and coverage of wireless networks, particularly in environments with poor direct links, such as dense IoT deployments and primary-secondary coexistence scenarios. Advanced MA schemes, such as NOMA, provide a promising solution for enabling simultaneous communication between the base station, relays, and end devices, thereby maximizing spectral utilization. However, integrating these schemes into relaying systems requires addressing challenges such as time and power allocation and ensuring fairness between primary and secondary users. This question explores how MA techniques can improve the performance and reliability of cooperative relaying while ensuring equitable resource distribution.

### Research Question 1.A

*What are the optimal strategies for integrating MA schemes, such as NOMA, into cooperative relaying to maximize spectral efficiency?*

Cooperative relaying introduces additional layers of complexity in terms of how users share the spectrum. NOMA, with its power-domain separation, enables primary and secondary users to simultaneously utilize spectrum resources, but it requires precise power allocation to minimize interference. This subquestion focuses on identifying strategies to integrate NOMA into relaying protocols.

### Research Question 1.B

*How can power and time allocation be optimized in relaying to ensure rate parity and rate guarantees in primary-secondary coexistence?*

Ensuring rate parity between primary and secondary users is critical in relaying systems. Power and time allocation strategies must not only enhance the achievable rate but also ensure the reliability of both primary and secondary systems. This question investigates different relaying strategies and power allocation in NOMA-based systems and provides insights into achieving this balance by optimizing resource allocation in dynamic environments while maintaining rate parity.

## Research Question 2 (RQ2)

*How can resource-efficient frameworks leveraging advanced MA techniques enable latency-critical applications, such as AR, to meet stringent QoS requirements in resource-constrained networks?*

AR applications require ultra-reliable communication with low latency and high data rates to overlay real-time virtual content onto real-world environments. These applications are highly data-intensive and impose strict QoS requirements, especially in industrial and collaborative AR scenarios. Advanced MA schemes, such as NOMA, provide the flexibility to dynamically allocate resources, balancing energy consumption and data throughput. This question addresses the challenges of designing resource-efficient frameworks for AR applications while ensuring video quality and low latency in constrained networks.

### Research Question 2.A

*How can NOMA-based frameworks dynamically balance data rates, and energy efficiency for AR applications?*

NOMA's ability to serve multiple users on the same spectrum makes it ideal for AR environments with varying QoS requirements. However, balancing data rates, and ensuring energy efficiency in dynamic conditions is complex. This research question explores the industrial AR frameworks by leveraging NOMA to optimize resource allocation in scenarios with heterogeneous user demands, demonstrating how dynamic adjustments improve communication efficiency.

### Research Question 2.B

*What techniques can be employed to minimize distortion and maximize video quality in collaborative AR systems under varying network conditions?*

Video quality in AR is sensitive to both network conditions and the underlying transmission scheme. Techniques such as rate-distortion modeling and adaptive power control allow for maintaining high-quality video transmission while minimizing distortion. This subquestion focuses on exploring these techniques to improve user experience without overloading the network.

## Research Question 3 (RQ3)

*How can advanced MA schemes optimize resource allocation and QoS management in RAN slicing for heterogeneous services?*

RAN slicing enables the division of network resources into virtual slices, each catering to distinct applications such as eMBB, URLLC, and mMTC. The heterogeneous nature of these applications introduces challenges in managing conflicting QoS requirements, such as low latency for URLLC and high data rates for eMBB.

Advanced MA schemes like NOMA and RSMA provide mechanisms to optimize resource allocation and user pairing within and across slices. This question explores how these schemes can address the scalability, adaptability, and QoS requirements of RAN slicing.

#### Research Question 3.A

*How can NOMA and RSMA improve resource allocation to support diverse application demands, such as eMBB and MC?*

NOMA and RSMA excel at handling heterogeneous traffic by enabling efficient resource sharing among users with different QoS requirements. For example, NOMA's power-domain separation can prioritize low-latency MC traffic while ensuring high throughput for eMBB users.

#### Research Question 3.B

*What are the trade-offs in selecting MA techniques for managing conflicting QoS requirements across heterogeneous network slices?*

The choice of MA technique—whether NOMA, RSMA, or puncturing—depends on the specific trade-offs between freshness of information, data rates, and resource utilization. For instance, RSMA provides flexibility in interference management, making it suitable for applications with overlapping demands. This thesis provides a foundation for understanding these trade-offs, particularly in scenarios where MC and eMBB coexist.

## 1.4 Theis Methodology

The methodology of this thesis is structured to systematically address the research questions and ensure the development of robust and efficient solutions for next-generation wireless networks. The research process is guided by a structured framework, illustrated in the methodology flowchart (Fig. 1.3), which progresses from literature analysis and solution design to modeling, simulation, and validation. The study began with a comprehensive literature review, which focused on multiple access techniques, cooperative networks, collaborative AR, and latency-aware hybrid access schemes. This phase was instrumental in identifying critical knowledge gaps and formulating research questions targeting key challenges such as scalability, spectral efficiency, and reliability in wireless networks. This phase reflects a qualitative approach, as it synthesizes insights from prior studies to develop hypotheses and define the scope of the research.

Building on these insights, innovative frameworks and optimization strategies were proposed, including PF-NOMA schemes for enhancing primary-secondary coexistence, resource-efficient NOMA-based transmission frameworks for collabora-

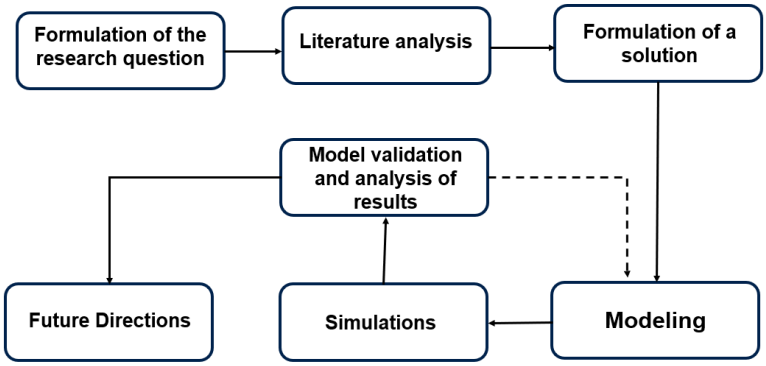



Figure 1.3: Methodology employed in the research work

tive AR, and hybrid access strategies (NOMA, RSMA, and puncturing) to balance trade-offs between age of information (AoI) and eMBB throughput in heterogeneous environments. The research employed quantitative approaches in developing mathematical models and simulation environments to evaluate these solutions. Stochastic models were designed to capture dynamic cooperative relaying for PF-NOMA, rate-distortion models ensured video quality in collaborative AR, and discrete-time markov chain models analyzed AoI and PAoI performance under hybrid access schemes. These models, grounded in mathematical rigor, provided closed-form formulations for performance metrics like spectral efficiency, reliability, and energy consumption.

This research adopts a quantitative approach, with its core methodologies focused on mathematical modeling, optimization, and simulation. Although a thorough literature review was conducted to inform hypothesis formulation and identify gaps in existing systems, this process served as a foundational element of the quantitative framework rather than a qualitative methodological component. The primary focus of the study was the development and validation of models through iterative simulations, which facilitated the precise optimization of critical parameters such as relay placement, power allocation, and access mechanisms.

The validation phase demonstrated the proposed solutions' advantages over baseline approaches through an in-depth quantitative analysis of performance metrics. Finally, the research concluded by synthesizing the findings, including the practical feasibility of PF-NOMA, the scalability of collaborative NOMA for AR applications, and the adaptability of hybrid access strategies to heterogeneous networks. Limitations were acknowledged, and future research directions, such as extending the frameworks to 6G applications and incorporating hardware-based validation, were outlined. This combination of qualitative synthesis for hypothesis generation and quantitative analysis for testing and validation ensures the reliability, robustness, and practical relevance of the contributions. By integrating these paradigms, the research addresses both theoretical and applied aspects of wireless communication research, making it comprehensive and impactful.

## 1.5 Contributions

This thesis makes significant contributions to advancing MA techniques and resource allocation strategies for 5G and beyond wireless networks, addressing critical challenges in cooperative relaying, AR, and RAN slicing. Through a systematic exploration of innovative frameworks and methodologies, the research achieves the following key contributions

1. **Enhancing Cooperative Relaying with Advanced MA Techniques:**
    - The thesis introduces novel cooperative relaying schemes leveraging NOMA to optimize spectral efficiency and reliability in primary-secondary coexistence scenarios. By formulating and solving optimization problems for power and time allocation, the research ensures fairness and performance for both primary and secondary users in shared-spectrum networks.
    - A piecewise relaying protocol is developed, combining decode-and-forward (DF) and amplify-and-forward (AF) schemes to dynamically adapt to varying channel conditions. This hybrid approach improves achievable rate regions and ensures reliable communication under stringent coexistence constraints, as demonstrated in the proposed PF-NOMA framework.
    - These contributions address **RQ1**, **RQ1.A**, and **RQ1.B**, with detailed exploration in *Paper 1*.
2. **Resource-Efficient Frameworks for Latency-Critical AR Applications:**
    - For data-intensive and latency-critical AR applications, the thesis proposes resource allocation frameworks that dynamically optimize power, throughput, and energy efficiency. By leveraging NOMA, the research addresses the unique challenges of balancing high data rates and low latency in resource-constrained environments.
    - Techniques such as rate-distortion modeling and adaptive power allocation are employed to minimize video distortion and ensure high-quality AR experiences. These methods are validated through simulation, demonstrating their effectiveness in industrial and collaborative AR systems.
    - These contributions address **RQ2**, **RQ2.A**, and **RQ2.B**, with detailed exploration in *Paper 2*. *Paper 5* also indirectly supports **RQ2** by addressing resource allocation challenges for latency-critical applications.
3. **Optimizing Resource Allocation in RAN Slicing for Heterogeneous Services:**
    - The research investigates the application of advanced MA schemes, including NOMA and RSMA, in managing RAN slices with diverse QoS requirements.
    - By analyzing trade-offs among latency, throughput, and resource utilization, the research demonstrates how NOMA and RSMA can improve resource allocation and user pairing across heterogeneous slices. These

Table 1.1: Mapping of Papers to Research Questions

| Paper | RQ1 | RQ1.A | RQ1.B | RQ2 | RQ2.A | RQ2.B | RQ3 | RQ3.A | RQ3.B |
|---|---|---|---|---|---|---|---|---|---|
| *Paper 1* | ✓ | ✓ | ✓ | | | | | | |
| *Paper 2* | | | | ✓ | ✓ | ✓ | | | |
| *Paper 3* | | | | | | | ✓ | ✓ | ✓ |
| *Paper 4* | | | | | | | ✓ | ✓ | ✓ |
| *Paper 5* | | | | ✓ | ✓ | | ✓ | ✓ | ✓ |

techniques enable scalability and adaptability in RAN slicing for services ranging from low-latency URLLC to high-throughput eMBB. To achieve optimal performance, threshold-based selection mechanisms, such as distance based and SNR gap based thresholds, play a crucial role in dynamically determining the most suitable access strategy. These thresholds ensure that latency constraints for MC users are met while maximizing data rates for eMBB users, effectively balancing the conflicting QoS requirements in multi-service networks.

- These contributions address **RQ3**, **RQ3.A**, and **RQ3.B**, with insights drawn from *Papers 3, 4,* and *5*. Specifically Paper 3 investigates the trade-offs between NOMA and puncturing for eMBB and URLLC coexistence, analyzing how these access techniques impact performance metrics. Paper 4 explores the balance between AoI and achievable data rates in heterogeneous services by leveraging advanced access strategies such as NOMA and RSMA. Finally, Paper 5 proposes adaptive user pairing strategies to optimize QoS management in RAN slicing, providing scalable solutions for multi-service environments.

This thesis develops a unified MA framework that integrates NOMA, RSMA, and puncturing to enhance multi-service coexistence and collaboration in next-generation wireless networks. By optimizing spectrum utilization, energy efficiency, and reliability, this work enables seamless resource sharing across diverse applications in 5G and beyond. Through rigorous modeling and simulation, the research demonstrates how these techniques can be applied to optimize system performance across diverse scenarios, such as cooperative relaying, AR, and RAN slicing. The unified framework highlights the strengths and trade-offs of each MA scheme, offering practical design guidelines for their implementation. This integration synthesizes findings across all research questions (**RQ1, RQ2, and RQ3**), leveraging insights from all five papers to provide a holistic approach to addressing the demands of heterogeneous wireless systems. The detailed mapping of papers to research questions is summarized in Table 1.1, which illustrates the alignment of the research outcomes with the thesis objectives.

## 1.6 Outline

This dissertation is organized to provide a comprehensive overview of the research conducted. The structure is as follows

- **Chapter 2: Multiple Access Techniques and RAN Slicing for 5G and Beyond.** This chapter provides an overview of the evolution of wireless communication systems, highlighting key principles and techniques that shape 5G-and-beyond networks. It discusses multiple access techniques with a focus on NOMA and RSMA, along with the concept of RAN slicing. These topics frame the challenges of scalability, spectral efficiency, freshness of information, and reliability, offering a comprehensive foundation for the research.
- **Chapter 3: Multiple Access Strategies for Cooperative Relaying and AR Applications.** This chapter provides detailed insights into three key sections: optimizing cooperative relaying in C-NOMA networks, achieving maximum rate in C-NOMA, and resource-efficient transmission strategies for collaborative AR. It highlights the novelty of these approaches in comparison to existing literature.
- **Chapter 4: Balancing QoS in RAN Slicing with Access Techniques.** This chapter explores coexistence strategies for eMBB and MC users in heterogeneous networks, focusing on leveraging hybrid access schemes to balance the freshness of information and throughput trade-offs, adaptive user pairing with non-orthogonal medium access choices, and performance evaluation of the proposed user pairing algorithm.
- **Chapter 5: Summary of Scientific Publications.** This chapter provides a summary of the five scientific publications forming the core of this dissertation, highlighting their contributions to the overall research.
- **Chapter 6: Conclusions and Future Work.** This chapter summarizes the key findings of the research, identifies limitations, and provides guidelines for future exploration and development in the field.

# Chapter 2

# Multiple Access Techniques and RAN Slicing for 5G and Beyond

Since the advent of wireless communications in the late 19th century, wireless technologies have undergone remarkable development, progressing through five generations of innovation. These advancements have transformed communication methods, profoundly impacting human life and reshaping society. At the core of this evolution lies MA, a foundational concept in wireless networks. MA encompasses the techniques that enable multiple users or devices to efficiently share finite radio resources—such as frequency, time, or code—within a network. By allowing simultaneous data transmission and reception while minimizing interference, MA ensures effective resource allocation and interference management. Broadly, existing MA schemes are classified into orthogonal and non-orthogonal transmission strategies, depending on whether multiple users are assigned the same frequency, time, or code resources [28, 4, 21, 16, 33].

As wireless networks transition to the next generation, 6G is envisioned to meet unprecedented demands far exceeding those of its predecessors. These include peak data rates reaching terabits per second, ultra-low latency, exceptional reliability, massive device connectivity, and seamless integration of intelligent functionalities [34, 35]. Additionally, 6G networks aim to connect diverse types of devices, offer ubiquitous coverage, and support highly dynamic use cases across various industries [36, 37]. To achieve these ambitious goals, multiple access will play a critical role by advancing beyond traditional approaches into the realm of next-generation multiple access (NGMA). NGMA encompasses innovative strategies that cater to the stringent performance, scalability, and adaptability requirements of 5G and beyond [38, 39].

Complementing NGMA, RAN slicing has emerged as a transformative concept that enables efficient allocation of network resources. By creating multiple logical

slices within a single physical infrastructure, RAN slicing allows the network to meet the specific needs of diverse services, such as eMBB, mMTC, and URLLC. This innovation ensures that the distinct requirements of modern communication systems are addressed while optimizing resource utilization, making it a cornerstone of future networks. In this thesis, particular focus is placed on RAN slicing between eMBB and mission-critical services, which are closely associated with URLLC. These services exemplify the contrasting requirements of high data throughput and ultra-low latency with extreme reliability, presenting a significant challenge for next-generation wireless systems [40].

This chapter explores the principles, advantages, and variants of advanced MA techniques, focusing on NOMA and RSMA. It also delves into the integration of MA techniques with RAN slicing, highlighting their combined potential to address the diverse and dynamic requirements of eMBB and mission-critical services in 5G and beyond.

## 2.1 Multiple Access Techniques

MA techniques have been a cornerstone of wireless communication systems, evolving continuously to meet growing demands for connectivity, efficiency, and QoS [16]. These techniques have shaped the development of each generation of wireless networks, driving innovation in network design and standardization [41, 42]. By enabling multiple users or devices to share finite radio resources—such as frequency, time, and code—MA facilitates simultaneous data transmission and reception, ensuring efficient resource utilization while minimizing interference.

OMA techniques, such as frequency division multiple access (FDMA), time division multiple access (TDMA), and orthogonal frequency division multiple access (OFDMA), have historically been the backbone of wireless communication systems. FDMA, introduced in 1G of wireless networks, divided the spectrum into frequency bands, allowing each user to occupy a dedicated channel. While this improved system reliability, its analog nature limited scalability and efficiency. With the transition to 2G networks, TDMA emerged as a digital approach that divided resources in the time domain, enabling more efficient spectrum utilization and better user management, especially in global system for mobile communications (GSM) systems [43].

The advent of 3G networks introduced code division multiple access (CDMA), which allowed multiple users to share the same time and frequency resources by employing unique spreading codes for each user [44]. This significantly improved spectral efficiency but required high chip rates, increasing implementation complexity. In the 4G era, OFDMA became the dominant MA technology. By allocating orthogonal subcarriers to users, OFDMA minimized interference and supported high-speed data services, such as video streaming and online gaming [45].

While OMA techniques ensure minimal interference by allocating distinct resources to each user, they are inherently constrained in terms of scalability and resource utilization. NOMA emerged as a solution to these limitations, allowing mul-

tiple users to share the same resource block. NOMA leverages power-domain or code-domain differentiation to distinguish between users. Power-domain NOMA employs superposition coding (SC) at the transmitter, where multiple users' signals are superimposed into a single transmitted signal by assigning different power levels to each user based on their channel conditions. At the receiver, successive interference cancellation (SIC) is used to decode the signals in sequence, starting with the strongest signal (usually assigned the highest power) and subtracting it from the combined received signal before decoding the next. This approach improves spectral efficiency and supports diverse QoS requirements [46]. Code-domain NOMA includes approaches like trellis-coded multiple access (TCMA) [47], interleave-division multiple access (IDMA) [48], and low-density signature (LDS) CDMA [49]. Despite its advantages, NOMA faces challenges in scenarios with high interference and complex user demands, requiring precise power control and interference management.

RSMA has emerged as a promising solution to address the limitations of both OMA and NOMA. RSMA dynamically manages interference through flexible power allocation, bridging the gap between these techniques and spatial division multiple access (SDMA). SDMA utilizes spatial diversity and advanced antenna systems, such as multiple-input multiple-output (MIMO), to serve multiple users simultaneously by separating their signals in the spatial domain. This approach is highly effective in mitigating inter-user interference, improving spectral efficiency, and supporting massive connectivity, making it a cornerstone of 5G and beyond. By adapting to varying interference levels, RSMA can function as either OMA or NOMA depending on the situation, offering unmatched flexibility in managing diverse application demands [39]. This adaptability makes RSMA particularly well-suited for advanced scenarios in 5G and beyond, including IoT, autonomous systems, and edge computing.

The conceptual foundations of NOMA and RSMA can be traced back to pioneering works by Shannon, Ahlswede, and Cover. In the 1960s and 1970s, researchers developed key capacity regions for multiple access channels (MACs) and broadcast channels (BCs) using techniques such as SC and SIC [50, 51]. NOMA was formally introduced in 2013 by Saito et al., who demonstrated its potential to improve system throughput and user fairness over traditional OMA approaches [52]. RSMA, as a natural extension of these advancements, represents the next step in multiple access evolution.

As wireless networks progress towards 6G, there is a clear trend toward adopting non-orthogonal transmission strategies to meet the demands of future systems [39]. schemes like NOMA, RSMA, and SDMA are expected to play pivotal roles in enabling URLLC, mMTC, and eMBB. Table 2.1 summarizes the evolutionary trajectory of MA technologies from 1G to 6G, highlighting their features and limitations. This progression underscores the critical role of MA in shaping modern wireless communication systems and meeting the demands of an increasingly connected world.

Table 2.1: Evolution of MA Schemes from 1G to 6G

| Generation | MA Scheme | Features | Limitations |
|---|---|---|---|
| 1G | FDMA | Analog-based; divides resources in frequency domain | Limited scalability and low spectral efficiency |
| 2G | TDMA | Digital approach; divides resources in time domain | Fixed time slots lead to inefficient utilization |
| 3G | CDMA | Spreading codes; enables shared time and frequency resources | High complexity due to high chip rates |
| 4G | OFDMA | Orthogonal subcarriers; minimizes interference; supports high-speed data | Efficiency constrained by orthogonality limits |
| 5G | NOMA | SC and SIC enhance resource sharing; improves spectral efficiency | Interference-sensitive; requires precise power control |
| 6G | RSMA | Adapts dynamically; bridges OMA, NOMA, and SDMA for scalability | Complexity in managing dynamic resources |

## 2.2 Non-Orthogonal Multiple Access (NOMA)

Power-domain NOMA enhances spectral efficiency and system capacity by enabling multiple users to share the same frequency and time resources with different power levels. This approach optimizes bandwidth utilization and supports massive connectivity, making it a fundamental technique for next-generation wireless networks.

### 2.2.1 Basic Principles of NOMA

At its core, NOMA is built upon two fundamental techniques: SC at the transmitter and SIC at the receiver. These principles, rooted in information theory, have been adapted to meet the needs of modern wireless systems, enabling efficient utilization of resources while maintaining system reliability.

SC was first introduced by Cover in 1972 as a method for achieving the capacity of gaussian broadcast channels [51]. This technique encodes messages for users with poorer channel conditions at lower data rates and then superimposes these signals with those intended for users experiencing better channel conditions. By doing so, multiple users can be simultaneously served within the same resource block, ensuring efficient spectrum utilization without requiring orthogonal separation. The strength of SC lies in its ability to adapt the power allocation among users based on their channel conditions, ensuring that weaker users still receive sufficient signal strength.

On the receiver side, SIC is employed to decode signals while managing interference [53]. SIC operates in a sequential manner, beginning with the user whose signal has the strongest received power. This signal is decoded first because it is least affected by interference. Once decoded, the signal is remodulated and subtracted from the composite received signal, effectively reducing interference for the

subsequent user. This process continues until the weakest user's signal is decoded without interference. By leveraging this step by step decoding and interference management approach, SIC ensures reliable signal detection, even for users with weaker channel conditions.

The combination of SC and SIC allows NOMA to achieve a delicate balance between system performance and computational complexity. By multiplexing users in the power domain, NOMA enhances spectral efficiency while maintaining fairness across users with varying channel conditions.

### 2.2.2 Variants of NOMA

This section examines the main variations of NOMA: downlink NOMA (D-NOMA), which focuses on efficiently managing resources and interference in downlink communication; uplink NOMA (U-NOMA), which tackles the complexities of interference in the uplink; and cooperative NOMA (C-NOMA), which uses collaboration between users to boost coverage.

### 2.2.3 Downlink and Uplink NOMA

In NOMA, power-domain multiplexing and SIC are the key techniques used to facilitate communication between multiple users sharing the same time and frequency resources. Fig. 2.1 illustrates the operation of both downlink and uplink NOMA. In such cases, users are categorized as the strong user (U1) with better channel conditions and the weak user (U2) with poorer channel conditions. While the underlying principles of NOMA remain consistent, the differences in the roles of the BS and users lead to distinct implementations in downlink and uplink scenarios.

#### Downlink NOMA

In D-NOMA, the BS transmits a composite signal containing the superimposed signals intended for multiple users. To ensure fairness and reliable communication, power allocation is used, where weaker users receive a higher power allocation than stronger users. The composite signal can be expressed as

$$x = \sqrt{P_1 \alpha_1} s_1 + \sqrt{P_2 \alpha_2} s_2,$$

where $s_1$ and $s_2$ are the signals for different users, and $\alpha_1 + \alpha_2 = 1$ with $\alpha_1 < \alpha_2$. The received signals at the users can be expressed as

$$y_1 = h_1 x + n_1, \quad y_2 = h_2 x + n_2,$$

where $h_1$ and $h_2$ are the channel coefficients for the users, and $n_1, n_2$ represent noise components.

At stronger users, SIC is performed. The signal for weaker users is first decoded and subtracted from the received composite signal. Once interference from

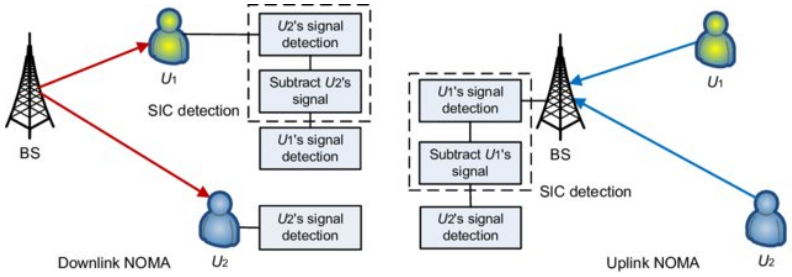



Figure 2.1: Downlink and uplink NOMA network [3].

the weaker user's signal is removed, the stronger user decodes its intended signal. Conversely, weaker users directly decode their signals, treating signals for stronger users as interference. This ensures fairness by allowing weaker users to access the system despite unfavorable channel conditions.

### Uplink NOMA

In U-NOMA, multiple users simultaneously transmit their signals to the BS over the same frequency and time resources. To facilitate SIC at the BS, users transmit at different power levels. Weaker users are assigned a higher transmission power to ensure their signals can be distinguished at the receiver. The received signal at the BS is

$$y = h_1\sqrt{P\alpha_1}s_1 + h_2\sqrt{P\alpha_2}s_2 + n,$$

where $P$ represents the maximum transmission power, and $n$ is noise at the BS.

The BS performs SIC to decode the signals. It first decodes the strongest received signal (usually from the weaker user with higher power allocation) and subtracts it from the total received signal. After removing this interference, the BS decodes the remaining signal. This ensures that weaker users benefit from interference-free transmission, while stronger users' signals are decoded in the presence of residual interference.

### Key Differences Between Downlink and Uplink NOMA

While downlink and uplink NOMA share the foundational principles of power-domain multiplexing and SIC, their implementations exhibit key differences. In downlink NOMA, the BS allocates power to users, with weaker users receiving higher power to ensure reliable communication. Conversely, in uplink NOMA, users control their transmission power levels, with weaker users transmitting at higher power to facilitate SIC at the BS. Another difference lies in where SIC is performed: in downlink NOMA, SIC is carried out at the user devices, with stronger users removing interference from weaker users' signals. In uplink NOMA, SIC is performed at the BS, where the strongest received signal is decoded first, and subsequent signals are recovered iteratively. Additionally, in downlink NOMA, weaker users must decode their signals in the presence of interference, while in uplink NOMA, weaker

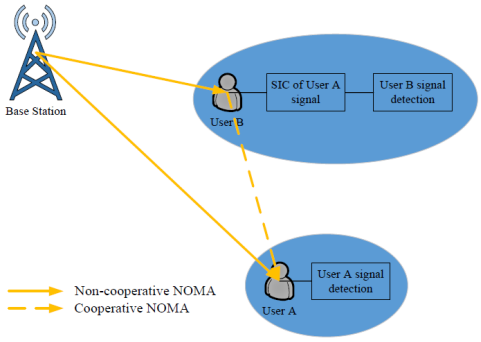



Figure 2.2: Illustration of non-cooperative and cooperative NOMA with a two-user case [4].

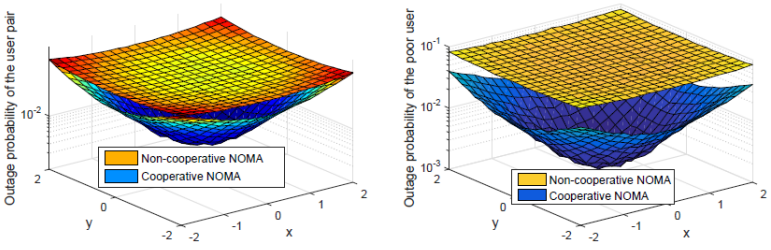
users enjoy interference-free decoding after SIC at the BS. These differences highlight the distinct roles of the BS and users in managing interference and ensuring efficient communication in each scenario [54, 55].

### 2.2.4 Cooperative NOMA (C-NOMA)

C-NOMA integrates the principles of NOMA with cooperative communication to address challenges such as weak channel conditions, limited network coverage, and fairness among users. By leveraging collaboration between users with stronger and weaker channel conditions, C-NOMA enhances network performance, particularly for users at the cell edge. This approach is particularly advantageous in scenarios where users with weaker channel conditions face significant signal degradation [56].

C-NOMA operates through two primary phases: the direct transmission phase and the cooperative transmission phase. During the direct transmission phase, the BS broadcasts a superimposed signal containing messages for both stronger and weaker users, assigning different power levels to their signals based on their channel conditions. Weaker users are assigned higher power to ensure their signals are easier to decode. Stronger users decode not only their own messages but also those intended for weaker users using SIC. In the cooperative transmission phase, stronger users act as relays for weaker users by retransmitting the decoded signals [56]. This process allows weaker users to receive two copies of their messages: one directly from the BS and another via a strong user acting as a relay. Fig. 2.2 illustrates the basic operation of C-NOMA in a two-user downlink scenario.

By employing this collaborative process, C-NOMA significantly improves the reception reliability of weaker users. Since stronger users decode the weaker users' messages during the SIC process, recruiting stronger users as relays is both natural and efficient. This cooperative mechanism enhances network coverage, fairness, and throughput for weaker users without introducing significant system overhead. Weaker users benefit from the additional support provided by stronger users, ensuring more reliable communication even under poor channel conditions. C-NOMA mitigates the imbalance in throughput between stronger and weaker users by redis-




(a) Outage probability of the user pair (b) Outage probability of the poor user

Figure 2.3: Performance of Cooperative NOMA transmission [4].

tributing network resources and leveraging cooperation. The cooperative transmission phase also introduces diversity gain, which helps mitigate multipath fading and improves system robustness. Furthermore, the relaying process is efficient, as strong users decode weaker users' messages during SIC, adding minimal complexity to the system.

Figs. 2.3 (a) and 2.3 (b) illustrate the performance improvements achieved by C-NOMA in terms of outage probability. Specifically, C-NOMA outperforms non-cooperative NOMA by achieving lower outage probabilities for both the user pair and the weaker user, due to the diversity gain provided by the cooperative transmission phase [57].

Despite its advantages, implementing C-NOMA introduces practical challenges. Errors in decoding at the stronger user during SIC can propagate, affecting the relayed signal's accuracy. Coordinating multi-user networks for cooperative NOMA can also increase overhead, particularly in large networks, and the cooperative transmission phase may consume additional time slots, impacting latency. To address these challenges, hybrid multiple access systems have been proposed, incorporating user pairing or grouping to reduce system complexity. In such systems, users within a cell are divided into groups, with cooperative NOMA applied within each group and OMA applied between groups. This hybrid approach balances performance and complexity, achieving significant sum-rate gains, particularly when users with distinctive channel conditions are paired.

C-NOMA has been extensively studied for its ability to optimize system performance in both single-antenna and multi-antenna configurations. Power allocation strategies and relay selection protocols play critical roles in enhancing outage performance and ensuring reliable communication. For instance, two-stage relay selection techniques have been proposed to maximize diversity gain and minimize outage probability.

While C-NOMA faces challenges such as error propagation and system overhead, it remains a promising solution for next-generation wireless networks. By combining cooperative relaying with NOMA, C-NOMA offers a robust framework for improving coverage, fairness, and reliability in scenarios with unfavorable channel conditions.

### 2.2.5 Advancements and Standardization of NOMA

NOMA has gained significant attention for its flexibility and lower complexity compared to alternative non-orthogonal schemes such as sparse code multiple access (SCMA) and pattern division multiple access (PDMA). Despite its conceptual simplicity, NOMA delivers strong performance, making it a promising solution for addressing modern wireless network challenges. Its integration into next-generation networks has been a key focus of standardization efforts, particularly through the 3rd generation partnership project (3GPP).

A major initiative in this direction is the downlink multi-user superposition transmission (MUST) study item, introduced in 3GPP LTE Release 13, which explores multi-user non-orthogonal transmission schemes, advanced receiver designs, and associated signaling mechanisms. MUST classifies non-orthogonal transmission schemes into three categories based on their characteristics: superposition transmission with non-gray-mapped composite constellations, superposition transmission with gray-mapped composite constellations, and superposition transmission with label-bit assignments on composite constellations. Initial evaluations demonstrated throughput gains of nearly 20% in both cell-average and cell-edge scenarios, reinforcing NOMA's potential to enhance network performance and spectral efficiency.

Beyond power-domain NOMA, various other non-orthogonal access techniques extend its core principles. SCMA employs multi-dimensional constellation codebooks and non-orthogonal spreading, enabling massive connectivity and grant-free access. In this approach, each user occupies a subset of orthogonal subcarriers, and a low-complexity message-passing algorithm is used at the receiver for detection. Another promising scheme, MUSA, builds on multi-carrier code division multiple access (MC-CDMA) by leveraging low-correlation spreading sequences and successive interference cancellation (SIC) at the receiver, making it particularly effective in scenarios with high user overloading. PDMA, in contrast, allocates resources across multiple domains, including code, power, and spatial domains, to improve spectral efficiency, reduce latency, and enhance user quality of experience (QoE).

These advanced MA schemes share NOMA's fundamental principle of supporting multiple users within the same channel while ensuring efficient spectrum utilization. SCMA introduces sparsity in subcarrier allocation, allowing multiple users to share the same subcarriers while managing interference effectively. As standardization efforts continue to evolve, the advancements in NOMA and its variants further solidify its role as a key enabler for multi-user access in 5G and beyond, contributing to the development of future wireless communication standards.

## 2.3 Rate Splitting Multiple Access (RSMA)

As wireless communication systems evolve to meet increasing data demands, managing interference efficiently has become a key challenge. Traditional MA techniques such as OMA allocate resources orthogonally, ensuring minimal interference but suffering from spectral inefficiency. Later advancements, including SDMA and NOMA,

addressed some of these limitations. However, SDMA treats interference as noise, which reduces efficiency in high-interference scenarios, while NOMA requires complex decoding through SIC. To address these challenges, RSMA has emerged as a flexible and robust solution, bridging the gap between existing methods and enabling superior interference management.

### 2.3.1 Basic Principle of RSMA

RSMA introduces a versatile approach to interference management by splitting user messages into common parts and private parts. The common part of a message is intended for multiple users and is decoded by all users, while the private part is decoded individually by each user. By combining these two components, RSMA enables a balance between decoding interference and treating it as noise, providing a dynamic solution that adapts to varying interference levels and channel conditions.

In comparison to existing techniques, RSMA stands out for its ability to unify OMA, SDMA, and NOMA into a single framework. It also improves upon NOMA, which decodes interference entirely but introduces significant complexity. RSMA effectively blends these approaches, offering partial interference decoding where only the common message undergoes SIC, while private messages are decoded directly. This reduces decoding complexity compared to NOMA and enhances system flexibility.

Furthermore, RSMA is particularly advantageous in managing varying interference regimes, as shown in Fig. 2.4. In low-interference scenarios, RSMA can behave like SDMA, treating interference as noise. In medium-interference conditions, it partially decodes interference while suppressing the rest. For high-interference scenarios, it approaches NOMA-like performance by decoding interference more aggressively. This adaptability allows RSMA to achieve optimal spectral efficiency across a broad range of user and network conditions.

Despite its clear advantages, RSMA is not without challenges. The performance of RSMA relies on accurate CSI, which can be compromised by feedback delays, user mobility, or estimation errors. Additionally, users must perform SIC to decode the common part of the message, which increases computational complexity, particularly as the number of users grows. Optimizing the decoding order and resource allocation remains crucial for maintaining system performance in practical deployments.

### 2.3.2 Variants of RSMA

RSMA adapts its operation to both downlink and uplink transmissions, offering efficient interference management in each case. Downlink RSMA (D-RSMA) focuses on delivering signals from the base station to users, while Uplink RSMA (U-RSMA) handles interference in user-to-base station transmissions. Although U-RSMA is within the scope of this thesis, D-RSMA is briefly described for the sack of completeness.

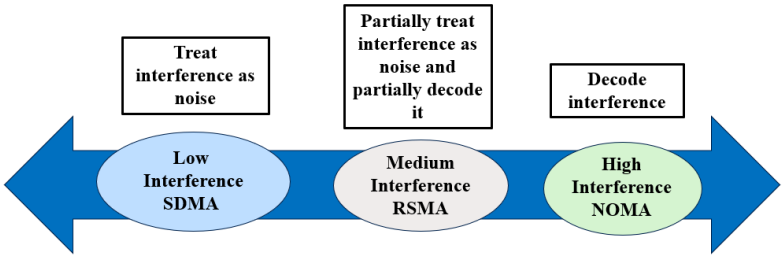



Figure 2.4: Adaptive change in information to be decoded and treated as noise with respect to change in interference levels.

### 2.3.3 Downlink and Uplink RSMA

The differences between downlink and uplink RSMA implementations arise from the distinct roles of the BS and users.

#### Downlink RSMA

In downlink RSMA, the BS transmits a signal that includes both common and private message components. As illustrated in Fig. 2.5, the process begins with a message splitter at the BS, which divides the messages for users into common and private parts. For a two-user scenario, the transmitted signal is expressed as

$$x = \sqrt{P_c}s_c + \sqrt{P_1}s_{p,1} + \sqrt{P_2}s_{p,2},$$

where $s_c$ is the common message intended to be decoded by both users, while $s_{p,1}$ and $s_{p,2}$ are private messages for user 1 and user 2, respectively. The total transmission power $P$ is allocated such that $P_c + P_1 + P_2 = P$.

At the receiver side, the process for decoding the received signal is detailed in Fig. 2.5. Both users first decode the common message $s_c$, treating the private messages as noise. The received signal for user $i$ can be expressed as

$$y_i = h_i x + n_i = h_i(\sqrt{P_c}s_c + \sqrt{P_1}s_{p,1} + \sqrt{P_2}s_{p,2}) + n_i,$$

where $h_i$ is the channel coefficient for user $i$, and $n_i$ is the additive gaussian noise.

After successfully decoding $s_c$, the users subtract it from their respective received signals using SIC. Each user then decodes its respective private message. For example, user 1 decodes $s_1$ after removing the common message, while treating $s_2$ as noise. Once the decoding is complete, the decoded private and common messages are recombined to reconstruct the original message for each user.

This hierarchical decoding structure ensures that RSMA efficiently utilizes the downlink capacity region while managing interference effectively.

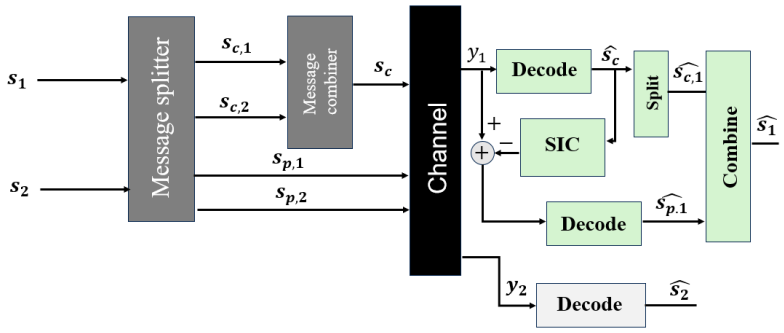



Figure 2.5: Downlink RSMA network.

### Uplink RSMA

In uplink RSMA, users split their messages into sub-messages, enabling efficient utilization of the gaussian multiple access channel (MAC) capacity region. This technique avoids the limitations of traditional approaches, such as time-sharing and joint decoding, which either introduce high overhead or are impractical due to complexity. RSMA allows all points within the MAC capacity region, including intermediate points along the line segment $A$–$B$ shown in Fig. 2.6, to be achieved without relying on time-sharing.

For a two-user uplink scenario, the transmitted signals from the users are

$$x_1 = \sqrt{P_{1,1}}s_{1,1} + \sqrt{P_{1,2}}s_{1,2}, \quad x_2 = \sqrt{P_2}s_2,$$

where user 1 splits its message into two sub-messages, $s_{1,1}$ and $s_{1,2}$, with respective power allocations $P_{1,1}$ and $P_{1,2}$, while user 2 transmits a single message $s_2$ with power $P_2$.
At the BS, the received signal is expressed as

$$y = h_1x_1 + h_2x_2 + n = h_1(\sqrt{P_{1,1}}s_{1,1} + \sqrt{P_{1,2}}s_{1,2}) + h_2\sqrt{P_2}s_2 + n,$$

where $h_1$ and $h_2$ are the channel coefficients for user 1 and user 2, respectively, and $n$ represents additive gaussian noise.

The BS decodes the signals sequentially using SIC. Higher-priority sub-messages, such as $s_{1,1}$, are decoded first, treating other components as noise. Once a sub-message is decoded, it is subtracted from the received signal, enabling the recovery of lower-priority components like $s_{1,2}$ and $s_2$. This hierarchical decoding process ensures that the BS can achieve any point within the MAC capacity region efficiently.
Fig. 2.7 illustrates the uplink RSMA process, showing how message splitting, power allocation, and SIC decoding enable efficient interference management and reliable communication. RSMA effectively balances spectral efficiency and practicality in uplink scenarios, making it a key solution for next-generation wireless systems.

RSMA's ability to balance common and private message components makes it an attractive solution for next-generation wireless systems. By adapting to various

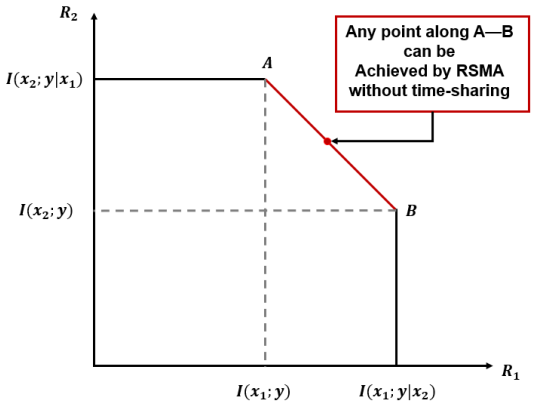



Figure 2.6: Two-user Gaussian multiple-access capacity region. $x_1$ and $x_2$ are the information symbols of user-1 and user-2, respectively. $y$ is the received signal. $I(x_2;y)$ is the mutual information between $x_2$ and $y$. $I(x_2;y|x_1)$ is the conditional mutual information between $x_2$ and $y$ given $x_1$ [5, 6].

channel conditions and user requirements, RSMA ensures robust interference management, high spectral efficiency, and effective utilization of the capacity region in both downlink and uplink scenarios.

### 2.3.4 Advancements and Standardization of RSMA

RSMA has emerged as a transformative MA technique, addressing many of the limitations faced by traditional MA schemes such as SDMA and NOMA. By introducing a dynamic strategy for managing interference, RSMA provides a robust, flexible, and high-performing solution for next-generation communication systems, including 6G networks.

One of RSMA's most significant advantages lies in its inherent robustness to imperfect CSI, a persistent challenge in practical wireless networks. Achieving perfect CSI is nearly impossible due to factors such as pilot contamination, feedback delays, overhead constraints, and user mobility. These challenges often degrade the performance of conventional schemes like SDMA and NOMA, which heavily rely on accurate CSI to manage interference effectively. In contrast, RSMA dynamically balances interference suppression and partial decoding, allowing it to operate reliably even with imperfect CSI. This makes RSMA particularly suitable for real-world multi-antenna systems, where acquiring precise channel information is a persistent issue [58].

Another key strength of RSMA is its flexibility in managing interference across a wide range of network conditions. Traditional schemes take rigid approaches to interference: SDMA treats all interference as noise, which becomes inefficient under high-interference regimes, while NOMA requires full decoding of interference using SIC, introducing significant decoding complexity and potential error propagation.

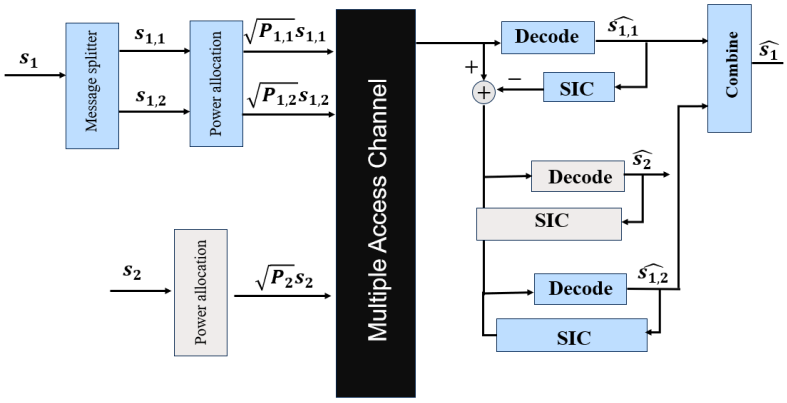



Figure 2.7: Uplink RSMA network.

RSMA, however, adopts a hybrid approach, enabling partial decoding of interference while treating the remaining interference as noise. This ability to seamlessly transition between treating interference as noise and decoding it allows RSMA to adapt to varying interference levels and channel conditions, ensuring optimal performance in both low- and high-interference scenarios. This versatility makes RSMA far more efficient and adaptable than its predecessors [59].

RSMA's generalized framework further sets it apart from SDMA and NOMA, which operate at fixed extremes of interference management. By splitting messages into common and private parts, RSMA allows the common part to be decoded by multiple users, while the private part is decoded individually. This mechanism not only improves spectral efficiency but also enhances interference management compared to conventional schemes. RSMA inherently encompasses and outperforms SDMA and NOMA as special cases, positioning itself as a superior and more versatile framework for modern wireless networks.

Despite its demonstrated advantages in academic and research studies, RSMA has not yet been officially incorporated into 3GPP standards as of January 2025. RSMA has been extensively studied for its potential to enhance spectral and energy efficiency in wireless communications. For instance, [60] highlights RSMA's advantages over traditional methods like SDMA, currently used in 3GPP's 5G new radio (NR) and IEEE 802.11 standards. While RSMA has shown significant promise, its adoption into 3GPP standards would require comprehensive evaluations, stakeholder consensus, and alignment with existing technologies. The standardization process is intricate and involves multiple stages of review and approval, which may delay the integration of RSMA into future standards.

## 2.4 Radio Access Network (RAN) Slicing

RAN slicing is a transformative concept in next-generation wireless communication systems, enabling efficient allocation of network resources by creating multiple log-

ical slices within a single physical infrastructure. Each slice is tailored to the specific requirements of different services, such as eMBB, mMTC, and URLLC. This innovation allows for the coexistence of diverse services while optimizing resource utilization and maintaining service-level isolation [7, 6, 5].

### 2.4.1 Overview of RAN Slicing

In scenarios where devices with varying service requirements connect to a shared BS, resource allocation becomes a complex task due to the lack of coordination among users. RAN slicing addresses this challenge through orthogonal and non-orthogonal slicing strategies, as illustrated in Fig. 2.8.

Orthogonal slicing, depicted in Fig. 2.8 (a), allocates distinct frequency channels to eMBB, mMTC, and URLLC services. For instance, URLLC transmissions, which prioritize low latency and high reliability, are distributed across several frequency channels. This ensures their stringent requirements are met but may lead to underutilization of resources during periods of inactivity. Intra-service resource sharing, such as within eMBB users, is often implemented using OMA techniques [7].

In contrast, non-orthogonal slicing, shown in Fig. 2.8 (b), dynamically reuses frequency resources across services. For example, frequency resources assigned to URLLC or mMTC may also be allocated to eMBB users. This approach, referred to as heterogeneous non-orthogonal multiple access (H-NOMA), improves spectral efficiency by leveraging the intermittent nature of mMTC and URLLC traffic. However, it introduces additional interference among services, requiring sophisticated interference management techniques [7].

### 2.4.2 Integration with Multiple Access Techniques

Integrating RAN slicing with advanced multiple access techniques enhances its flexibility and efficiency. Three prominent approaches are heterogeneous orthogonal multiple access (H-OMA), H-NOMA, and heterogeneous rate-splitting multiple access (H-RSMA).

- **H-OMA:** In this method, resources are divided orthogonally among different services, maintaining isolation between eMBB, mMTC, and URLLC. Intra-service resource sharing within each slice can use OMA or NOMA for efficiency. While this ensures reliable performance, resource utilization can be suboptimal due to idle resources during low traffic periods [7].

- **H-NOMA:** This approach enables the same frequency resources to be shared across slices. NOMA-based resource sharing within eMBB slices allows simultaneous transmission by employing power-domain multiplexing and SIC. Dynamic adaptation to traffic demands significantly improves spectral efficiency but requires advanced interference management [6].

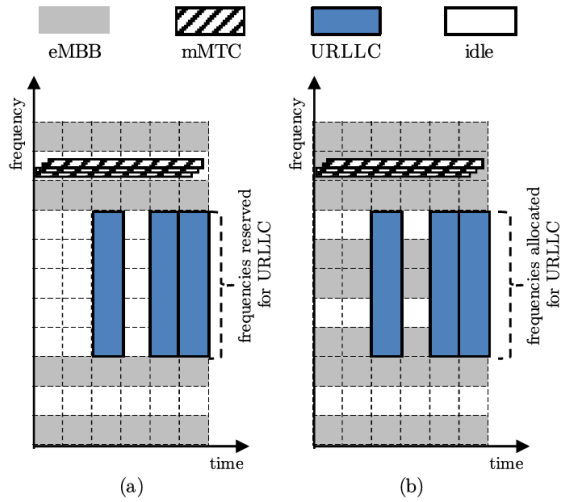



Figure 2.8: Illustration of the slicing of the wireless resources in a time-frequency frame for supporting the three generic services with: (a) H-OMA (b) H-NOMA. The idle time-frequency blocks are not used for transmission due to absence of traffic. With H-OMA, some of the frequency channels are reserved to URLLC traffic, whereas with H-NOMA the same channels are allocated to both URLLC and eMBB [7].

- **H-RSMA:** H-RSMA extends the benefits of NOMA by incorporating message splitting and rate splitting. Each user's message is divided into common and private components, allowing concurrent transmission across slices with minimal interference. By dynamically allocating power to these components and leveraging SIC for decoding, H-RSMA achieves improved spectral efficiency and interference management compared to H-NOMA. It is particularly advantageous in scenarios where user channel conditions vary significantly, as it provides greater flexibility in handling interference and optimizing performance [6].

### 2.4.3 Advancements and Standardization of RAN Slicing

As mobile networks transition toward 5G and beyond, the standardization of RAN slicing has become a key focus to enable seamless interoperability and optimized network efficiency. A significant milestone in this direction is the work undertaken by the 3GPP, particularly within the scope of 5G system architecture standardization. The specifications introduced in 3GPP Release 15 laid the foundational framework for network slicing, defining core principles for slice orchestration, slice selection, and resource isolation. Subsequent releases, including Release 16 and Release 17, have refined RAN slicing mechanisms by introducing enhanced slice-aware radio resource management, inter-slice coordination, and dynamic slice scaling to accommodate varying network demands [61, 62].

Beyond 3GPP, initiatives from other standardization bodies, such as the open

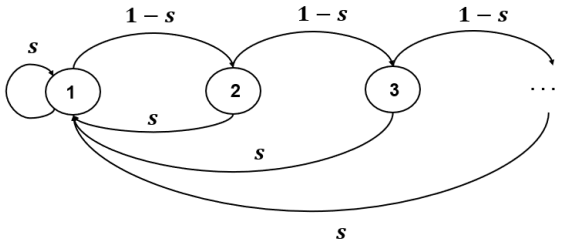



Figure 2.9: The DTMC model representing the AoI evolution of MC traffic.

radio access network (O-RAN) Alliance and the ITU, have further contributed to the evolution of RAN slicing. The O-RAN architecture introduces open interfaces and intelligent control mechanisms, leveraging real-time analytics and AI-driven automation to optimize slice performance. Additionally, advancements in software-defined networking (SDN) and network function virtualization (NFV) have played a crucial role in enabling flexible and scalable RAN slicing implementations [63].

As research and industry collaborations continue, RAN slicing is expected to drive significant advancements in network efficiency, enabling operators to deliver customized services with stringent performance requirements. Future developments will focus on enhancing cross-domain slicing, improving end-to-end slice orchestration, and integrating AI-powered automation to further streamline slice management, reinforcing RAN slicing as a fundamental enabler of next-generation mobile networks [64].

## 2.5 Age of Information (AoI)

AoI is a key performance metric for time-sensitive applications, particularly in URLLC and mission-critical services. AoI quantifies the freshness of information by measuring the time elapsed since the generation of the last successfully received packet. This metric is especially relevant for applications requiring real-time updates, such as industrial automation and autonomous systems [6, 65].

Several models have been proposed to study the AoI in different communication scenarios. These include the generate-at-random model, where updates arrive stochastically; time-slotted or scheduled models, which assume updates are sent at fixed or coordinated intervals; energy harvesting models, where update transmissions are constrained by energy availability; queue-based models, which consider the impact of waiting times; and multi-source models, involving multiple updating nodes. Among these, the generate-at-will model, where the source can generate and transmit updates at any chosen time, offers the most control and is widely used in theoretical analysis. In this thesis, we adopt the generate-at-will model as our framework, and therefore provide a detailed explanation of its characteristics and relevance in the following sections.

The evolution of AoI in generate at will model can be visualized in Fig. 2.10. The

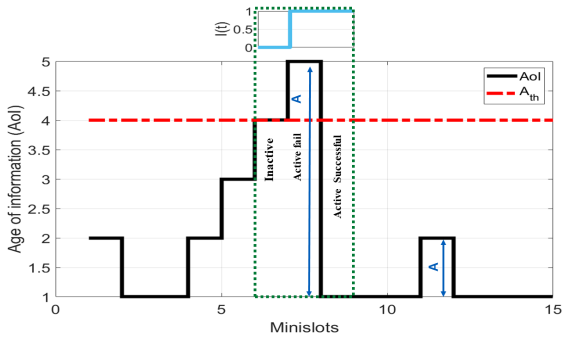



Figure 2.10: An example of the AoI evolution for the MC user.

instantaneous AoI at any time is defined as

$$\Delta(t+1) = \begin{cases} \Delta(t)+1, & \text{if no new packet is successfully received at } t, \\ 1, & \text{if a new packet is successfully received at } t. \end{cases}$$

Among AoI-based metrics, the average AoI provides a long-term measure of freshness, reflecting the overall performance of the update mechanism. On the other hand, the peak AoI captures the worst-case scenario just before a new update is received, making it valuable for systems that are sensitive to stale data. Therefore, using both average and peak AoI offers a comprehensive view of system timeliness and serves as a robust QoS metric in the context of this thesis.

The **average AoI**, denoted as $\overline{\Delta}$, is calculated as the long-term time average of the instantaneous AoI. It provides a measure of the typical freshness of information in a system. Mathematically, it can be expressed as

$$\overline{\Delta} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \Delta(t).$$

This metric is influenced by the packet arrival rate, transmission success probability, and service prioritization policies. A lower average AoI indicates fresher updates, which is critical for mission-critical services. Efficient scheduling and resource allocation mechanisms can significantly reduce the average AoI by prioritizing time-sensitive packets and ensuring reliable transmission [6, 5].

The **Peak AoI (PAoI)** violation probability, denoted as $p_v$, measures the likelihood that AoI exceeds a predefined threshold $A_{th}$ [66]. Mathematically, it can be expressed as

$$p_v = \mathbb{P}(A > A_{th}),$$

where $A$ represents the PAoI.

Additionally, the success probability of each user, denoted as $s$, plays a crucial role in determining AoI metrics. It is defined as the probability that the achievable data rate $R$ exceeds the required transmission rate $v$

$$s = \mathbb{P}(R \geq v).$$

A higher success probability implies more reliable communication, leading to lower AoI and improved service performance.

The evolution of AoI can also be modeled using a discrete-time markov chain (DTMC) [67]. Each state in the DTMC corresponds to a specific AoI value, and transitions between states depend on the packet arrival process and the success of packet transmissions. If a packet is not successfully transmitted, the AoI increases by one unit, corresponding to a transition to the next state. Conversely, if a packet is successfully transmitted, the AoI resets to one, corresponding to a transition to the initial state. Fig. 2.9 highlights periods of inactivity, failed transmissions, and successful updates, illustrating the dynamic behavior of AoI. The DTMC framework enables precise analysis of average AoI and PAoI violation probability, helping design resource allocation strategies that minimize AoI and ensure reliable communication.

For mission-critical services, maintaining low AoI and minimizing PAoI violations are critical to ensuring timely and reliable communication [6].

### 2.5.1 Advancements and Standardization of AoI

While the concept of AoI has gained significant attention in academic research over the past decade, its standardization as a formal QoS metric is still in its early stages. However, there is a growing recognition in both academia and industry of its relevance, particularly for latency-sensitive and real-time applications such as industrial automation, autonomous systems, and IoT networks. Organizations like the IEEE and 3GPP have begun to explore AoI-related metrics in the context of URLLC and mission-critical services. This trend indicates a shift toward integrating AoI into formal performance benchmarks, which may eventually lead to its inclusion in networking standards and protocol designs. The use of AoI in this thesis aligns with this broader direction, highlighting its practical importance beyond theoretical studies.

## 2.6 Short Summary of Chapter

This chapter explored advanced MA techniques and RAN slicing, emphasizing their roles in 5G and beyond. It covered OMA, NOMA, and RSMA, highlighting NOMA's spectral efficiency through SIC and RSMA's superior interference management. The analysis positioned RSMA as a flexible and efficient solution for future networks. A key insight is that no single MA technique is universally optimal; hybrid approaches integrating OMA, NOMA, and RSMA are essential for maximizing efficiency, fairness, and scalability.

The discussion also examined RAN slicing, enabling network resource partitioning for diverse services like eMBB, URLLC, and mMTC. H-OMA, H-NOMA, and H-RSMA were introduced, with H-RSMA recognized for its balance of spectral efficiency and interference management. However, key challenges remain in cooperative relaying and AR applications, including interference mitigation, fair resource allocation, and energy-efficient transmission. NOMA's interference in relaying net-

works demands advanced power control, while AR applications require low-latency, high-quality transmission under strict energy constraints. Addressing these issues is crucial for optimizing NOMA-based cooperative relaying and AR communication.

The next chapter focuses on performance evaluation and optimization of NOMA-based cooperative relaying and AR applications. Through mathematical modeling and simulations, it will explore power allocation, interference management, and energy-efficient transmission strategies to enhance spectral efficiency and system reliability. These insights will guide the practical implementation of advanced MA techniques in real-world wireless networks.

# Chapter 3

# Multiple Access Strategies for Cooperative Relaying and AR Applications

This chapter explores advanced multiple access techniques for cooperative relaying and AR applications in next-generation wireless networks. It focuses on optimizing resource allocation, interference management, and access coordination to enhance efficiency, scalability, and reliability in 5G and beyond. The proposed strategies address the unique challenges of multi-service coexistence, ensuring seamless collaboration among heterogeneous network users. Specifically, the discussions and solutions presented directly contribute to addressing RQ.1 and RQ.2, providing a framework for dynamic spectrum sharing, adaptive relaying protocols, and QoS-aware access mechanisms. These innovations lay the groundwork for next-generation communication systems, supporting high-data-rate, low-latency, and ultra-reliable applications such as AR-driven services.

## 3.1 Multiple Access Schemes for Efficient Cooperative Relaying in Shared Spectrum

Cooperative relaying has emerged as a cornerstone of modern wireless communication, providing enhanced coverage and spectral efficiency, particularly in shared-spectrum environments where primary and secondary systems coexist. In such scenarios, licensed (primary) users and unlicensed (secondary) users share spectrum resources, introducing challenges such as interference management, fairness, and resource optimization. Addressing these challenges is critical for next-generation wireless networks, including 5G and beyond, which must support massive connectivity, low latency, and reliable communication.

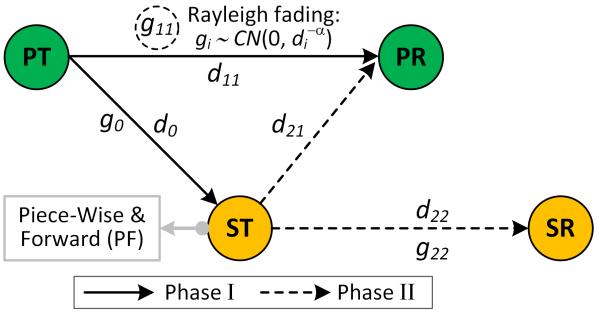



Figure 3.1: The cooperative network consisted of a primary transmitter (PT) and secondary transmitter (ST). The ST used PF as a relaying protocol.

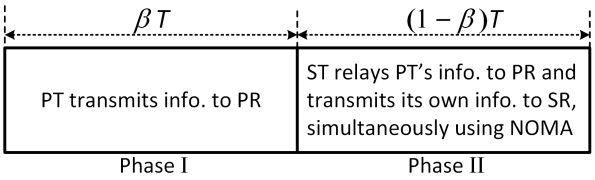



Figure 3.2: Time block of cooperative NOMA.

Advanced MA schemes, such as NOMA, offer a promising solution to these challenges by enabling multiple users to share the same frequency and time resources simultaneously. By integrating NOMA with cooperative relaying, wireless networks can improve spectral utilization, coverage, and overall reliability. However, integrating NOMA into relaying systems introduces complexities such as optimal power and time allocation, fairness assurance, and protocol selection, which must be carefully addressed to ensure robust system performance. This chapter explores these challenges and provides insights into strategies for optimizing cooperative relaying systems in shared-spectrum networks, addressing **RQ.1**, **RQ.1.A**, and **RQ.1.B**.

### 3.1.1 Framework and Challenges

The integration of cooperative relaying with NOMA addresses critical challenges in resource management and coverage extension, making it a vital component of modern wireless systems. Cooperative relaying improves communication reliability by enabling nearby devices to assist in data transmission, providing spatial diversity that mitigates fading, shadowing, and path loss. On the other hand, NOMA enhances spectral efficiency by allowing simultaneous transmissions through power-domain multiplexing.

Fig. 3.1 depicts a typical shared-spectrum cooperative relaying system, where the secondary transmitter (ST) not only relays data for the primary transmitter (PT) but also transmits its own data to the secondary receiver (SR). The two-phase structure of this system is further illustrated in the time-block diagram (Fig. 3.2). During Phase I, the PT communicates directly with the primary receiver (PR) over a fraction of

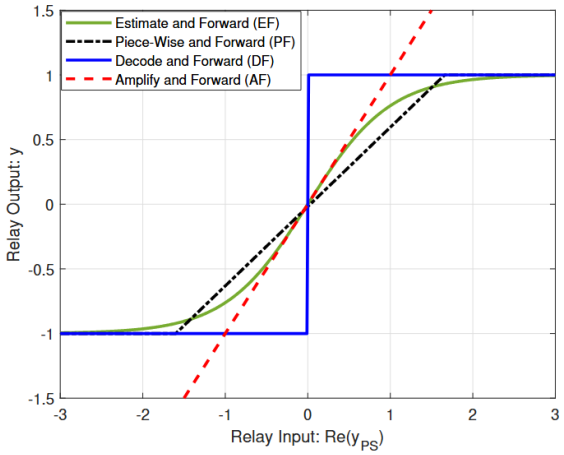



Figure 3.3: A comparison of the input output relationship between different relaying protocols.

the time ($\beta T$). During Phase II, the ST relays the PT's data to the PR and transmits its own data to the SR using NOMA. This dual functionality ensures efficient spectrum utilization and improved coverage for both primary and secondary users, as discussed in [3].

This framework provides the foundation for addressing **RQ.1** by integrating NOMA into cooperative relaying systems and exploring their potential to enhance efficiency and reliability in shared-spectrum environments.

### 3.1.2 Relaying Protocols

To gain deeper insight into the role of relaying in cooperative NOMA systems, it is essential to examine the fundamental characteristics of various relaying protocols. Each protocol differs in how it processes and forwards signals, leading to distinct performance profiles in terms of complexity, latency, and robustness to channel conditions. Fig. 3.3 illustrates the input-output behavior of four key protocols amplify and forward (AF), decode and forward (DF), estimate and forward (EF), and piecewise and forward (PF) which form the basis of this discussion. By analyzing these protocols, we can better understand how their inherent trade-offs influence the design of efficient and reliable cooperative communication systems. The following sections provide a detailed explanation of each protocol and its relevance in the context of integrating NOMA into relaying frameworks.

**Amplify-and-Forward (AF)**: In the AF protocol, the relay amplifies the received signal and forwards it to the receiver without performing any decoding. As illustrated by the red dashed line in Fig. 3.3, AF exhibits a linear relationship between the relay input (Re($y_{PS}$)) and the output ($y$). While AF is computationally simple and introduces minimal delay, it amplifies noise along with the signal, making it less reliable in low-SNR scenarios. The simplicity of AF makes it a viable option in scenarios where computational efficiency is prioritized over reliability [68].

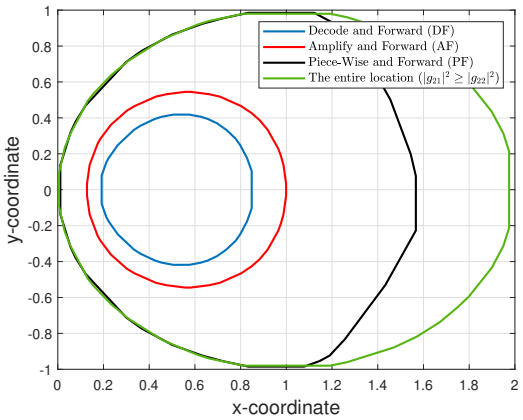



Figure 3.4: A comparison of the locations that PF-NOMA, DF-NOMA and AF-NOMA can cover for the rate greater than or equal to 0.5 bit/s/Hz at the second receiver (SR).

**Decode-and-Forward (DF)**: In the DF protocol, the relay fully decodes the received signal, processes it, and then re-encodes it for transmission. The blue step-like curve in Fig. 3.3 represents the behavior of DF. Unlike AF, DF eliminates noise at the relay stage, ensuring reliable signal transmission to the destination. However, this comes at the cost of higher computational complexity and latency due to the decoding and re-encoding process. DF is particularly effective in high-SNR scenarios, where the relay can decode the signal with minimal errors [69, 70].

**Estimate-and-Forward (EF)**: EF, represented by the green curve in Fig. 3.3, represents the optimal input-output relationship for a relay in a memoryless network. While EF achieves the best possible performance in terms of maximizing the SNR at the destination, its implementation is computationally prohibitive due to the hyperbolic tangent function involved in the relay output calculation [71]. **Piecewise-and-Forward (PF)**: PF strikes a balance between the simplicity of AF and the reliability of DF by approximating the idealized EF protocol using a piecewise-linear approach. The black dashed line in Fig. 3.3 shows how PF divides the relay input into three regions, enabling a computationally efficient approximation of EF. PF's piecewise-linear relationship ensures robustness in varying SNR regimes, making it well-suited for dynamic and heterogeneous network conditions. This protocol offers an attractive compromise between complexity and performance, making it a practical choice for real-world implementations [72].

Fig. 3.4 illustrates the regions where PF-NOMA, DF-NOMA, and AF-NOMA can support a transmission rate of at least 0.5 bit/s/Hz at the SR, whose position is shown in Fig. 3.1. Among the schemes, PF-NOMA achieves the widest coverage, indicating its superior ability to maintain reliable communication under varying channel conditions. This comparison highlights the robustness and adaptability of PF-NOMA, especially in dynamic environments where effective interference and fading management are essential.

### 3.1.3 Optimizing Power and Time Allocation

**Power Allocation**: Power allocation plays a crucial role in ensuring fairness and reliability in shared-spectrum networks, directly addressing **RQ.1.B**. NOMA's power-domain multiplexing allows stronger users to receive lower power while allocating higher power to weaker users. Fig. 3.5, 3.6, and 3.7 provide insights into the impact of power allocation on the achievable rate ($R_s$) at the SR. For PF-NOMA, the optimal power fraction ($\omega$) dynamically adjusts to maximize the secondary user's rate while ensuring reliable communication for the primary user. In contrast, AF-NOMA and DF-NOMA exhibit less flexibility, with DF requiring higher power for the primary signal to ensure decoding reliability [73].

**Time Allocation**: Time allocation is another critical factor in NOMA-based relaying systems, contributing to fairness and efficiency. The time-block diagram in Fig. 3.2 highlights the division of the communication process into two phases. By optimizing the time fraction ($\beta$) between Phase I (direct PT-to-PR transmission) and Phase II (relaying and secondary transmission), PF-NOMA achieves minimal time fractions for relaying, allowing more time for simultaneous transmission and improving overall efficiency. These strategies directly address **RQ.1.B**, ensuring fairness and reliability in primary-secondary coexistence [3].

Integrating advanced MA schemes, such as NOMA, into cooperative relaying systems offers significant benefits in terms of spectral efficiency, coverage, and reliability. Among the examined relaying protocols, PF-NOMA stands out as the most adaptable and efficient, combining the strengths of AF and DF while avoiding their limitations. Its ability to dynamically adjust power and time allocation ensures optimal performance under diverse channel conditions, making it a versatile solution for shared-spectrum networks.

By addressing **RQ.1**, **RQ.1.A**, and **RQ.1.B**, this section highlights the potential of PF-NOMA to meet the demands of next-generation wireless systems, particularly in scenarios requiring fairness between primary and secondary users.

## 3.2 Collaborative Resource Management for AR Applications

AR applications demand low latency, high throughput, and reliable communication to ensure seamless user experiences, particularly in collaborative environments such as industrial automation, telemedicine, and smart manufacturing. These demands introduce significant challenges in wireless networks, especially with constraints on spectrum availability and energy efficiency. Advanced MA strategies, specifically NOMA, have emerged as a promising solution. NOMA enables multiple users to share the same spectrum resources simultaneously, addressing the complex requirements of AR applications.

This section explores how NOMA-based frameworks can dynamically balance data rates and energy efficiency while maintaining high-quality video transmission

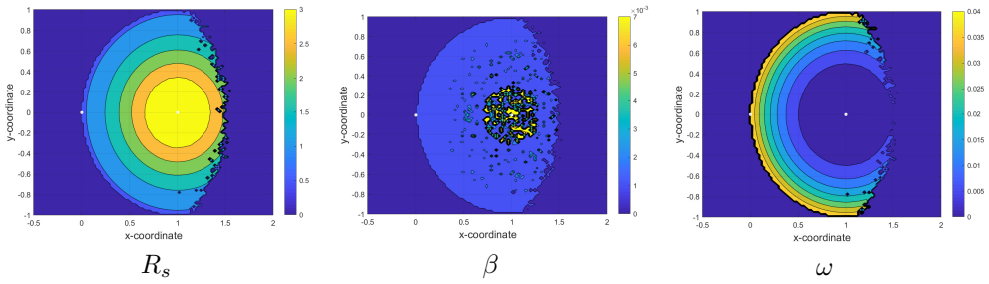



Figure 3.5: Maximizing rate at the second receiver (SR) by finding the optimum value for time fraction ($\beta$) and power fraction ($\omega$) in PF-NOMA protocol.

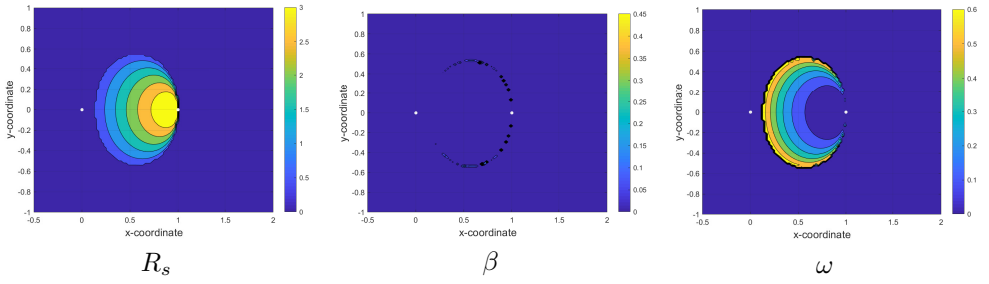



Figure 3.6: Maximizing rate at the second receiver (SR) by finding the optimum value for time fraction ($\beta$) and power fraction ($\omega$) in AF-NOMA protocol.

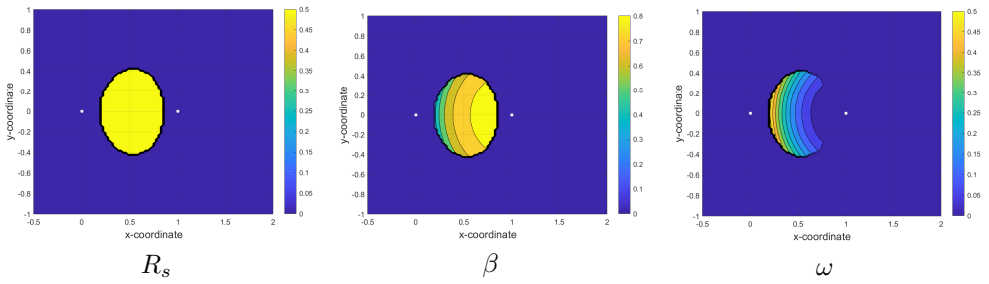



Figure 3.7: Maximizing rate at the second receiver (SR) by finding the optimum value for time fraction ($\beta$) and power fraction ($\omega$) in DF-NOMA protocol.

under varying network conditions. It addresses **RQ2.A** and **RQ2.B**. Through collaborative transmission frameworks, rate-distortion optimization, and energy-efficient resource management, this chapter demonstrates the potential of NOMA-based solutions for AR applications in dynamic and resource-constrained environments.

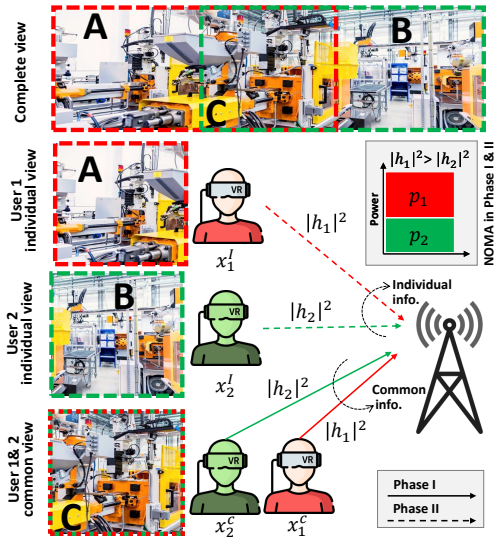



Figure 3.8: Two-step NOMA-enabled collaborative AR transmission system.

### 3.2.1 Efficient Collaborative Transmission Strategies with NOMA

Collaborative NOMA transmission frameworks play a critical role in addressing the stringent resource requirements of AR applications. These frameworks allow multiple AR users to share spectrum resources efficiently while ensuring both low latency and high throughput.

Fig. 3.8 illustrates a two-phase collaborative NOMA system tailored for industrial AR scenarios. In the first phase, known as *collaborative uplink transmission*, AR users jointly transmit shared (common) information to the BS using NOMA. This approach leverages channel conditions to optimize the transmission of overlapping views or shared data. In the second phase, referred to as *individual content transmission*, each AR user sends unique (individualized) data to the BS, ensuring that personalized requirements are met without compromising spectral efficiency.

The system operates under a block fading channel model, where the channel conditions remain constant during a single time slot but vary independently across slots. The BS employs *SIC* to decode user signals, prioritizing users with higher channel gains ($|h_1|^2 > |h_2|^2$). This decoding strategy minimizes interference and maximizes throughput, making the framework particularly effective for latency-sensitive industrial AR tasks such as machine maintenance or real-time monitoring.

By dynamically allocating resources across shared and individual transmissions, this collaborative NOMA framework addresses **RQ2.A** by enabling the system to adapt to varying user demands while optimizing both data rates and energy efficiency. This adaptability ensures resource-efficient communication, even in scenarios with heterogeneous QoS requirements.

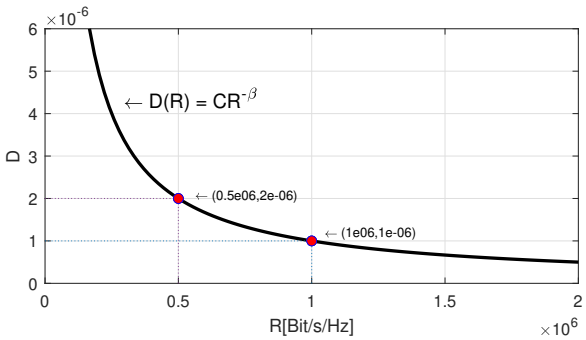



Figure 3.9: Relation between distortion (D) and rate (R) for hyperbolic model.

### 3.2.2 Optimizing Resources with Rate-Distortion Models in AR

Efficient resource utilization is vital for maintaining high-quality communication in AR applications, particularly when balancing data rates with acceptable levels of distortion. Rate-distortion theory provides the foundation for optimizing resource allocation in such scenarios, addressing **RQ2.B** by minimizing distortion and maintaining video quality under varying network conditions.

The relationship between distortion ($D$) and the data rate ($R$) is defined by the rate-distortion function:

$$D(R) = CR^{-\beta},$$

where $C$ and $\beta$ are system-dependent parameters. For simplicity, $C = 1$ and $\beta = 1$ are assumed, resulting in $D(R) = R^{-1}$. As shown in Fig. 3.9, this model allows dynamic adjustments in resource allocation to achieve specific distortion targets. For instance, achieving a distortion level of $2 \times 10^{-6}$ requires a data rate of $0.5 \times 10^{6}$ [bits/s/Hz], while stricter distortion levels demand higher rates. These insights guide power allocation strategies to ensure high-quality shared information transmission without overloading the network.

By optimizing both power and spectral resources for shared and individual content, rate-distortion models enable AR systems to deliver high-quality collaborative content while minimizing resource usage. This section highlights the critical role of rate-distortion optimization in balancing communication quality and resource efficiency, directly addressing **RQ2.B**.

### 3.2.3 Energy-Efficiency and Quality Trade-Offs

Balancing energy consumption with communication quality is a fundamental challenge in collaborative AR scenarios, particularly in NOMA-enabled 5G-and-beyond networks. Efficient resource allocation strategies are essential to ensure scalable and reliable communication while minimizing resource usage. Figs. 3.10 and 3.11 illustrate how advanced multiple access and resource management techniques address

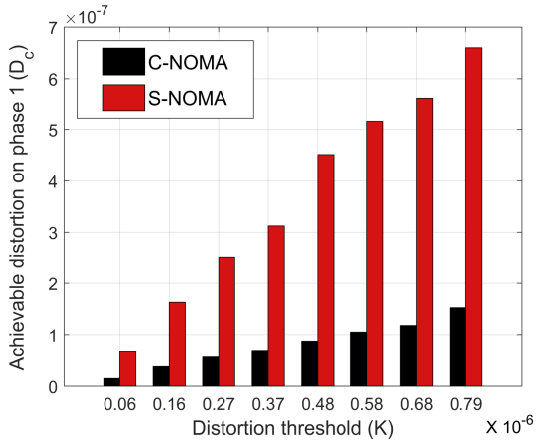



Figure 3.10: Achievable distortion on the common information in C-NOMA and S-NOMA versus allowable maximum distortion on the common information scheme by proposed method for N = 5.

these trade-offs, focusing on optimizing performance in data-intensive and latency-sensitive applications.

Fig. 3.10 examines the transmission quality of common information ($D_c$) for C-NOMA and S-NOMA schemes, addressing the research question: **RQ2**. C-NOMA significantly reduces distortion in common information by $76.9\% - 77.28\%$ compared to S-NOMA. This improvement stems from C-NOMA's ability to prioritize users with the best channel conditions for transmitting the majority of common information. By focusing on these high-quality links, C-NOMA ensures reliable communication and efficient resource utilization. In contrast, S-NOMA relies on a random selection process for Phase 1 transmissions, which can result in suboptimal performance when users with weaker channel conditions are chosen.

Fig. 3.11 explores the energy allocation dynamics across two phases of transmission, addressing the research question: **RQ2**. The results reveal that as the distortion threshold ($K$) increases, the energy required for uploading common information in phase 1 decreases, while the energy needed for transmitting individual information in phase 2 rises. C-NOMA demonstrates superior energy efficiency, consuming $21.45\% - 23.94\%$ less energy in phase 1 and $27.87\% - 52.43\%$ less energy in phase 2 compared to S-NOMA. Despite this reduced energy consumption, C-NOMA achieves better distortion performance, underscoring its effectiveness in balancing energy efficiency with communication quality.

The energy consumption behavior across the two phases highlights an inverse relationship: reducing energy usage in phase 1 increases it in phase 2, and vice versa. This interplay underscores the need for carefully balancing resource allocation to maintain overall energy efficiency while satisfying QoS requirements. By dynamically adjusting power and time allocation, systems can effectively manage these trade-offs to meet the specific demands of collaborative AR scenarios.

Leveraging collaborative strategies like those employed in C-NOMA offers sig-

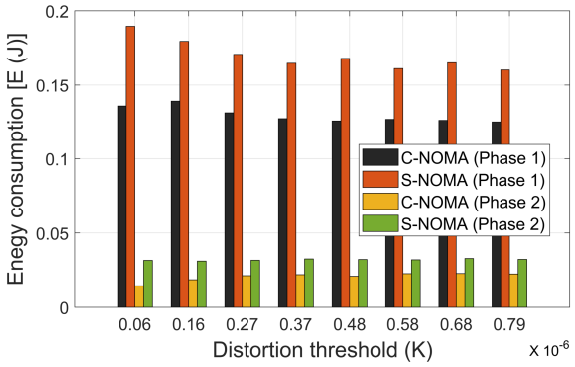



Figure 3.11: The energy allocation for proposed method (N = 5).

nificant advantages for efficiency, reliability, and scalability. By prioritizing users with stronger channel conditions and optimizing resource distribution, C-NOMA ensures high-quality communication with minimal energy expenditure. This makes it particularly well-suited for applications that require robust and low-latency communication, such as collaborative AR. These advancements not only enhance communication quality but also address the growing energy efficiency demands of modern wireless networks.

## 3.3 Short Summery of Chapter

This chapter explored advanced multiple access strategies for cooperative relaying and AR applications, emphasizing their role in optimizing spectrum use, enhancing connectivity, and balancing energy efficiency with communication quality. By integrating cooperative relaying with NOMA, various relaying protocols (AF, DF, PF) were analyzed, with PF-NOMA emerging as the most adaptable and efficient for shared-spectrum networks by optimizing power and time allocation while ensuring fairness.

For AR applications, collaborative NOMA-based frameworks were shown to enhance QoS through dynamic resource allocation, rate-distortion optimization, and energy-efficient transmission, ensuring high-throughput communication. By balancing power and time allocation, these frameworks offer a promising solution for real-time AR environments.

Despite advancements, challenges remain in interference mitigation, adaptive resource allocation, and energy-efficient frameworks under fluctuating network conditions. Addressing these issues is crucial for the large-scale deployment of cooperative relaying and AR-based communication in next-generation networks.

The next chapter explores QoS optimization in RAN slicing using advanced multiple access techniques. As network slicing becomes central to 5G and beyond, balancing high throughput for eMBB and low latency for URLLC is critical. The discussion will focus on hybrid access strategies like NOMA, RSMA, and puncturing,

leveraging threshold-based selection and dynamic resource allocation to optimize access scheme selection and QoS trade-offs, ensuring efficient and scalable network communication.

# Chapter 4

# Balancing QoS in RAN Slicing with Advanced Access Techniques

RAN slicing is a pivotal enabler in 5G-and-beyond networks, allowing efficient partitioning of network resources into virtual slices tailored to heterogeneous service requirements. Each slice caters to distinct applications, such as eMBB, URLLC, and mMTC. However, managing conflicting QoS requirements—such as high throughput for eMBB and low latency for URLLC—within shared resources is a significant challenge. Advanced MA schemes, such as NOMA, RSMA, and puncturing, present promising solutions to address these challenges.

This chapter explores how advanced MA schemes optimize resource allocation and QoS management in RAN slicing, focusing on the coexistence of eMBB and MC users. By leveraging hybrid access strategies and threshold-based selection mechanisms, these techniques ensure scalable, reliable communication while addressing trade-offs between latency, throughput, and resource utilization. Research questions addressed in this chapter include RQ3, RQ3.A, RQ3.B, RQ.2 and RQ2.A.

## 4.1 Hybrid Access Strategies for QoS Optimization in RAN Slicing

Efficient resource allocation in RAN slicing requires intelligent access strategies to balance the trade-offs between latency-sensitive and high-throughput applications. MC applications, such as industrial automation and autonomous driving, demand ultra-low latency, while eMBB services require high data rates. Traditional orthogonal access schemes struggle to accommodate these diverse requirements efficiently.

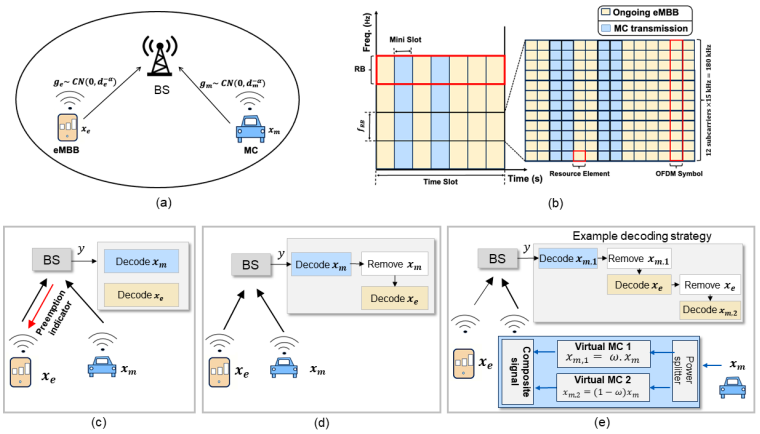



Figure 4.1: System model of eMBB-MC coexistence: (a) Single-cell uplink transmission scenario of eMBB and MC users, (b) Time-frequency resource grid with prescheduled eMBB user and generic overlapping access scheme for a "generate-at-will" MC user, (c) Puncturing-based access, (d) NOMA-based access, and (e) RSMA-based access scheme.

To address this challenge, hybrid multiple access techniques, including NOMA, RSMA, and puncturing, offer promising solutions. These strategies enable flexible resource sharing, ensuring that both low-latency and high-throughput requirements are met without significantly compromising network performance.

The objective of this thesis is to develop an access strategy that ensures the freshness of the information delivered by the MC user remains as close as possible to the best achievable value, while simultaneously maximizing the data rate for the eMBB user. To achieve this, threshold-based selection mechanisms—such as distance-based and SNR gap-based thresholds—are employed to guide the choice of the most appropriate access scheme. The system under consideration, as illustrated in Fig.4.1(a), consists of a single-cell uplink scenario with one eMBB user and one MC user. These users share the same RBs on the time-frequency grid, shown in Fig.4.1(b), resulting in conflicting requirements for latency and throughput. To address this coexistence challenge, decoding strategies such as puncturing, NOMA, and RSMA—depicted in Figs. 4.1(c), (d), and (e), respectively—are considered. By leveraging hybrid access strategies and threshold-based selection mechanisms, these techniques ensure scalable and reliable communication, while effectively managing the trade-offs between latency, throughput, and resource utilization. This section explores how these mechanisms contribute to optimizing QoS in heterogeneous networks.

### 4.1.1 Threshold-Based Strategy Selection

Selecting the appropriate multiple access scheme is crucial for optimizing the coexistence of eMBB and MC users in RAN slicing. Given the conflicting QoS requirements, a dynamic approach is necessary to determine the best strategy based on real-time network conditions. As illustrated in Fig. 4.1(b), the shared time-frequency

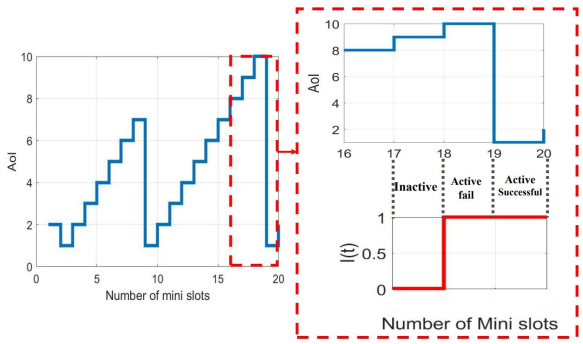



Figure 4.2: Evaluation of instantaneous AoI for MC versus minislots index.

grid requires intelligent strategies to ensure effective coexistence. Threshold-based selection mechanisms provide an efficient way to assess the suitability of puncturing, NOMA, or RSMA by evaluating user distance and SNR variations. AoI emerges as a key QoS metric in this context—capturing the freshness of information by tracking the time since the last successful update, as illustrated in Fig. 4.2—and these thresholds ensure that AoI constraints for MC users are met while maintaining high data rates for eMBB users. To achieve this balance, the system employs distance-based thresholds, which exploit user proximity to the base station to prioritize low-latency transmissions, and SNR gap-based thresholds, which leverage channel quality differences to dynamically switch between access schemes. These threshold-based mechanisms serve as the foundation for intelligent access strategy selection under varying network conditions. In the following sections, each of these strategies is described in detail, highlighting their design principles, operational logic, and impact on system performance.

### Distance-Based Thresholds

Distance-based thresholds evaluate the relative proximity of users to the BS and their impact on QoS metrics such as the average AoI. These thresholds guide the selection of NOMA or puncturing as follows

- **Puncturing:** The BS temporarily halts eMBB transmissions to prioritize MC packets, minimizing AoI for MC services. Puncturing is optimal when the MC user is farther from the BS ($\geq$ threshold distance $d_m^{\text{th}}$) or the eMBB user is closer to the BS ($\leq$ threshold distance $d_e^{\text{th}}$). These conditions ensure minimal transmission delays, as illustrated in Fig. 4.3.
- **NOMA:** NOMA allows simultaneous transmission for eMBB and MC users via power-domain multiplexing. It is preferred when the eMBB user is farther from the BS ($> d_e^{\text{th}}$) or the MC user is closer to the BS ($< d_m^{\text{th}}$). NOMA balances eMBB throughput and MC latency while reducing interference for MC users in close proximity to the BS, as shown in Fig. 4.3.

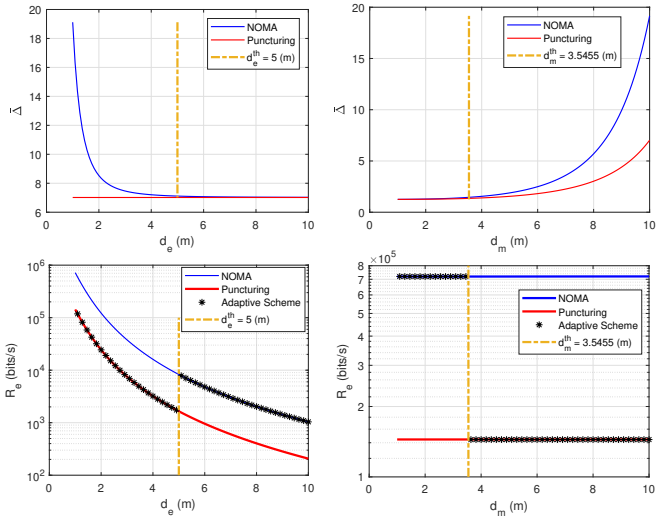



Figure 4.3: Achievable rate ($R_e$) for eMBB user under NOMA and puncturing as a function of $d_m$ for distance based.

### SNR Gap-Based Thresholds

SNR gap thresholds analyze the channel gain difference between eMBB and MC users. The SNR gap, defined as

$$\Gamma_d = 10 \log_{10} \left( \frac{P_e d_e^{-\alpha} |g_e|^2}{P_m d_m^{-\alpha} |g_m|^2} \right),$$

represents the SNR gap, accounting for both user distances and channel conditions. The objective of identifying the threshold is to select the optimal access scheme such that the freshness of the information delivered by the MC user is kept as close as possible to the best achievable value while simultaneously maximizing the eMBB user's data rate. These thresholds guide the selection of puncturing, NOMA, or RSMA

- **Puncturing:** Effective for high SNR gaps ($\Gamma_d \geq 31.90$ dB), where MC reliability is prioritized. This ensures low AoI and high reliability, as shown in Fig. 4.4.
- **NOMA:** Suitable for moderate SNR gaps ($-44.3$ dB $\leq \Gamma_d \leq 17.12$ dB). NOMA achieves a favorable trade-off between eMBB throughput and MC latency, leveraging power-domain multiplexing to meet diverse demands.
- **RSMA:** Optimal for intermediate SNR gaps (17.12 dB $\leq \Gamma_d \leq 31.90$ dB). RSMA enhances flexibility by splitting user messages into streams with distinct power and rate-splitting factors, achieving ideal AoI-throughput trade-offs, as illustrated in Fig. 4.4.

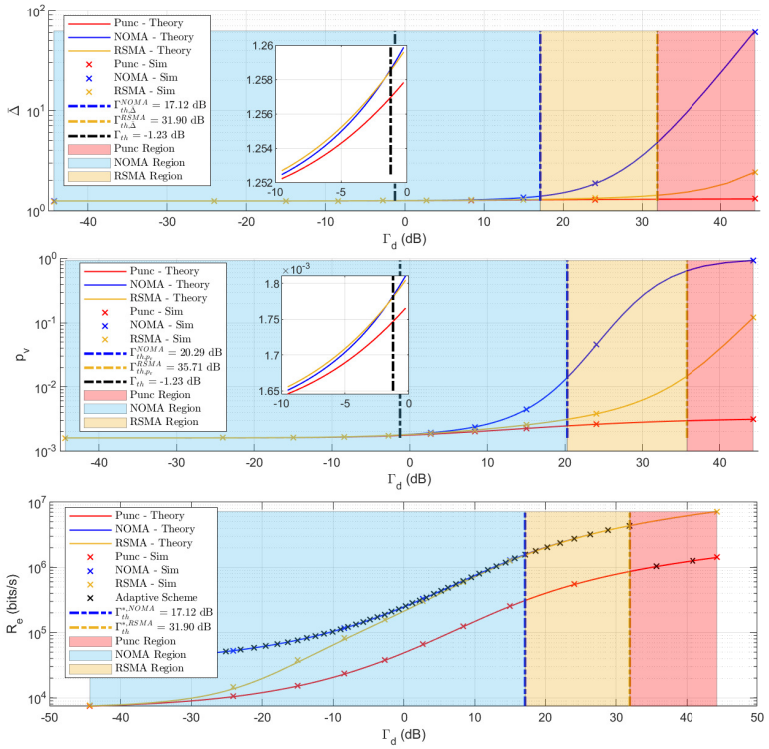



Figure 4.4: Probability of successful transmission ($s_m$), average AoI ($\bar{\Delta}$), and PAoI violation probability ($p_v$) for the MC user, as functions of the SNR gap ($\Gamma_d$).

### 4.1.2 Threshold-Guided Access for Mixed QoS Demands

This study builds upon prior analyses of NOMA and puncturing by introducing RSMA as a candidate solution for eMBB-MC coexistence. A comprehensive evaluation of AoI and data rate trade-offs across puncturing, NOMA, and RSMA was conducted, identifying the conditions where RSMA or NOMA outperforms puncturing based on SNR gaps, distance thresholds, and user requirements. Advanced MA schemes leverage hybrid access strategies to address conflicting QoS requirements in RAN slicing. Techniques like NOMA, RSMA, and puncturing dynamically adapt to user demands, optimizing resource allocation and improving scalability.

By dynamically selecting hybrid access schemes based on distance and SNR thresholds, this approach ensures scalable, reliable communication for MC applications while maintaining high eMBB data rates. NOMA's power-domain separation prioritizes low-latency URLLC traffic while ensuring high throughput for eMBB users, whereas RSMA enhances flexibility with its message-splitting mechanism, achieving optimal trade-offs between throughput and latency. Puncturing minimizes MC freshness of information but reduces eMBB throughput, whereas RSMA provides

superior flexibility, adapting to diverse demands at the cost of increased complexity. These findings highlight the importance of carefully selecting MA techniques to balance latency, throughput, and resource utilization, enabling scalable and reliable communication for 5G-and-beyond applications.

This chapter explicitly addresses the research questions as follows: For RQ3, advanced MA schemes like NOMA, RSMA, and puncturing are shown to dynamically optimize resource allocation and manage QoS in heterogeneous services by leveraging threshold-based mechanisms (distance- and SNR-based thresholds). For RQ3.A, NOMA improves resource allocation through power-domain multiplexing, while RSMA employs message splitting for better flexibility and trade-offs, effectively supporting the diverse requirements of eMBB and URLLC. Finally, for RQ3.B, the trade-offs in selecting MA techniques are discussed, where puncturing ensures minimal latency at the expense of throughput, while RSMA provides superior adaptability with increased complexity, emphasizing the importance of strategic decisions for QoS management in RAN slicing.

## 4.2 Adaptive User Pairing with advanced MA Choices

Efficient resource allocation in uplink communication is crucial for managing the coexistence of eMBB and MC traffic. These two services have distinct QoS requirements—eMBB demands high data rates, while MC applications prioritize low latency and reliability. This section introduces an adaptive pairing framework that dynamically selects the most suitable MA scheme, NOMA, RSMA, or puncturing based on instantaneous SNR thresholds and QoS metrics such as AoI. By addressing the challenges posed by heterogeneous traffic in uplink scenarios, the proposed method ensures optimal performance for next-generation communication systems.

The system consists of multiple eMBB and MC users distributed around a BS. Each eMBB user operates with a full-buffer traffic model, while MC users generate sporadic packets based on a bernoulli distribution. Both user types share the same time-frequency resources, which are divided into RBs and further subdivided into minislots. The proposed pairing scheme dynamically assigns MC users to eMBB users based on real-time channel conditions and QoS demands. The framework is illustrated in Fig. 4.5, where the pairing strategy and resource allocation are depicted across spatial user distributions and the time-frequency grid.

### 4.2.1 Adaptive Pairing and Access Scheme Selection

The adaptive pairing algorithm is designed to minimize interference and maximize throughput by pairing an eMBB user with the MC user exhibiting the largest channel gain difference. The selection of the appropriate access scheme is guided by SNR gap thresholds, which define the conditions under which NOMA, RSMA, or puncturing is most effective. NOMA is preferred due to its ability to enable simultaneous transmissions through power-domain multiplexing, ensuring high data rates for eMBB

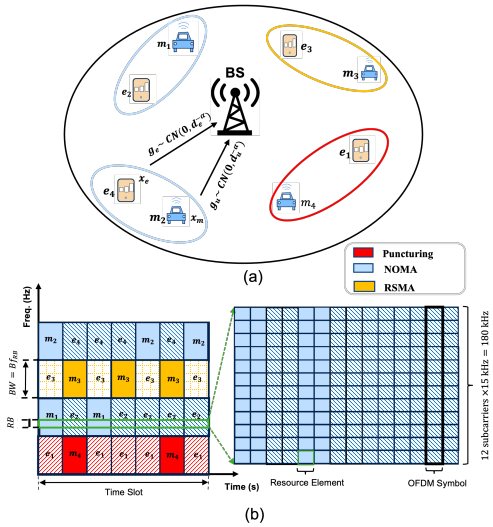



Figure 4.5: System model of eMBB-MC coexistence: (a) Single-cell uplink transmission scenario of eMBB and MC users, and (b) Time-frequency resource grid with prescheduled eMBB users $(e_1, e_2, e_3, e_4)$ and generic overlapping access scheme for "generate-at-will" MC users $(m_1, m_2, m_3, m_4)$.

users while maintaining acceptable latency for MC users. When NOMA is not feasible, RSMA is employed as an alternative, leveraging message splitting to manage interference efficiently. Puncturing serves as a fallback mechanism, prioritizing MC transmissions by temporarily suspending eMBB transmissions to minimize AoI.

The pairing strategy dynamically adjusts based on real-time variations in network conditions and user activity, ensuring robust and efficient performance across different traffic loads and deployment scenarios.

### 4.2.2 System-Level Analysis of Adaptive Pairing for Heterogeneous QoS in RAN Slicing

The performance of adaptive pairing is evaluated using a MATLAB-based simulator, with comparisons made against two baseline strategies: *Random Grouping with NOMA* and *Traditional NOMA Grouping*. The evaluation focuses on key metrics, including average AoI, PAoI violation probability, and eMBB data rates, under varying spatial user distributions and network conditions.

The proposed adaptive user pairing algorithm in this thesis uses an optimally derived SNR gap threshold to intelligently pair users in a way that satisfies both latency-sensitive QoS constraints for MC users and high data rate demands for eMBB users.

To benchmark the performance of the proposed algorithm, two baseline strategies are considered

- **Random Grouping with NOMA:** eMBB and MC users are randomly paired, and each pair uses NOMA for uplink transmission to the base station. This represents a worst-case scenario, serving as a baseline to highlight the impact of uncoordinated user pairing on MC users' QoS, particularly in terms of outage probability.
- **Traditional NOMA Grouping:** eMBB users are paired with MC users who have the highest SNR gap. If a particular MC user is already paired, the next best option is selected. This method prioritizes eMBB throughput by minimizing interference from MC users and reflects a conventional NOMA strategy.

Since the effectiveness of user pairing is highly dependent on the users' spatial positions relative to the BS, the algorithm's performance is evaluated under three realistic deployment scenarios, shown in the top row of Fig. 4.6:

- **Concave Distribution:** eMBB users are concentrated close to the BS ($R_{\text{eMBB}} = 1$ m), while MC users are located farther away ($R_{\text{MC}} = 3$ m), creating a high-interference scenario.
- **Uniform Distribution:** Both eMBB and MC users are evenly distributed within the same radius ($R_{\text{eMBB}} = R_{\text{MC}} = 3$ m), representing a balanced layout.
- **Convex Distribution:** MC users are located close to the BS ($R_{\text{MC}} = 1$ m), while eMBB users are farther away ($R_{\text{eMBB}} = 3$ m), favoring MC transmission reliability.

Outage probability is a key performance indicator used to measure the proportion of MC users that fail to meet their QoS thresholds, such as average AoI or PAoI. It is defined as the ratio of MC users who do not satisfy their QoS requirement to the total number of MC users:

$$p(\text{QoS}_{\text{MC}} < \text{QoS}_{\text{th}}) = \frac{\sum_{j=1}^{M} \mathbb{I}_A(j)}{M}, \tag{4.1}$$

where $M$ is the total number of MC users, and $\mathbb{I}_A(j)$ is an indicator function identifying users who do not meet the QoS threshold. This metric provides insight into the system's reliability and the effectiveness of the pairing algorithm in minimizing QoS violations.

As shown in Fig. 4.6, the proposed adaptive pairing algorithm significantly reduces the QoS outage probability for MC users compared to both baseline approaches. In interference-prone environments, such as the concave distribution, the algorithm achieves an **85% reduction** in outage probability at low activation probabilities ($p_m = 0.1$). In the uniform distribution scenario, the algorithm consistently ensures **zero QoS outage probability**, demonstrating its robustness across different network configurations.

Moreover, the algorithm shows strong **resilience to moderate CSI (Channel State Information) estimation errors**. Even with a 5% error margin, it continues to outperform traditional grouping strategies. However, as CSI errors become more severe,

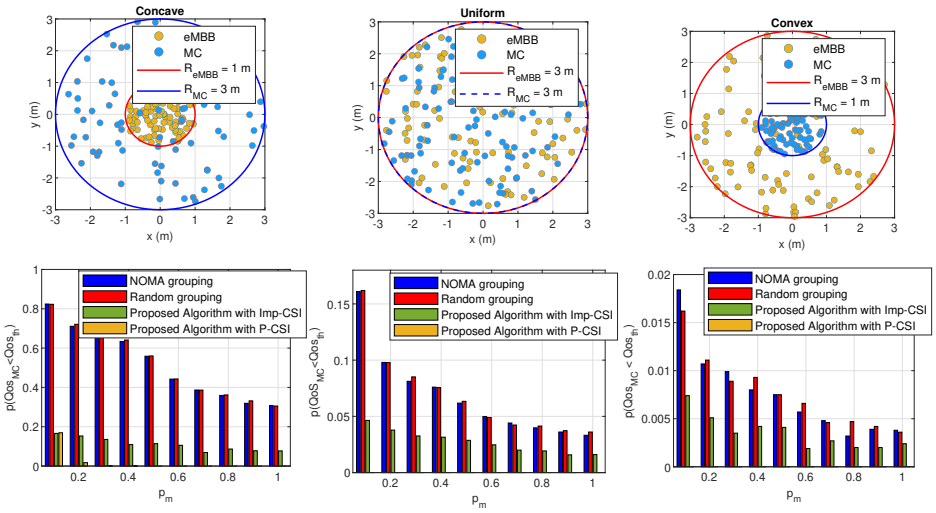



Figure 4.6: Performance results of the proposed pairing strategy: *top row*–user distribution scenarios, *middle row*–MC user outage probability as a function of $p_m$, *last row*–eMBB user data rate as a function of SNR gap ($\Gamma_d$).

some degradation in MC users' QoS is observed—particularly in high-interference scenarios—highlighting the importance of accurate channel information for maintaining performance.

This study directly addresses the research questions by demonstrating how advanced MA techniques—namely NOMA, RSMA, and puncturing—can be dynamically integrated to optimize resource allocation, balance QoS requirements, and support the coexistence of heterogeneous services in RAN slicing. RQ3 is tackled through an adaptive framework that selects access schemes based on real-time conditions, while RQ3.A is explored by analyzing how NOMA and RSMA meet diverse latency and throughput demands under varying SNR thresholds and user distributions. RQ3.B is addressed through performance evaluations that reveal trade-offs: NOMA favors high throughput, RSMA offers flexibility with added complexity, and puncturing minimizes latency but limits eMBB performance. Furthermore, RQ2 and RQ2.A are implicitly covered through the system's use of NOMA to support AoI-sensitive applications efficiently. Overall, the findings highlight the proposed framework's effectiveness in enabling scalable, adaptable, and QoS-aware communication for future wireless networks.

## 4.3 Short Summary of Chapter

This chapter examined advanced multiple access strategies for balancing QoS in RAN slicing, focusing on the coexistence of eMBB and MC applications. It highlighted the limitations of traditional orthogonal access and the need for hybrid solu-

tions like NOMA, RSMA, and puncturing to meet contrasting requirements.

A key contribution was the introduction of threshold-based selection mechanisms, using distance-based and SNR gap-based thresholds to optimize MC and eMBB coexistence. RSMA proved highly flexible for interference management, NOMA efficiently shared spectrum, and puncturing ensured ultra-low latency at the cost of eMBB throughput. The evaluation of AoI underscored the impact of access scheme choices on system performance.

The chapter also introduced an adaptive user pairing framework that dynamically selects access schemes based on SNR variations, significantly improving reliability, reducing outages, and optimizing network performance. However, challenges remain in real-time decision-making, mitigating computational complexity, and ensuring robust performance in dynamic multi-cell networks—critical hurdles for deploying RAN slicing and non-orthogonal access in next-generation networks.

The next chapter summarizes key research contributions, highlighting advancements in NOMA, RSMA, and puncturing across various wireless scenarios. It reviews novel relaying schemes, collaborative AR transmission, access scheme selection for eMBB-URLLC coexistence, and adaptive user pairing for mission-critical applications, providing insights into their impact on 5G and beyond.

# Chapter 5

# Summary of Publications

## 5.1 Paper 1 : NOMA Relaying for Optimal Access and Reliability

**Khodakhah F**, Mahmood A, Österberg P, Gidlund M. Multiple access-enabled relaying with piece-wise and forward NOMA: Rate optimization under reliability constraints. Sensors. 2021 Jul 13;21(14):4783.

### 5.1.1 Motivation and Contributions

Future IoT and 5G networks must support the coexistence of primary and secondary services while ensuring reliability and high data rates. This paper is motivated by the need to optimize relaying schemes in NOMA systems, particularly under strict reliability constraints, to support both primary (e.g., eMBB) and secondary (e.g., IoT) services. The paper proposes a novel PF-NOMA scheme that adapts its relaying strategy based on the quality of the received signal. The primary contribution is the development of an optimization framework for time and power allocation that maximizes the secondary user rate while meeting the reliability constraints of the primary service. Simulation results show that PF-NOMA outperforms existing DF and AF NOMA schemes in terms of rate regions and reliability.

### 5.1.2 Novelty

The introduction of the PF-NOMA scheme represents a significant novel contribution by combining aspects of both DF and AF relaying based on the quality of the signal. This hybrid approach enhances the performance of NOMA relays in terms of both data rate and reliability. Furthermore, the development of an optimization framework aimed at maximizing secondary user rates while adhering to reliability

constraints is another innovative aspect. This framework is particularly relevant and applicable to IoT environments and heterogeneous networks.

### 5.1.3 Limitation

The paper operates under the assumption of perfect knowledge of CSI for optimization purposes, which may not be entirely realistic in dynamic or fast-fading environments. Additionally, the proposed scheme primarily focuses on rate maximization and does not account for AoI or other latency-sensitive metrics that could be crucial in mission-critical applications.

### 5.1.4 Author Contribution

Farnaz Khodakhah proposed the PF-NOMA scheme, developed the mathematical models, and led the optimization analysis and in writing the paper. Aamir Mahmood contributed to the design and formulation of the optimization problem and in writing the paper. Patrik Österberg and Mikael Gidlund provided valuable feedback on the methodology, assisted in refining the mathematical models, and contributed to validating the results.

## 5.2 Paper 2: Efficient NOMA Transmission for Industrial AR Systems

**Khodakhah F**, Mahmood A, Abedin SF, Thar K, Österberg P, Gidlund M. Design and Resource Allocation of NOMA-based Transmission Scheme for Industrial Collaborative AR. In2022 IEEE Globecom Workshops (GC Wkshps) 2022 Dec 4 (pp. 1604-1609). IEEE.

### 5.2.1 Motivation and Contributions

The need for real-time, collaborative AR in industrial environments requires robust transmission schemes capable of supporting high data rates for multiple users simultaneously. This paper is motivated by the challenge of meeting the stringent data rate and distortion requirements of collaborative AR applications in resource-constrained wireless systems. The paper introduces a collaborative NOMA-based transmission scheme specifically tailored for multi-user AR environments, where common and individual views are transmitted with different priorities. The main contribution is the formulation of an optimization problem for joint time and power allocation to minimize the maximum distortion of individual views while ensuring a target distortion for common views. The solution leverages primal-dual interior-point algorithms and successive convex approximation (SCA), demonstrating significant gains in energy efficiency and video quality.

### 5.2.2 Novelty

The novelty of this work lies in the development of a collaborative NOMA scheme tailored specifically for industrial AR applications, leveraging the separation between common and individual views for multiple users. Additionally, the paper introduces a unique optimization framework that effectively balances energy consumption and rate-distortion performance, ensuring high-quality AR experiences even in resource-constrained environments.

### 5.2.3 Limitation

The scope of this work is confined to collaborative AR environments, leaving the generalizability of the proposed scheme to other industrial IoT applications largely unexplored. Furthermore, the optimization problem is formulated under the assumption of static channel conditions, which may not accurately represent the rapidly changing conditions encountered in real-world AR applications.

### 5.2.4 Author Contribution

Farnaz Khodakhah proposed the collaborative NOMA scheme, formulated the optimization problem, and conducted the performance analysis and writing the paper. Aamir Mahmood and Sarder Fakhrul Abedin contributed to the development of the optimization techniques and provided insights into the application of SCA. Kyi Thar, Patrik Österberg, and Mikael Gidlund provided feedback on the applicability of the scheme to industrial AR use cases.

## 5.3 Paper 3: NOMA or Puncturing for Uplink eMBB-URLLC from an AoI View

**Khodakhah F**, Stefanović Č, Mahmood A, Farag H, österberg P, Gidlund M. NOMA or Puncturing for Uplink eMBB-URLLC Coexistence from an AoI Perspective?. In GLOBECOM 2023-2023 IEEE Global Communications Conference 2023 Dec 4 (pp. 4301-4306). IEEE.

### 5.3.1 Motivation and Contributions

As URLLC and eMBB traffic must coexist in 5G networks, ensuring low AoI for URLLC traffic without severely compromising eMBB throughput is essential. This paper explores whether NOMA, which allows for simultaneous transmission, can provide comparable AoI to puncturing, a scheme that halts eMBB transmission for URLLC traffic, while maintaining higher eMBB data rates.
The paper presents a detailed analysis of the conditions under which NOMA and

puncturing should be employed. It develops a closed-form expression for average AoI under both schemes and identifies a threshold for switching between NOMA and puncturing based on user proximity to the base station. It also quantifies the performance gap between NOMA and puncturing in terms of AoI and demonstrates how NOMA can offer up to five times higher data rates for eMBB users under certain conditions.

### 5.3.2 Novelty

The novelty of this work lies in the formalization of operational thresholds for switching between NOMA and puncturing schemes based on user distance and channel conditions, offering a practical framework for real-time decision-making in 5G networks. Additionally, the application of DTMC for analyzing the AoI in the context of NOMA and puncturing represents another significant and innovative contribution.

### 5.3.3 Limitation

The model primarily focuses on average AoI and does not consider more advanced QoS metrics that could capture worst-case delays experienced by URLLC users. Moreover, the results are derived from simulations using a simplified traffic model, which may not adequately reflect the variability and complexity of real-world URLLC and eMBB traffic scenarios.

### 5.3.4 Author Contribution

Farnaz Khodakhah led the research, developed the DTMC model, derived the analytical expressions for AoI, and played the primary role in writing and coordinating the paper. Čedomir Stefanović and Hossam Farag contributed to the mathematical modeling and DTMC analysis, providing key insights into the switching thresholds between NOMA and puncturing. Aamir Mahmood contributed to defining the direction and scope of the problem, in addition to providing critical feedback on the overall research framework. Patrik Österberg and Mikael Gidlund provided essential guidance on the research direction and validated the analytical models through simulation.

## 5.4 Paper 4: AoI-Rate Tradeoff for Mission-Critical and eMBB Uplink

**Khodakhah F**, Mahmood A, Stefanovic C, Farag H, Osterberg P, Gidlund M. Balancing AoI and Rate for Mission-Critical and eMBB Coexistence with Puncturing, NOMA, and RSMA in Cellular Uplink. IEEE Transactions on Vehicular Technology. 2024 Sep 2.

### 5.4.1 Motivation and Contribution

With the growing need for MC applications, ensuring low-latency, highly reliable communication for MC services alongside the bandwidth-heavy demands of eMBB services presents a critical challenge. This paper is motivated by the need to optimize multiple access schemes to ensure that both services can coexist effectively while maintaining low AoI for MC user and high data rates for eMBB users. The paper systematically analyzes and compares three multiple access schemes puncturing, NOMA, and RSMA in terms of their ability to balance AoI for MC applications and data rates for eMBB users. The key contributions include the derivation of closed form expressions for average AoI and peak AoI violation probabilities, identification of optimal power and rate allocation strategies for RSMA, and demonstrating how RSMA can achieve near puncturing level AoI performance with better eMBB rates.

### 5.4.2 Novelty

This paper is the first to comprehensively examine the performance of RSMA in the context of AoI for mission-critical services, offering detailed insights into its application in 5G uplink scenarios. It introduces a novel power and rate allocation strategy designed to optimize the trade-off between the timeliness requirements of MC and the throughput demands of eMBB users.

### 5.4.3 Limitation

Although the paper offers extensive theoretical analysis, it does not fully address the practical implementation challenges of RSMA, such as decoding complexity. Additionally, the analysis is limited to a single-cell uplink scenario, which may not adequately capture the complexities and impact of multi-cell interference in real-world deployments.

### 5.4.4 Author Contribution

Farnaz Khodakhah conceptualized the study, developed the mathematical models, conducted the performance analysis, and played a key role in writing the paper. Aamir Mahmood provided significant insights into NOMA and RSMA schemes and contributed to refining the optimization approach. Čedomir Stefanović assisted in the derivation of key metrics and provided feedback on the modeling approaches. Hossam Farag, Patrik Österberg, and Mikael Gidlund contributed to the conceptual framework and validated the results through simulation, offering important guidance throughout the research process.

## 5.5 Paper 5: Adaptive Pairing and Access for Cellular IoT Services

**Khodakhah, F.**, Mahmood, A., Österberg, P., & Gidlund, M. (2024). Adaptive User Pairing with Non-orthogonal Medium Access Choices for Balanced Coexistence of Mission-Critical and eMBB Services in Cellular IoT. *IEEE Open Journal of the Communications Society* (Under Review)

### 5.5.1 Motivation and Contributions

The increasing coexistence of heterogeneous traffic services in 5G networks, such as eMBB and MC services, presents a major challenge in ensuring the QoS requirements of both services. MC services, such as MC, require strict timeliness and reliability, while eMBB demands high data rates. In cellular IoT, this coexistence requires a balanced multiple access approach to optimize performance for both services. This paper is motivated by the need to develop adaptive user pairing strategies and access scheme selection mechanisms that ensure MC users meet stringent QoS metrics like AoI, while maximizing the throughput of eMBB users. The paper derives an SNR gap threshold that determines when to use puncturing, NOMA, or RSMA based on SNR gap between eMBB and MC users and activation probability. It proposes an adaptive pairing algorithm that balances MC users QoS requirements (minimizing AoI and Peak AoI violation probability) while maximizing eMBB data rates. It conducts extensive simulations under different user deployment scenarios (concave, uniform, convex) and demonstrates how the proposed pairing strategy outperforms traditional NOMA and random pairing, reducing MC users QoS outage by 85% in some cases. It provides insights into network wide performance, identifying conditions (e.g., activation probability and network size) where RSMA and NOMA become preferable access methods.

### 5.5.2 Novelty

The paper introduces a novel SNR gap threshold that facilitates switching between multiple access schemes puncturing, NOMA, and RSMA based on user proximity to the base station and activation probability. This threshold is crucial for achieving optimal QoS for MC users while maximizing eMBB rates. Additionally, the paper presents the first adaptive user pairing algorithm designed for uplink MC and eMBB coexistence under various non orthogonal access schemes. This algorithm dynamically determines the optimal pairing and access scheme by considering the SNR gap and user distribution, consistently demonstrating superior performance across diverse scenarios. Furthermore, the paper uniquely integrates AoI and Peak AoI violation probabilities into the decision making process, prioritizing MC services that demand data freshness and low latency an approach that contrasts with existing literature focused primarily on metrics like sum rate or fairness.

### 5.5.3 Limitation

The analysis in this paper is confined to single cell network scenarios, without accounting for multi cell environments where inter cell interference could significantly impact the practical performance of NOMA and RSMA schemes. Additionally, the computational complexity of the proposed pairing algorithm, which scales as $O(M^3)$ for larger user sets, poses challenges for real-time implementation in large-scale IoT networks. The study also assumes perfect CSI, an assumption that may not hold in fast-fading or dynamically changing wireless environments, potentially reducing the accuracy of the SNR gap threshold in practical applications. Furthermore, while the paper's focus is on uplink transmission, its findings may not be directly applicable to downlink scenarios, where the dynamics of user pairing and interference management can differ significantly.

### 5.5.4 Author Contributions

Farnaz Khodakhah led the research, developed the mathematical models for the SNR gap thresholds, designed the adaptive pairing algorithm, conducted simulations, and played a primary role in writing the paper. Aamir Mahmood contributed to the conceptual framework, particularly in the development of the pairing strategy and the access scheme selection process. Provided feedback on the optimization approach and reviewed the manuscript. Patrik Österberg provided critical insights into the evaluation of QoS metrics such as AoI and Peak AoI violation probabilities. Assisted in refining the simulation methodology and validation of results. Mikael Gidlund provided overall guidance, particularly in the practical applicability of the proposed algorithm to 5G IoT scenarios. Reviewed and edited the manuscript, ensuring alignment with the latest advances in non-orthogonal access technologies.

# Chapter 6

# Conclusions

This thesis has thoroughly investigated advanced medium access techniques and resource allocation strategies to address critical challenges in modern wireless communication systems, specifically within the context of 5G and beyond. Aligned with the core themes of multi-service coexistence and collaboration strategies, the research presented here explores cooperative relaying, AR-driven applications, and RAN slicing as key enablers of flexible and efficient medium access. By systematically addressing the heterogeneous requirements of diverse service categories—such as eMBB, URLLC, and mMTC—the thesis proposes innovative, context-aware methodologies that enhance coexistence, optimize spectrum utilization, and support seamless collaboration across next-generation network architectures.

Research question 1 (RQ1), along with its sub-questions (RQ1.A and RQ1.B), was answered in Paper 1. The research focused on developing cooperative relaying strategies for primary-secondary coexistence in shared-spectrum networks, utilizing advanced MA schemes such as PF-NOMA. The study proposed optimal strategies for integrating NOMA into cooperative relaying to maximize spectral efficiency while addressing power and time allocation challenges. These techniques ensured fairness and reliability in relaying systems by dynamically adapting to varying network conditions. The advantage of this approach lies in its ability to significantly improve spectral efficiency and reliability in primary-secondary coexistence scenarios. However, the increased complexity of real-time power and time allocation remains a limitation in large-scale deployments.

For research question 2 (RQ2), including RQ2.A and RQ2.B, the primary contributions were presented in [Paper 2], with additional indirect support from [Paper 5]. These works proposed resource-efficient frameworks for latency-critical applications, particularly AR. [Paper 2] provided a detailed exploration of rate-distortion models and adaptive power control mechanisms, demonstrating how NOMA-based strategies can dynamically balance data rates and energy efficiency. These results ensured high video quality and minimized distortion in resource-constrained environments, addressing the requirements of industrial and collaborative AR applications. Additionally, [Paper 5] indirectly supported RQ2 by tackling broader resource allo-

cation challenges in latency-critical scenarios. The advantage of these approaches lies in their ability to meet stringent QoS requirements while maintaining scalability. However, challenges remain in accurately modeling rate-distortion characteristics and adapting resource allocation to diverse AR use cases, potentially limiting the generalizability of these methods.

Research question 3 (RQ3), addressed in sub-questions RQ3.A and RQ3.B, was explored in Paper 3, Paper 4, and Paper 5. These works focused on hybrid access schemes, such as NOMA, RSMA, and puncturing. By investigating the trade-offs between timeliness (measured as AoI) and data rate, these papers highlighted the superior performance of RSMA in balancing the freshness of information and data rates. Closed-form expressions for average and peak AoI were derived, providing valuable insights into resource allocation and interference management in RAN slicing. The advantage of this approach is its ability to achieve efficient coexistence of mission-critical and eMBB services. However, the main disadvantage lies in the increased computational overhead of implementing RSMA, particularly in dense network scenarios with multiple slices.

Overall, the thesis demonstrated the potential of advanced MA techniques in addressing the challenges of scalability, adaptability, and efficiency in next-generation wireless systems. The proposed methodologies offer significant improvements in spectral efficiency, reliability, and timeliness for diverse applications. However, the thesis has certain limitations. While the proposed methods are validated through extensive simulations and mathematical modeling, real-world implementation and validation were not conducted. Additionally, the computational complexity of some methods, such as RSMA and PF-NOMA, may present challenges for practical deployment in ultra-dense or highly dynamic environments. Despite these limitations, the contributions of this thesis provide a robust foundation for advancements in wireless communication systems. By addressing the posed research questions with innovative and effective solutions, the thesis advances the state of the art and sets a benchmark for future explorations in the field.

## 6.1 Ethical and Societal Considerations

Throughout the research work in this thesis, the author has remained critical of positive outcomes and has endeavored to present findings with transparency, avoiding overstating their applicability beyond the evaluated context. The research emphasizes stringent safety and reliability measures to ensure that the developed methodologies, when applied in real-world systems, do not compromise human safety or environmental integrity.

A key ethical concern is related to the safety-critical nature of MC communication systems and eMBB coexistence. Failure to satisfy the timing and reliability requirements for MC applications could lead to hazardous consequences, such as communication delays in industrial or healthcare scenarios. This research acknowledges the stringent reliability standards required in such applications and incorporates mechanisms, such as adaptive user pairing and hybrid access schemes, to minimize po-

tential risks in real-world deployments.

Security aspects have not been explicitly addressed in the scope of this thesis, as it assumes no intentional intrusion or malicious activity during the communication process. However, it is recognized that implementing a complete system for real-world deployment would necessitate integrating the proposed solutions with robust security mechanisms to mitigate risks such as unauthorized access, data breaches, or service disruptions.

From a societal perspective, the advancements in hybrid access strategies and NOMA techniques discussed in this work align with the broader digital transformation trends enabled by 5G and beyond networks. These technologies have the potential to drive innovation across industries, creating new applications and opportunities in areas like industrial automation, AR, and healthcare monitoring. They can also enhance productivity by reducing latency and improving data rate efficiencies, enabling transformative applications that positively impact economic and social development.

However, the ubiquity of such advancements may also lead to societal challenges, such as job displacement due to increased automation in industrial and other sectors. While the rise of automation and IoT technologies augments human capabilities, it necessitates proactive strategies to manage their implications, including workforce reskilling and policy frameworks to ensure equitable societal benefits. The integration of these technologies must, therefore, balance economic growth with ethical responsibility towards societal welfare.

## 6.2 Future Directions

While this thesis has made significant contributions to the optimization of multiple access techniques and resource allocation in 5G/6G networks, several directions remain open for future research. A critical next step is the practical implementation and real-world validation of hybrid access schemes, particularly RSMA and NOMA, to evaluate their feasibility in large-scale and computationally constrained environments. The study was primarily limited to single-cell scenarios, and extending the analysis to multi-cell environments while accounting for inter-cell interference, dynamic user mobility, and collaborative resource management will provide a more comprehensive understanding of the proposed schemes in real-world deployments. Additionally, the integration of multiple access techniques with emerging technologies such as reconfigurable intelligent surfaces (RIS), unmanned aerial vehicles (UAVs), and satellite-terrestrial hybrid networks presents an exciting avenue for enhancing coverage, energy efficiency, and adaptability in future wireless systems.

Another crucial research direction lies in the development of AI-driven optimization techniques for resource allocation, leveraging machine learning and reinforcement learning to enable intelligent, dynamic, and scalable decision-making for access scheme selection, power control, and QoS management. As wireless networks become increasingly complex, such data-driven approaches can play a key role in

optimizing resource distribution in highly dynamic environments. An additional extension of this work could involve proposing a more efficient and realistic user grouping strategy — specifically, by grouping as many MC users as possible on the eMBB resource blocks, without compromising the QoS requirements of MC users and while ensuring the minimum rate requirements of eMBB users. One potential solution to this problem may involve the use of machine learning approaches such as reinforcement learning, although further exploration is needed to confirm its effectiveness.

Moreover, expanding the scope of QoS metrics beyond AoI and PAoI to include worst-case delays, energy efficiency, and fairness considerations would provide a more holistic framework for resource allocation strategies. Future research should also explore optimized downlink and bidirectional communication strategies, which are essential for applications such as remote healthcare, autonomous driving, and smart city infrastructure. By addressing these future research areas, the work initiated in this thesis can be expanded to contribute to the development of next-generation communication systems that are more intelligent, efficient, and resilient. As the evolution toward 6G and beyond continues, interdisciplinary research incorporating AI, security, and sustainability principles will be crucial in shaping the future of wireless networks to meet the growing demands of emerging applications and services.

# Bibliography


[1] AFM Shahen Shah, Ahmed Nidham Qasim, Muhammet Ali Karabulut, Haci Ilhan, and Md Baharul Islam. Survey and performance evaluation of multiple access schemes for next-generation wireless communication systems. *IEEE Access*, 9:113428–113442, 2021.

[2] Samsung Research. The next hyper-connected experience for all - 6g vision. Technical report, Samsung Research, 2020. White Paper.

[3] Lu Lv, Jian Chen, Qiang Ni, Zhiguo Ding, and Hai Jiang. Cognitive non-orthogonal multiple access with cooperative relaying: A new wireless frontier for 5g spectrum sharing. *IEEE Communications Magazine*, 56(4):188–195, 2018.

[4] Zhiguo Ding, Yuanwei Liu, Jinho Choi, Qi Sun, Maged Elkashlan, I Chih-Lin, and H Vincent Poor. Application of non-orthogonal multiple access in lte and 5g networks. *IEEE Communications Magazine*, 55(2):185–191, 2017.

[5] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[6] Yijie Mao, Onur Dizdar, Bruno Clerckx, Robert Schober, Petar Popovski, and H Vincent Poor. Rate-splitting multiple access: Fundamentals, survey, and future research trends. *IEEE Communications Surveys & Tutorials*, 24(4):2073–2126, 2022.

[7] Petar Popovski, Kasper Fløe Trillingsgaard, Osvaldo Simeone, and Giuseppe Durisi. 5g wireless network slicing for embb, urllc, and mmtc: A communication-theoretic view. *Ieee Access*, 6:55765–55779, 2018.

[8] Mostafa Zaman Chowdhury, Md Shahjalal, Shakil Ahmed, and Yeong Min Jang. 6g wireless communication systems: Applications, requirements, technologies, challenges, and research directions. *IEEE Open Journal of the Communications Society*, 1:957–975, 2020.

[9] Mohammed H Alsharif and Rosdiadee Nordin. Evolution towards fifth generation (5g) wireless networks: Current trends and challenges in the deployment of millimetre wave, massive mimo, and small cells. *Telecommunication Systems*, 64:617–637, 2017.

[10] Klaus David and Hendrik Berndt. 6g vision and requirements: Is there any need for beyond 5g? *IEEE vehicular technology magazine*, 13(3):72–80, 2018.

[11] Howard Benn. Vision and key features for 5th generation (5g) cellular. *Samsung R&D Institute UK*, 2014.

[12] Adrian Kliks, Lukasz Kulacz, Pawel Kryszkiewicz, Hanna Bogucka, Marcin Dryjanski, Magnus Isaksson, George P Koudouridis, and Per Tengkvist. Beyond 5g: Big data processing for better spectrum utilization. *IEEE Vehicular Technology Magazine*, 15(3):40–50, 2020.

[13] Souvik Maiti and Sonam Juneja. Energy efficiency techniques in 5g/6g networks: Green communication solutions. In *International Conference on Advances in Data-driven Computing and Intelligent Systems*, pages 407–423. Springer, 2023.

[14] Lauri Lovén, Miguel Bordallo López, Roberto Morabito, Jaakko Sauvola, Sasu Tarkoma, et al. Large language models in the 6g-enabled computing continuum: a white paper. 2025.

[15] Wei Jiang, Bin Han, Mohammad Asif Habibi, and Hans Dieter Schotten. The road towards 6g: A comprehensive survey. *IEEE Open Journal of the Communications Society*, 2:334–366, 2021.

[16] Yuanwei Liu, Zhijin Qin, Maged Elkashlan, Zhiguo Ding, Arumugam Nallanathan, and Lajos Hanzo. Nonorthogonal multiple access for 5g and beyond. *Proceedings of the IEEE*, 105(12):2347–2381, 2017.

[17] Maice Costa, Marian Codreanu, and Anthony Ephremides. On the age of information in status update systems with packet management. *IEEE Transactions on Information Theory*, 62(4):1897–1910, 2016.

[18] Qing He, Di Yuan, and Anthony Ephremides. On optimal link scheduling with min-max peak age of information in wireless systems. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–7. IEEE, 2016.

[19] Jeffrey G Andrews, Stefano Buzzi, Wan Choi, Stephen V Hanly, Angel Lozano, Anthony CK Soong, and Jianzhong Charlie Zhang. What will 5g be? *IEEE Journal on selected areas in communications*, 32(6):1065–1082, 2014.

[20] Akhil Gupta and Rakesh Kumar Jha. A survey of 5g network: Architecture and emerging technologies. *IEEE access*, 3:1206–1232, 2015.

[21] Zhiguo Ding, Xianfu Lei, George K Karagiannidis, Robert Schober, Jinhong Yuan, and Vijay K Bhargava. A survey on non-orthogonal multiple access for 5g networks: Research challenges and future trends. *IEEE Journal on Selected Areas in Communications*, 35(10):2181–2195, 2017.

[22] Anup Mishra, Yijie Mao, Onur Dizdar, and Bruno Clerckx. Rate-splitting multiple access for 6g—part i: Principles, applications and future works. *IEEE Communications Letters*, 26(10):2232–2236, 2022.

[23] Vipin Balyan. Cooperative relay to relay communication using noma for energy efficient wireless communication. *Telecommunication systems*, 77(2):271–281, 2021.

[24] Sun Mao, Jinsong Wu, Lei Liu, Dapeng Lan, and Amir Taherkordi. Energy-efficient cooperative communication and computation for wireless powered mobile-edge computing. *IEEE Systems Journal*, 16(1):287–298, 2020.

[25] Yushan Siriwardhana, Pawani Porambage, Madhusanka Liyanage, and Mika Ylianttila. A survey on mobile augmented reality with 5g mobile edge computing: Architectures, applications, and technical aspects. *IEEE Communications Surveys & Tutorials*, 23(2):1160–1192, 2021.

[26] Xiuquan Qiao, Pei Ren, Guoshun Nan, Ling Liu, Schahram Dustdar, and Junliang Chen. Mobile web augmented reality in 5g and beyond: Challenges, opportunities, and future directions. *China Communications*, 16(9):141–154, 2019.

[27] Ejder Bastug, Mehdi Bennis, Muriel Médard, and Mérouane Debbah. Toward interconnected virtual reality: Opportunities, challenges, and enablers. *IEEE Communications Magazine*, 55(6):110–117, 2017.

[28] SM Riazul Islam, Nurilla Avazov, Octavia A Dobre, and Kyung-Sup Kwak. Power-domain non-orthogonal multiple access (noma) in 5g systems: Potentials and challenges. *IEEE Communications Surveys & Tutorials*, 19(2):721–742, 2016.

[29] Xenofon Foukas, Georgios Patounas, Ahmed Elmokashfi, and Mahesh K Marina. Network slicing in 5g: Survey and challenges. *IEEE communications magazine*, 55(5):94–100, 2017.

[30] Haijun Zhang, Na Liu, Xiaoli Chu, Keping Long, Abdol-Hamid Aghvami, and Victor CM Leung. Network slicing based 5g and future mobile networks: Mobility, resource management, and challenges. *IEEE communications magazine*, 55(8):138–145, 2017.

[31] Wen Wu, Nan Chen, Conghao Zhou, Mushu Li, Xuemin Shen, Weihua Zhuang, and Xu Li. Dynamic ran slicing for service-oriented vehicular networks via constrained learning. *IEEE Journal on Selected Areas in Communications*, 39(7):2076–2089, 2020.

[32] Phuong Luu Vo, Minh NH Nguyen, Tuan Anh Le, and Nguyen H Tran. Slicing the edge: Resource allocation for ran network slicing. *IEEE Wireless Communications Letters*, 7(6):970–973, 2018.

[33] Yuanwei Liu, Chongjun Ouyang, Zhiguo Ding, and Robert Schober. The road to next-generation multiple access: A 50-year tutorial review. *arXiv preprint arXiv:2403.00189*, 2024.

[34] Zhengquan Zhang, Yue Xiao, Zheng Ma, Ming Xiao, Zhiguo Ding, Xianfu Lei, George K Karagiannidis, and Pingzhi Fan. 6g wireless networks: Vision, requirements, architecture, and key technologies. *IEEE vehicular technology magazine*, 14(3):28–41, 2019.

[35] Xiaohu You, Cheng-Xiang Wang, Jie Huang, Xiqi Gao, Zaichen Zhang, Mao Wang, Yongming Huang, Chuan Zhang, Yanxiang Jiang, Jiaheng Wang, et al. Towards 6g wireless communication networks: Vision, enabling technologies, and new paradigm shifts. *Science China Information Sciences*, 64:1–74, 2021.

[36] Walid Saad, Mehdi Bennis, and Mingzhe Chen. A vision of 6g wireless systems: Applications, trends, technologies, and open research problems. *IEEE network*, 34(3):134–142, 2019.

[37] Cheng-Xiang Wang, Xiaohu You, Xiqi Gao, Xiuming Zhu, Zixin Li, Chuan Zhang, Haiming Wang, Yongming Huang, Yunfei Chen, Harald Haas, et al. On the road to 6g: Visions, requirements, key technologies, and testbeds. *IEEE Communications Surveys & Tutorials*, 25(2):905–974, 2023.

[38] Yuanwei Liu, Wenqiang Yi, Zhiguo Ding, Xiao Liu, Octavia A Dobre, and Naofal Al-Dhahir. Developing noma to next generation multiple access: Future vision and research opportunities. *IEEE Wireless Communications*, 29(6):120–127, 2022.

[39] Yuanwei Liu, Shuowen Zhang, Xidong Mu, Zhiguo Ding, Robert Schober, Naofal Al-Dhahir, Ekram Hossain, and Xuemin Shen. Evolution of noma toward next generation multiple access (ngma) for 6g. *IEEE Journal on Selected Areas in Communications*, 40(4):1037–1071, 2022.

[40] Salah Eddine Elayoubi, Sana Ben Jemaa, Zwi Altman, and Ana Galindo-Serrano. 5g ran slicing for verticals: Enablers and challenges. *IEEE Communications Magazine*, 57(1):28–34, 2019.

[41] Lajos Hanzo, Mohammed El-Hajjar, and Osamah Alamri. Near-capacity wireless transceivers and cooperative communications in the mimo era: Evolution of standards, waveform design, and future perspectives. *Proceedings of the IEEE*, 99(8):1343–1385, 2011.

[42] Yunlong Cai, Zhijin Qin, Fangyu Cui, Geoffrey Ye Li, and Julie A McCann. Modulation and multiple access for 5g networks. *IEEE Communications Surveys & Tutorials*, 20(1):629–646, 2017.

[43] Raymond Steele and Lajos Hanzo. *Mobile radio communications: Second and third generation cellular and WATM systems: 2nd*. IEEE Press-John Wiley, 1999.

[44] Klein S Gilhousen, Irwin M Jacobs, Roberto Padovani, Andrew J Viterbi, Lindsay A Weaver, and Charles E Wheatley. On the capacity of a cellular cdma system. *IEEE transactions on vehicular technology*, 40(2):303–312, 1991.

[45] Junyi Li, Xinzhou Wu, and Rajiv Laroia. *OFDMA mobile broadband communications: A systems approach*. Cambridge University Press, 2013.

[46] Peng Wang, Jun Xiao, and Li Ping. Comparison of orthogonal and non-orthogonal approaches to future wireless cellular systems. *IEEE Vehicular Technology Magazine*, 1(3):4–11, 2006.

[47] Fredrik Brannstrom, Tor M Aulin, and Lars K Rasmussen. Iterative detectors for trellis-code multiple-access. *IEEE Transactions on Communications*, 50(9):1478–1485, 2002.

[48] Lihai Liu, Jun Tong, and Li Ping. Analysis and optimization of cdma systems with chip-level interleavers. *IEEE Journal on Selected Areas in Communications*, 24(1):141–150, 2005.

[49] Reza Hoshyar, Ferry P Wathan, and Rahim Tafazolli. Novel low-density signature for synchronous cdma systems over awgn channel. *IEEE Transactions on Signal Processing*, 56(4):1616–1626, 2008.

[50] Rudolf Ahlswede. On two-way communication channels and a problem by zarankiewicz. *Probl. of Control and Inform. Theory*, 1971.

[51] Thomas Cover. Broadcast channels. *IEEE Transactions on Information Theory*, 18(1):2–14, 1972.

[52] Yuya Saito, Yoshihisa Kishiyama, Anass Benjebbour, Takehiro Nakamura, Anxin Li, and Kenichi Higuchi. Non-orthogonal multiple access (noma) for cellular future radio access. In *2013 IEEE 77th vehicular technology conference (VTC Spring)*, pages 1–5. IEEE, 2013.

[53] Pulin Patel and Jack Holtzman. Analysis of a simple successive interference cancellation scheme in a ds/cdma system. *IEEE journal on selected areas in communications*, 12(5):796–807, 1994.

[54] Ferdi Kara and Hakan Kaya. Ber performances of downlink and uplink noma in the presence of sic errors over fading channels. *Iet Communications*, 12(15):1834–1844, 2018.

[55] Md Shipon Ali, Hina Tabassum, and Ekram Hossain. Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (noma) systems. *IEEE access*, 4:6325–6343, 2016.

[56] Ming Zeng, Wanming Hao, Octavia A Dobre, and Zhiguo Ding. Cooperative noma: State of the art, key techniques, and open challenges. *IEEE Network*, 34(5):205–211, 2020.

[57] Zhiguo Ding, Mugen Peng, and H Vincent Poor. Cooperative non-orthogonal multiple access in 5g systems. *IEEE Communications Letters*, 19(8):1462–1465, 2015.

[58] Jihye An, Onur Dizdar, Bruno Clerckx, and Wonjae Shin. Rate-splitting multiple access for multi-antenna broadcast channel with imperfect csit and csir. *arXiv preprint arXiv:2102.08738*, 2021.

[59] Yijie Mao. Rate-splitting multiple access for downlink communication systems. *HKU Theses Online (HKUTO)*, 2018.

[60] Sundar Aditya, Onur Dizdar, David Vargas, Bruno Clerckx, Stephen Wang, Xinze Lyu, and Sibo Zhang. Rate splitting multiple access: Prototypes, experiments and standardization efforts. In *2023 IEEE Conference on Standards for Communications and Networking (CSCN)*, pages 376–376, 2023.

[61] 3rd Generation Partnership Project (3GPP). 3GPP Release 16 Description. Technical Report 3GPP TR 21.916, 3GPP, 2020.

[62] 3rd Generation Partnership Project (3GPP). 3GPP Release 17 Description. Technical Report 3GPP TR 21.917, 3GPP, 2022.

[63] Michele Polese, Leonardo Bonati, Salvatore D'oro, Stefano Basagni, and Tommaso Melodia. Understanding o-ran: Architecture, interfaces, algorithms, security, and research challenges. *IEEE Communications Surveys & Tutorials*, 25(2):1376–1411, 2023.

[64] Milan Stolic, Julie Ann Connary, Wei Yan, Arghya Mukherjee, Mohammad Salaheldin, and Rob Piasecki. 5g automation architecture. *Cisco White Paper*, 2020.

[65] Roy D Yates, Yin Sun, D Richard Brown, Sanjit K Kaul, Eytan Modiano, and Sennur Ulukus. Age of information: An introduction and survey. *IEEE Journal on Selected Areas in Communications*, 39(5):1183–1210, 2021.

[66] Basel Barakat, Hachem Yassine, Simeon Keates, Ian Wassell, and Kamran Arshad. How to measure the average and peak age of information in real networks? In *European Wireless 2019; 25th European Wireless Conference*, pages 1–5. VDE, 2019.

[67] Qamar Abbas, Syed Ali Hassan, Haris Pervaiz, and Qiang Ni. A markovian model for the analysis of age of information in iot networks. *IEEE Wireless Communications Letters*, 10(7):1596–1600, 2021.

[68] Dinh-Thuan Do and Anh-Tu Le. Noma based cognitive relaying: Transceiver hardware impairments, relay selection policies and outage performance comparison. *Computer Communications*, 146:144–154, 2019.

[69] Huu-Phuc Dang, Minh-Sang V Nguyen, Dinh-Thuan Do, and Hong-Lien Pham. Exploiting hybrid decode-and-forward–amplify-and-forward in noma: An application to device-to-device networks. *International Journal of Communication Networks and Distributed Systems*, 25(2):145–163, 2020.

[70] Hongwu Liu, Zhiguo Ding, Kyeong Jin Kim, Kyung Sup Kwak, and H Vincent Poor. Decode-and-forward relaying for cooperative noma systems with direct links. *IEEE transactions on Wireless Communications*, 17(12):8077–8093, 2018.

[71] Krishna Srikanth Gomadam and Syed Ali Jafar. Optimal relay functionality for snr maximization in memoryless relay networks. *IEEE Journal on Selected Areas in Communications*, 25(2):390–401, 2007.

[72] Shuang Tian, Yonghui Li, and Branka Vucetic. Piecewise-and-forward relaying in wireless relay networks. *IEEE Signal Processing Letters*, 18(5):323–326, 2011.

[73] Nan Li, Ming Xiao, and Lars K Rasmussen. Optimized cooperative multiple access in industrial cognitive networks. *IEEE Transactions on Industrial Informatics*, 14(6):2666–2676, 2018.