UNICA

UNIVERSITÀ
DEGLI STUDI
DI CAGLIARI

Ph.D. DEGREE IN
ELECTRONIC AND COMPUTER ENGINEERING
Cycle XXXVI

TITLE OF THE Ph.D. THESIS

Multiuser Diversity Management for Multicast/Broadcast Services in 5G and Beyond Networks

Scientific Disciplinary Sector(s)

ING-INF/03

Ph.D. Student:	Ernesto Fontes Pupo
Supervisor	Prof. Maurizio Murroni

Final exam. Academic Year 2022/2023
Thesis defence: March 2024 Session

Dedicated to my mom and dad with their everyday phrases, "Tie your car to a star but let your feet on the earth" and "Things are done well or not done."

To my siblings, my perfect understanding of why, how, and what throughout my life. My paradigm.

To Claudia, the most beautiful person I ever met, my partner in life, my half, the other half, and more.

To my whole family and friends, who shape every bit of me.

Contents

List of Figures

List of Tables

Abstract

The envisaged fifth-generation (5G) and beyond networks represent a paradigm shift for global communications, offering unprecedented breakthroughs in media service delivery with novel capabilities and use cases. Addressing the critical research verticals and challenges that characterize the International Mobile Telecommunications (IMT)-2030 framework requires a compelling mix of enabling radio access technologies (RAT) and native softwarized, disaggregated, and intelligent radio access network (RAN) conceptions. In such a context, the multicast/broadcast service (MBS) capability is an appealing feature to address the ever-growing traffic demands, disruptive multimedia services, massive connectivity, and low-latency applications.

Embracing the MBS capability as a primary component of the envisaged 5G and beyond networks comes with multiple open challenges. In this research, we contextualize and address the necessity of ensuring stringent quality of service (QoS)/quality of experience (QoE) requirements, multicasting over millimeter-wave (mmWave) and sub-Terahertz (THz) frequencies, and handling complex mobility behaviors. In the broad problem space around these three significant challenges, we focus on the specific research problems of effectively handling the trade-off between multicasting gain and multiuser diversity, along with the trade-off between optimal network performance and computational complexity.

In this research, we cover essential aspects at the intersection of MBS, radio resource management (RRM), machine learning (ML), and the Open RAN (O-RAN) framework. We characterize and address the dynamic multicast multiuser diversity through low-complexity RRM solutions aided by ML, orthogonal multiple access (OMA) and non-orthogonal multiple access (NOMA) techniques in 5G MBS and beyond networks. We characterize the performance of the multicast access techniques conventional multicast scheme (CMS), subgrouping based on OMA (S-OMA), and subgrouping based on NOMA (S-NOMA). We provide conditions for their adequate selection regarding the specific network conditions (Chapter 4). Consequently, we propose heuristic methods for the dynamic multicast access technique selection and resource allocation, taking advantage of the multiuser diversity (Chapter 5.1). Moreover, we proposed a multicasting strategy based on fixed pre-computed multiple-input multiple-output (MIMO) multi-beams and S-NOMA (Chapter 5.2). Our approach tackles specific throughput requirements for enabling extended real-

ity (XR) applications attending multiple users and handling their spatial and channel quality diversity.

We address the computational complexity (CC) associated with the dynamic multicast RRM strategies and highlight the implications of fast variations in the reception conditions of the multicast group (MG) members. We propose a low-complexity ML-based solution structured by a multicast-oriented trigger to avoid overrunning the algorithm, a K-Means clustering for group-oriented detection and splitting, and a classifier for selecting the most suitable multicast access technique (Chapter 6.1). Our proposed approaches allow addressing the trade-off between optimal network performance and CC by maximizing specific QoS parameters through non-optimal solutions, considerably reducing the CC of conventional exhaustive mechanisms. Moreover, we discuss the insertion of ML-based multicasting RRM solutions into the envisioned disaggregated O-RAN framework (Chapter 6.2.5). We analyze specific MBS tasks and the importance of a native decentralized, softwarized, and intelligent conception.

We assess the effectiveness of our proposal under multiple numerical and link-level simulations of recreated 5G MBS use cases operating in μWave and mmWave. We evaluate various network conditions, service constraints, and users' mobility behaviors.

Chapter 1

Introduction

1.1 Research Motivation and Context

1.1.1 5G and Beyond

Even when the fifth-generation (5G) development recently started rolling out and seems at its beginning, Ericsson's mobile data traffic forecast [1] reports that the share of 5G mobile data traffic will grow to 69 % in 2028, where video traffic is estimated to account for 80 %. Moreover, by the end of 2028, five billion 5G subscriptions are forecasted globally, with 5G population coverage projected to reach 85 %. Its actual and predicted worldwide penetration levels have paved the way for shifting the research and industry attention toward future wireless networks. The 5G mobile system represents a step forward regarding spectrum usage, system capacity, network performance, and reliability [2]. Nevertheless, the new milestones are related to the enormous traffic growth, unprecedented heterogeneity, challenging applications, and anywhere/anytime best-connected expectations.

The envisioned sixth-generation (6G) era will represent a complete paradigm shift for global communications, merging the physical, digital, and virtual worlds through immersive human interaction. It must be developed to support a hyperconnected world and meet stringent requirements over challenging use cases that surpass the well-known enhanced Mobile Broadband (eMBB), Ultra-Reliable and Low-Latency Communications (URLLC), massive Machine Type Communications (mMTC), and Vehicular-to-Everything (V2X) [2]. A significant use case will be immersive and advanced experience-sharing communications, including extended reality (XR) (such as augmented reality (AR) and virtual reality (VR)), holographic communications, three-dimensional (3D) video delivery, and 4K/8K video streaming. 6G will enable extreme communication applications such as autonomous driving, telesurgery, mixing robotic technologies, flexible manufacturing, and seamless interaction with immersive applications. Such a wave of multimedia and experience delivery will align with an upcoming connected everything era. 6G massive

communications imply a hyperconnected resilient network infrastructure with an unprecedented diversity of end devices. These challenging use cases require meeting enhanced capabilities regarding throughput, end-to-end (E2E) latency, reliability, energy/capacity efficiency, and mobility [3]. Moreover, the future 6G network must embrace a green and sustainable approach with lower power consumption, longer life cycles, and less environmental impact [3].

The current study results of the International Mobile Telecommunications (IMT)-2030 promotion group position the *new networks with native artificial intelligence (AI)* as one of the ten prominent candidates 6G technologies [4, 5]. The 6G baseline network architecture must incorporate intelligence as an endogenous characteristic expanding artificial intelligence (AI)/machine learning (ML) solutions to the network E2E. Such vision is crucial for handling dynamic ultra-dense heterogeneous networks (HetNets) environments and enormous action spaces. This native conception will require a dynamic open radio access network (RAN) deployment with slicing support, such as the architecture promoted by the O-RAN Alliance, adding the virtualization, softwarization, and disaggregation expected in 6G [6].

Addressing the challenging research verticals shaping the 6G context will require a compelling mix of radio access technologies (RAT) and native softwarized/intelligent conceptions. In such a context, the multicast/broadcast service (MBS) [7] capability, introduced in Release (Rel)-17, brings alternative point-to-multipoint (PTM) delivery mechanisms for efficient resource utilization, overhead and delay reduction, load balancing, and reliability [2]. It enables considerable capacity gain through cost-effective and high-quality multicasting RRM solutions. The MBS delivery strategies are essential to face the requirements of massive multimedia content delivery and internet of things (IoT) deployments for disseminating early warnings and public safety. The following subsection presents the MBS evolution toward the envisioned 5G and beyond networks, potentials, and enabler technologies.

1.1.2 Multicast/Broadcast Service Evolution, Potentials and Enabling Technologies

One of the significant advancements in broadband service delivery was the development of Long Term Evolution (LTE) native evolved multimedia broadcast multicast service (eMBMS) over Rel-9 to 14 [8]. eMBMS paves the way for content providers and operators for cost-effective, high-quality service delivery to concurrent cellular users with common interests. It is based on single frequency network (SFN) over synchronized multi-cell transmissions providing over-the-air multicast-broadcast SFN (MBSFN) signals. MBSFN transmissions are time-interleaved with the unicast communication with pre-assigned and dedicated sub-frames over the radio frame and utilize the total system bandwidth [9]. Later, single cell point-to-multipoint (SC-PTM) was introduced in Rel-13 with a dynamic time and frequency resource utilization (even within a sub-frame) within a single-cell coverage [10]. The

dynamic resource utilization effectively allows the integration of broadcast service delivery with unicast physical channels [11]. Regardless, the drawbacks behind the LTE-based multicast/broadcast hurdle its acceptance and penetration among service providers.

The first phases of the 5G standardization (Rel-15 and 16) focused on the solo unicast capability [2], and the prominent use cases eMBB, URLLC, mMTC, and V2X [12]. Moreover, Rel-16 added the Terrestrial (T)-Broadcast targeting Enhanced Television (EN-TV) with an LTE multicast/broadcast capability. In Rel-17 [8], the 3rd Generation Partnership Project (3GPP) started standardizing the novel MBS paradigm as a native 5G multicast/broadcast capability. The standardization has been conducted for the overall system architecture, including next generation radio access network (NG-RAN) and 5G core network (5GC). Broadcast services over MBS inherit broadly the same requirements and design features as eMBMS and SC-PTM. However, MBS is built over novel 5G new radio (NR) characteristics such as bandwidth part (BWP), beamforming, absence of always-on reference signals, and variable subcarrier spacing (SCS). The key features of the standardized MBS can be summarized as follows [9]:

- Group-oriented scheduling mechanism which enables the user equipment (UE) to receive MBS applications, including simultaneous unicast service reception.
- Multicast/broadcast shared delivery in 5GC.
- Reliability enhancements by dynamic change of multicast service delivery between PTM and point-to-point (PTP).
- Supporting service continuity and lossless handover.
- Unicast/multicast/broadcast services simultaneous reception in the radio resource control (RRC) connected and idle states.
- MBS over legacy network nodes (3GPP Rel-15 and 16 networks).

Current Rel-18 [13] marks the start of the 5G-Advanced era. It aims to enhance the resource utilization efficiency in multicast/broadcast. It enables RAN sharing scenarios over heterogeneous networks and supporting the multicast reception in inactive RRC state [2, 14]. The envisaged NR MBS includes key performance indicators (KPIs) to reduce computational complexity and increase energy efficiency with ML solutions based on decentralized and softwarized network elements.

The 6G study and conceptualization will start from the time frame of Rel-20 with the first 6G specifications in Rel-21. The future 6G standard will respond to multiple key research verticals, merging the physical, digital, and virtual world through immersive human interaction; extreme disaggregation and virtualization over a scalable network architecture; ultra-secure and resilient communications; spectrum expansion to the THz bands and new spectrum sharing paradigms; green and sustainable

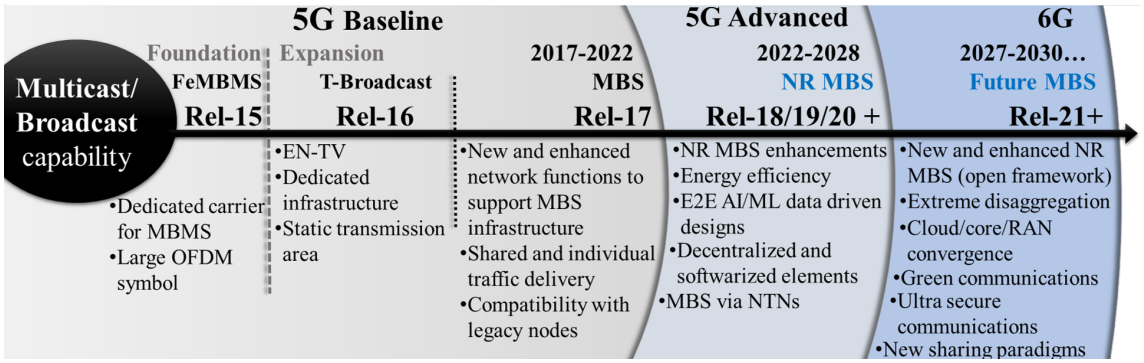


Figure 1.1: Multicast/Broadcast evolution toward 6G and main features of each release (further evolved multimedia broadcast multicast service (FeMBMS); non-terrestrial network (NTN); orthogonal frequency-division multiplexing (OFDM)).

communications with lower power consumption, longer life cycles, and less environmental impact [3].

The 6G RAN era will be characterized by groundbreaking use cases where a novel MBS conception will be prominent in reaching milestones through efficient resource utilization, overhead, and E2E delay reduction, load balancing, and reliability. Group-oriented communications allow for efficiently streaming content to large and small areas and offloading popular information to the network edge caching. The multicast/broadcast capability is identified as an essential technology for 6G massive vehicular IoT in disseminating early warnings and public safety information as a fundamental component of modern transportation systems. The envisioned MBS solutions must be embedded in the novel O-RAN framework with convergent cloud, core, and RAN through open interfaces [6]. Figure 1.1 illustrates the expected multicast/broadcast evolution from the baseline 5G to the upcoming 6G.

The seamless unicast/multicast/broadcast convergence in the 6G toolbox requires the integration of multiple enabling technologies such as those presented in Figure 1.2. These trending features belong to the longer-term new R&D wave toward 6G [15], summarized in the following:

- mmWave and THz communications with **MIMO** and **beamforming (BF)** are game-changing technologies for delivering high throughput group-oriented services while exploiting the users' spatial and channel diversity. Moreover, cell-free masive MIMO (mMIMO) allows extra spatial diversity and processing gain by simultaneously and coherently delivering unicast and multicast services through multiple geographically distributed base stations (BSs) [8].

- The **non-orthogonal multiple access (NOMA)** technologies are essential for future wireless networks' mixed unicast/multicast/broadcast service delivery [16]. NOMA empowers the network with seamless connectivity, secure

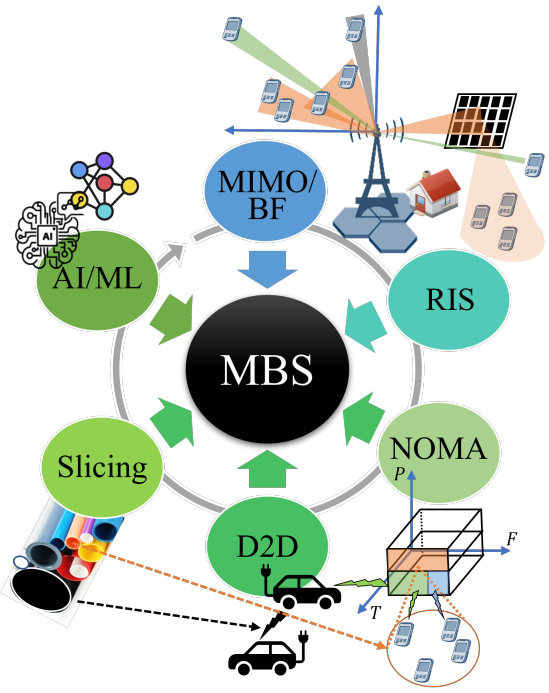


Figure 1.2: The MBS enabling technologies.

transmission strategies, improved fairness, and reduced outage probability.

- The **reflective intelligent surfaces (RISs)** can reduce blockage effects and improve the reception conditions of the worst channel quality users and the corresponding multicast group (MG) quality of service (QoS). Furthermore, RIS is considered a sustainable and ecologically friendly solution based primarily on passive components [3].

- The proximity technology **Device-to-Device (D2D)** communications and specifically D2D underlaid multicasting (D2DM) are cost-effective solutions for group-oriented communications with users in the vicinity, reducing latency, handling diversity, enabling alternative links, and extending the coverage [17].

- The **network slicing** paradigm adds flexibility, dedicated prioritization, dynamism, and isolation by creating independent network slices (NSs) over a physical infrastructure. NSs are critical in the upcoming MBS conception to manage differentiated traffic in real-time efficiently [2, 18].

- **AI/ML** solutions are crucial for handling ultra-dense ultra-dense heterogeneous networks (HetNets), enormous action spaces, and highly dynamic network/services/user setups. The effective integration of the above MBS-enabling technologies lies in a native intelligent RRM conception.

1.1.3 Multicast/broadcast Challenges

Despite the enormous advantage that the envisioned MBS supposes for future networks, its current development stage is far from the expected requirements with multiple open challenges. In this subsection, we discuss the challenges and the research problems identified under the MBS umbrella that motivate this investigation.

The pass forward from the current 5G technological development state to the future 6G requires meeting advanced KPIs such as 0.1 ms of latency, users' throughput of up to 1 Gbps, peak data rate of up to 1 Tbps, users mobility at high speeds of up to 1000 km/h, reliability of 99.9999999 %, a capacity density of 500 Mbps/m^2 with connection density of up to 10^8 devices per km^2. The energy and capacity efficiency are expected to improve the 5G numbers by up to five and three times, respectively [3]. In such a context, guaranteeing the stringent QoS requirements becomes a challenge regarding the contracted service level agreement (SLA) by content providers and tenants. Moreover, 6G needs to encompass the principle of human-centric networks targeting users' quality of experience (QoE) satisfaction aside from network-centric QoS optimization [19]. However, in 6G, quality of physical experience (QoPE) will be considered, incorporating human physiological factors [20]. In the following, we will focus on QoS optimizations, as QoE is mainly based on subjective evaluations that are out of the scope of this research.

Ensuring QoS requirements for MBS applications is even more complicated regarding unicast delivery due to the diversity in terms of reception conditions of the MG. Group-oriented multicast applications suffer a higher probability of throughput performance deterioration caused by the reception limitations of the group members with the worst channel conditions. Consequently, multicasting without a tailored strategy can degrade the QoS of the whole MG or produce an unfair resource allocation. It happens because the conventional multicast scheme (CMS) treats all the MG according to the lowest channel quality user, ignoring the users' diversity [21]. The CMS approach could allocate more radio resources (if possible) to back the lower-channel quality users, but this can produce an inefficient resource utilization or degrade the QoS of other applications.

To ensure the stringent QoS/QoE requirements, novel multicasting strategies must be aware of the trade-off between the available multicasting gain and the existing multiuser diversity [22]. The multicast gain results from multiple users requesting the same content, which can be treated as a MG. Moreover, these users experience diverse reception conditions. Effective trade-off management profiting from the MG users' diversity could maximize performance and improve resource utilization. In recent years, several works such as [23–27] have addressed this challenge with solutions based on splitting the MG members according to their specific reception conditions and using a multi-rate group-oriented modulation and coding scheme (MCS) to deliver the service.

Current broadband multimedia technologies and multi-rate applications are based on orthogonal multiple access (OMA) [28]. Nevertheless, NOMA enables an alter-

native waveform domain rather than the conventional time-frequency lattice [29]. Recent works, such as [30–32], compared the performance gain of NOMA over OMA in single-antenna and multi-antenna systems with single-cell and multi-cell deployments based on KPIs such as sum rate, fairness, and energy efficiency. These proposals presented NOMA as an essential technology in future wireless networks, improving efficiency and flexibility.

Over the years, multicasting has been associated with omnidirectional communication at sub-6 GHz frequency bands [33]. In the last years, multicasting has gained momentum in highly directional multi-beam communications in mmWave and sub-THz bands with massive MIMO [26]. High-frequency propagation brings new challenges related to high path loss, severe signal attenuation due to blockage (e.g., human blockage of 15 dB), and reduced coverage [34]. As the frequency increases, the variation in the users' reception conditions, signal-to-interference-plus-noise ratio (SINR), and feedback channel quality indicator (CQI) also increases. The ineffective handling of these impairments considerably reduces the MG's QoS. Multicasting at these high-frequency bands is complex because directional beams usually cover a small angular area and must be steered toward the right direction, dynamically adjusting the beamwidth, switching to multiple beams, and managing the beams' gain and power, subject to (s.t.) the users' distribution.

On the other hand, handling complex mobility behaviors at high frequencies adds extra difficulties. The mobility behavior and speed of the multicast users are directly correlated with the variations in the users' channel conditions and MG diversity, increasing the probability of throughput performance deterioration. Moreover, adequate directionality is necessary to achieve an optimal link budget, to sustain high-capacity connections, adequate beamforming, and precise beam alignment, including user/group tracking, handover, and radio link failure recovery [8]. Managing high-speed users implies effective beam-switching and tracking to ensure uninterrupted communication. During multicast service delivery, the beam should track the most significant part of the MG, which is more challenging since users may follow different mobility patterns [8].

Due to the strong directionality that high-frequency multicast multi-beam communications imply, even small movements of the MG members may drastically change the QoS due to blockage, misalignments, and severe variations in the SINR, and feedback CQI. Significant variations in the channel quality conditions of the MG members impose recalculating the optimal multicast access technique and RRM strategy. Under fast variations in the reception conditions of the MG members, non-optimized multicast RRM solutions could exponentially increase the computational complexity (CC) and associated delay. A continuous RRM recalculation toward an optimal solution could not be tolerated in multiple MBS applications.

Recent investigations about RRM have been chiefly oriented toward optimizing resource management, spectrum utilization, and interference mitigation, but the associated complexity has been given minor attention. Nevertheless, upcoming use cases will significantly increase the network complexity, adding new computa-

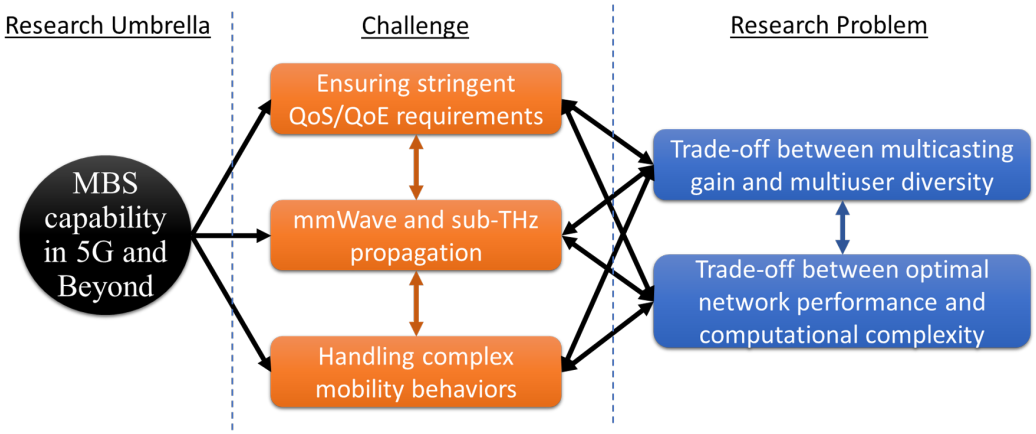


Figure 1.3: Challenges toward reaching the envisioned MBS capability.

tional levels, constraints, and hardware needs to provide seamless connectivity and real-time response [35]. For delay-aware services, high induced latency cannot be tolerated during the RRM, making the CC critical during the solution design [36]. The CC must be considered a KPI during the multicast RRM solution-finding. In such a context, ML-based solutions can relax the RRM-associated computational burden and achieve an acceptable trade-off between network performance and CC [37]. This intelligence-native conception will require flexible architectures such as O-RAN to add the virtualization, softwarization, and disaggregation expected in 6G.

Despite the enormous benefits that MBS suppose, the dimension of the envisioned 6G, the dynamic nature of the MBS use cases, and the stringent service requirements make an efficient multicast/broadcast service delivery challenging. Integrating MBS into the 6G toolbox requires embracing well-established PTM delivery mechanisms and novel top-notch technologies (Figure 1.2) with seamless PTP communications convergence. The resulting MBS capability will evolve into a complex technology from the RRM point of view. Additionally, the forthcoming transformation will be linked to a 3D ultra-dense HetNets with differentiated services, stringent requirements, and the always best-connected (ABC) paradigm. As a result, managing and allocating network resources will become even more complex [38].

Figure 1.3 illustrates the above-identified challenges and research problems under the MBS umbrella in 5G and beyond networks. We identify as a primary challenge the necessity of ensuring the stringent QoS/QoE requirements of the use cases where MBS is a core element. Meeting these strict requirements implies effectively embracing the mmWave and sub-THz propagation. These frequencies could enable many use cases but at the cost of higher hardware/software needs and increased final complexity. The third challenge is handling complex mobility behaviors that characterize an MG of multiple members with different mobility patterns and velocities.

As we analyzed, the impact of the MG mobility behavior is directly related to the operation frequency and the QoS/QoE requirements.

The two identified and discussed research problems in Figure 1.3 are linked to successful challenge management and future MBS deployment. Figure 1.3 describes a fully connected relationship among the identified challenges and the trade-offs between multicasting gain and multiuser diversity and between optimal network performance and CC. Then, addressing these research problems and dealing with these critical challenges means advancing the knowledge and state-of-the-art contributions regarding the multicast/broadcast capability and the envisioned future networks. The timing for being part of this research work is perfect since the technological development and user/industrial needs regarding a better and hyperconnected world are continuously ongoing. Navigating through these research problems and challenges under the multicast/broadcast umbrella, an umbrella covering me for multiple years, is a powerful motivation for this Ph.D. study.

1.2 Objective

Once the research overview and motivation have been explained, we present the main goal and specific objectives that guide this research. This investigation covers essential aspects at the intersection of MBS, RRM, ML, and the O-RAN framework. We aim to **characterize and address the dynamic multicast multiuser diversity through low-complexity RRM solutions aided by ML and OMA/NOMA in 5G MBS and beyond networks**. To reach this goal we define several specific objectives (SOs):

- **(SO-1) Analysis of the existing multicasting gain and multiuser diversity management solutions in the 5G MBS context and beyond**. Detailed state-of-the-art analysis and definition of the principal theoretical concepts covered in the research. Our study delves into the CC-aware solutions with sub-optimal methods.

- **(SO-2) Definition of the system model and problem formulation for addressing the dynamic multicast multiuser diversity**. Definition of the mathematical notations that will be used along the manuscript, as well as constraints and recreated network conditions. After stating the research conditions, we formulate the main problems to be addressed.

- **(SO-3) Characterization and identification of the conditions and variables interrelation for an effective coexistence of CMS, subgrouping based on OMA (S-OMA) and subgrouping based on NOMA (S-NOMA)**. Integral analysis of the variables that characterize the performance of the evaluated multicast access techniques. Identify the conditions and applications suited for each one. We consider the dynamic behavior of

these techniques regarding users' reception conditions, multimedia service constraints, and network parameters.

- **(SO-4) Proposal of a dynamic multicast access technique selection and resource allocation algorithms**. We propose heuristic-based algorithms to prove the QoS advantage of dynamically selecting among the traditional single-rate strategy CMS and the multi-rate subgrouping-based solutions. We extend the analysis scope by designing an algorithm that combines multicasting over fixed pre-computed MIMO multi-beams and multi-rate subgrouping, taking advantage of the users' spatial and channel quality diversity. We evaluate the proposals through numerical and link-level simulations recreating various network conditions. The CC analysis is not considered a KPI. We focus our solutions on maximizing the QoS.

- **(SO-5) Design of ML-based low-complexity multicast access technique selection and resource allocation algorithms and their insertion into the O-RAN framework**. We propose novel low-complexity multicast RRM strategies for dynamic access technique selection and resource allocation. Our proposal is oriented to address and contextualize the complexity associated with the multicast resource allocation process and the implications of fast variations in the reception conditions of the MG members due to the users' mobility behaviors and the impact of mmWave propagation. We propose sub-optimal solutions for the QoS improvement with particular attention to the CC as a KPI. We evaluate the proposals through link-level simulations recreating various network conditions. We analyze the insertion of ML-based multicasting RRM solutions in the O-RAN framework.

1.3 Contributions

The research's main contributions and related publications can be summarized as follows:

- We define equations to assess the performance among the considered multicast access techniques: CMS, S-OMA, and S-NOMA. We define the interrelations among the variables that shape the performance of the evaluated multicast access techniques and their dynamic behavior in terms of the users' reception conditions, multimedia service constraints, and network parameters. We provide conditions for an adequate dynamic selection of the multicast strategy that better suits the specific network characteristics. We prove how the evaluated multicast access techniques must coexist to provide an effective resource allocation subject to particular network conditions.

  Related Publications:

E. F. Pupo, C. C. González, L. Atzori and M. Murroni, "Thresholds of outperformance among Broadcast/Multicast access techniques in 5G networks," 2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Chengdu, China, 2021, pp. 1-6, doi: 10.1109/BMSB53066.2021.9547169 [39].

E. F. Pupo, C. C. González, E. Iradier, Montalban, J., Angueira, P., M. Murroni (2023). Dynamic Single/Multi-Rate Multicasting Aided NOMA for Addressing the Multiuser Diversity in 5G Networks. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.24312049.v1 [40].

- We provide a method for the dynamic multicast access technique selection and resource allocation based on the proposed performance equations. The proposal shows the advantages of dynamically selecting among CMS, S-OMA, and S-NOMA to effectively handle the trade-off between the available multicasting gain and the existing multiuser diversity. We propose a multicasting strategy based on fixed pre-computed MIMO multi-beams and S-NOMA, tackling specific throughput requirements for enabling XR applications to attend multiple users in a 5G MBS use case. The following approach allows us to dynamically take advantage of the users' spatial and channel quality diversity, maximizing specific QoS metrics.

  Related Publications:

  E. F. Pupo, C. C. González, L. Atzori and M. Murroni, "Dynamic Multicast Access Technique in SC-PTM 5G Networks: Subgrouping with OM/NOM," 2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Bilbao, Spain, 2022, pp. 1-6, doi: 10.1109/BMSB55706.2022.9828674 [23].

  E. F. Pupo, C. C. González, V. Popescu, D. Giusto and M. Murroni, "Beyond 5G Multicast for XR Communications aided by Pre-computed Multi-beams and NOMA," 2023 IEEE Global Communications Conference (GLOBECOM), Kuala Lumpur, Malaysia, 2023, pp. 1-6 (Accepted paper).

- We address the CC associated with the dynamic multicast RRM strategies in 5G MBS use cases and highlight the implications of fast variations in the MG members' reception conditions. We propose two novel solutions for the multicast access technique selection based on ML multiclass classification algorithms with multi-layer perceptron (MLP) and extra tree classifier (ETC). We propose using a K-means clustering unsupervised ML approach for detecting and splitting group-oriented MGs based on the CQI values. We propose a multicast-oriented trigger to avoid overrunning the entire algorithm subject to the temporal variations of the MG's CQI distribution, reducing the induced latency over the time slot lattice. Our proposed approaches allow addressing

the trade-off between optimal network performance and CC by maximizing specific QoS parameters through non-optimal solutions, considerably reducing the CC of conventional exhaustive mechanisms. We characterize the ML-based multicasting RRM insertion in the O-RAN framework under a softwarized and intelligent vision.

Related Publications:

Pupo, E. F., González, C. C., Montalban, J., Angueira, P., Murroni, M., & Iradier, E. (2023). Artificial Intelligence Aided Low Complexity RRM Algorithms for 5G-MBS. IEEE Transactions on Broadcasting [41].

Pupo, E. F, González, C. C., Iradier, E., Montalban, J., Angueira, P., Murroni, M. (2023). Machine Learning-based Multicasting Radio Resource Management over 6G O-RAN Framework. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.24408250.v1 [15].

- We assess the effectiveness of our proposal under multiple numerical and link-level simulations of recreated 5G MBS use cases operating in μWave and mmWave, evaluating a wide range of network conditions and users mobility behaviors. The carried link-level simulations were based on an implemented ad-hoc simulator developed in Python to facilitate the studies on 5G MBS. The simulator combines unicast/multicast/broadcast capabilities over terrestrial and airborne network deployments.

  Related Publications:

  Pupo, E. F, C. C. González, E. Iradier, J. Montalban and M. Murroni, "5G Link-Level Simulator for Multicast/Broadcast Services," 2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Beijing, China, 2023, pp. 1-6, doi: 10.1109/BMSB58369.2023.10211507 [42].

The defined research goal and specific objectives guided the investigation process, enabling the characterization and addressing of the identified research problems as illustrated in Figure 1.4. The figure shows the interrelation between the two research problems and the main contributions, classifying the contributions regarding the SOs. The contributions derived from each SO are explained and supported by the results and analysis in the upcoming chapters.

1.4 Organization of this Document

The thesis is structured in seven chapters plus the appendices, bibliography, and list of acronyms. First, Chapter 1 presents an overview of the research umbrella and the motivation, establishes the objective and contributions, and finally gives

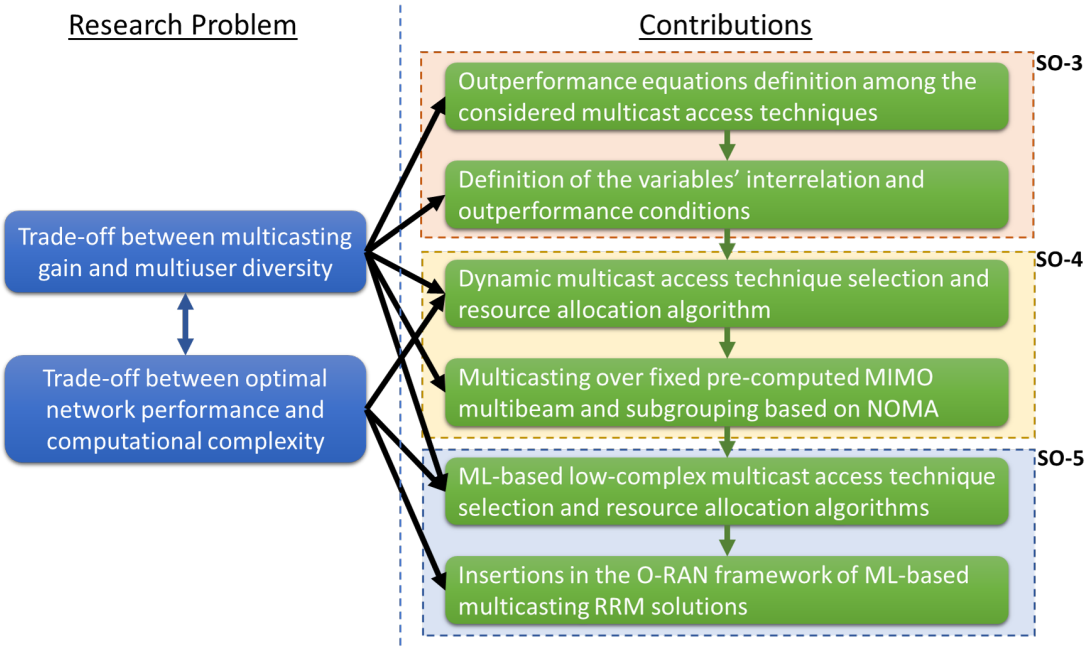


Figure 1.4: Main contributions.

the thesis outline. Chapter 2 overviews the existing related works and main theoretical concepts to better understand the remainder of the paper. Then, Chapter 3 establishes the system model and problem formulation. The core content regarding the objectives, contributions, and associated publications are detailed in Chapter 4, 5, and 6 as follows:

- Chapter 4 presents the characterization and identification of the conditions and variables interrelation for an effective coexistence of CMS, S-OMA, and S-NOMA. We realize the analysis of the variables that characterize the performance of the evaluated multicast access techniques.

- Chapter 5 proposes heuristic-based algorithms to prove the QoS advantage of dynamically selecting among the traditional single-rate strategy CMS and the multi-rate subgrouping-based solutions. Moreover, an algorithm that combines multicasting over fixed pre-computed MIMO multi-beams and multi-rate subgrouping is presented, taking advantage of the users' spatial and channel quality diversity. The proposals are evaluated through numerical and link-level simulations recreating various network conditions.

- Chapter 6 proposes novel low-complexity multicast RRM strategies for dynamic access technique selection and resource allocation. The chapter delves into the complex implications of fast variations in the reception conditions of the MG members due to the users' mobility behaviors and the impact of

mmWave propagation. The proposals are evaluated through link-level simulations recreating various network conditions. Moreover, the chapter analyses the insertion of ML-based multicasting RRM solutions in the O-RAN framework.

Finally, the general conclusions and future research directions are drawn in Chapter 7. The thesis concludes with the appendices, including the main mathematical notations and the list of acronyms. At the end, the Bibliography is presented.

Chapter 2

Related Works and Theoretical Background

In Chapter 1, we presented an overview of the research umbrella and the motivation, established the objective and contributions, and finally gave the thesis outline. As we analyzed in Subsection 1.1.3, addressing the trade-off between multicasting gain and multiuser diversity and between optimal network performance and CC are the baseline of our approach and contributions. Therefore, this Chapter gathers the related state-of-the-art publications and theoretical background, which provide the basis for the upcoming discussions and results.

2.1 Challenging Applications and Role of the Multicast Capability

The envisaged 5G and beyond networks offer unprecedented breakthroughs in media service delivery, novel capabilities, and enhancing disruptive services [43]. The next realm in communication technology is immersive communications, including XR, haptic, and holographic applications. The shift from two-dimensional (2D) to 3D displaying methods and additional multisensorial content will blur the boundaries of the physical and the virtual worlds [44]. Hence, immersive 3D communications are expected to profoundly impact the landscape of communication industries and our lifestyle for the following years while paving the way for the metaverse paradigm [45].

The required data rate for conveying live 3D images, especially for 360 high-resolution video streaming, will rise to terabits per second (Tbps), a value that can hardly be supported by 5G. Second, the system latency for delivering multisensorial data, such as haptic information, must be as low as a few milliseconds to guarantee an adequate user experience [46]. Last but not least, the synchronization issues that arise from the orchestration of the video streams provided by multiple cameras or

Use Case	Throughput	Latency	Reliability
HD	12 Mbs	$\leq$ 250 ms	$--$
UHD and 4K	20-50 Mbps	15-35 ms	$--$
8K	100-140 Mbps	15-35 ms	99.99999 %
XR	0.025-5 Gbps	5-7 ms	99.99999 %
VR entertainment	0.02-3 Gbps	5-10 ms	99.99999 %
AR entertainment	0.02-1 Gbps	20 ms	99.99999 %
Digital twin of smart city	10 Mbps	5-10 ms	99.99999 %
AR smart healthcare	10 Gbps	5 ms	99.999999 %
Holographic	4-10 Tbps	sub ms	$--$
eHealth/remote surgery	1 Gbps	$<$ 1 ms	99.99999 %
Industrial automation	Gbps order	0.1-1 ms	99.9999999 %
Autonomous mobility	$--$	$<$ 1 ms	99.99999 %

Table 2.1: Enhanced use cases' requirements.

data from various sensors, including haptic information, induce new challenges in the communication subsystems.

Most envisioned multimedia applications require a high data throughput, with negligible latency and extreme reliability, so that the users can experience a truly immersive experience. Table 2.1 summarizes some prominent multimedia use cases and their requirements as defined in [8, 19, 45, 47–49]. For an application such as XR, the user can experience dizziness and suffer from high QoE degradation if the requirements are not accomplished. These requirements must be addressed by embracing the best available and upcoming technologies at the physical and networking level [50].

For bandwidth-demanding applications, especially for a dense user deployment, multicast traffic delivery can significantly benefit communication and computing performance. This capability provides cost-effective and resource-efficient delivery mechanisms to multiple end-users requesting the same content [9, 51]. Tailored PTM communication strategies can provide considerable capacity gain into the beyond 5G ecosystem, being an essential element of the RRM toolbox [2]. In [52], the authors analyzed the integration of 6G NTNs and the MBS capability to enable scalable services through efficient broadcasting strategies, streaming content to large areas, and offloading popular content to the network edge caching. In [53], Zhou *et al.* delved into the importance of the MBS capability for massive vehicular IoT in emergency information delivery as a fundamental component of modern transportation systems.

In [54], the authors propose a smart mode selection using reinforcement learning for VR mixed broadband broadcasting aided D2DM in 5G HetNets. The proposed approach outperforms the benchmark solutions with a VR broadcasting strategy that improves the mean throughput of the system. In [55], Zhong *et al.* present a

decentralized optimization for multicast adaptive video streaming in edge-caching assisted networks. The authors analyze how applications based on omnidirectional video with ultra-large bandwidth, such as VR and holographic, heavily rely on adaptive video streaming to deliver multiple streams of high-definition content, s.t. the users field-of-view (FoV). In such a context, ubiquitous edge-caching and multicast support can enable a large scale of low latency video services, as above presented in Table 2.1. In this research, the authors emphasize the essential role of multicasting technologies and edge-caching for the next generation of communication technologies. Moreover, in [56], the authors present a multicast-aware optimization for resource allocation combined with edge computing and caching. In this approach, the authors realize personalized computing at the edge of users with different FoV, multicasting the service to the users interested in the same stream. Such mechanisms avoid delivering the whole content to all users, reducing bandwidth consumption s.t. the user FoV and reception conditions. In [57], the authors propose a scalable multicast solution for live 360-degree video streaming over mobile networks in a bandwidth-efficient manner.

In [48], the authors define that VR broadcasting is the most direct method that aids wireless users in getting access to the metaverse. The VR broadcast is transmitted in a heterogeneous manner in the metaverse due to the heterogeneous nature of wireless and edge devices. The authors discuss how the users could access the metaverse throughout macrocell broadcasting, mmWave small-cell unicasting, and D2DM. Additionally, in [45], it is discussed how D2DM allows to efficiently utilize the available communication resources of VR users. In this proposal, the users dynamically participate in multicast clusters, sending shared reused tracking signals to service providers and increasing the total bit rate within the cell.

In [58], a multicast and broadcast QoS enhancement is presented with a flexible service continuity configuration. The solution intends to minimize the signaling overhead while meeting the latency requirements through a dynamic service continuity over 5G. Moreover, in [59], the authors propose an ML-based 5G RAN slicing for broadcasting services. The solution applies ML to estimate and predict the channel status in mobile scenarios over the MBS context. In [60], a priority-aware resource allocation algorithm is applied for 5G mmWave MBS service delivery. The authors in [61] present a hybrid terrestrial-airborne connectivity for MBS application over beyond 5G networks. This work proposes an access network selection strategy for a softwarized terrestrial networks (TN)/airborne connectivity to satisfy multiple high-throughput demanding service requests. In [34], Brancati *et al.* presents a RIS deployment and orientation strategy aided by ML to deliver MBS applications over high-throughout mmWave multi-beam multicasting.

In [8], the authors comprehensively analyze how the multicast capability is crucial for approaching 6G trending use cases. The authors delve into the main multicast challenges and enabling solutions. Moreover, in [2], Carballo *et al.* examines the role of multicasting over 6G NTN. They evaluate the potential of exploiting the softwarization paradigm in the heterogeneous 3D TN–NTN architecture in the delivery

of multicast services. In [15], we discuss the insertion in the 6G O-RAN framework of ML-based multicasting RRM solutions. The research covers essential aspects at the intersection of MBS, ML-based RRM solutions, and the disaggregated O-RAN architecture, identifying possible scenarios as feature extensions of O-RAN.

2.2 Multicast Access Techniques

As defined in [62], PTM communications simultaneously serve all users interested in a multicast service through shared channels and physical resources. The PTM communications aim to improve the system capacity and, theoretically, serve an unlimited number of users [63]. The multicasting techniques are divided into single-rate and multi-rate [64]. The single-rate solutions are based on CMS [65], where the selection of the MCS focuses on the users with the worst channel quality conditions in the MG. In particular, according to the reception conditions experienced by the UEs belonging to the MG, the RRM performs the link adaptation procedures, that is, the selection of the most appropriate transmission parameters (i.e., the MCS) for delivering the multicast content. The CMS maximizes the system coverage by enabling serving the entire MG and exploiting the available multicasting gain with the MCS imposed by the users with the lowest SINR. In [66, 67], the authors propose the single-rate opportunistic multicasting scheme (OMS) to reduce the CMS drawbacks. This technique improves the system capacity by multicasting only to the portion of the MG with better channel quality conditions. This solution maximizes the QoS of the served users and the resulting system capacity, but it does not guarantee the reception of all the users requesting the same content. Although single-rate schemes present advantages in terms of simple implementation and low complexity, these traditional solutions do not handle the multiuser diversity and tend to suffer from a low capacity efficiency and unfair resource allocation [21, 68].

In [63], the authors define that two main strategies are considered for multi-rate multicast transmission: stream and group splitting. The first is based on splitting high-data rate multimedia contents into multiple lower data rate substreams [69]. During the service delivery, a base substream is received by all users, ensuring complete coverage of the MG. Then, the good channel quality users receive the additional enhancement streams, improving their QoS. Typically, scalable video coding (SVC) is the baseline technology of this approach [70, 71]. In [72], Mao *et al.* present a detailed survey including the future research trend and fundamentals of rate-splitting multiple access (RSMA). RSMA is a flexible and scalable framework to optimize non-orthogonal transmissions and facilitates interference management in heterogeneous environments. Nevertheless, the multicast rate-splitting strategies are out of the scope of this research.

In the traditional multi-rate schemes, the MG members are served with different MCSs according to the particularities in the channel quality diversity (CQD) defined by the spatial users' distribution (or spatial diversity (SD)) and the heterogeneity of

the wireless channels [26]. As described in [63], an effective group splitting approach consists of dividing the multicast UEs into subgroups and applying group-oriented resource allocation based on adaptive MCSs. This approach overcomes the limitations of the single-rate techniques by partially exploiting the multiuser diversity. The subgrouping formation may be performed according to an optimization problem that can be formulated to achieve different goals, such as maximizing system throughput, fairness index, and spectral or energy efficiencies. The main issue faced by the subgrouping approach is the CC. In [68], the authors propose the subgrouping techniques as a practical scheme to serve all the users requesting the same content. They suggest allocating group-oriented resources to each MG subgroup based on the user CQI diversity. This solution allocates the resources to the different users applying S-OMA as the baseline multi-rate technology.

In [73], the authors propose an RRM strategy for group-oriented services in LTE advanced systems. The proposal is based on policies such as serving all the MG members and maximizing the system throughput to support effective multicast content delivery. The main challenge of the subgrouping techniques is the CC for an effective subgroup creation and resource allocation that maximizes a specific cost function. To simplify such complexity, the authors in [74] prove that the subgrouping techniques achieve the best results when creating just two subgroups, splitting the low and high channel quality users in the MG. This research's outcomes align with the results presented in [64], where the authors analytically demonstrate that the optimal number of subgroups is not greater than two. Regarding this, in [75], the authors propose a method to tailor broadcasting and multicasting transmission in 5G NR. The primary outcome of this proposal is to validate through numerical simulations that increasing the number of subgroups does not proportionally increase the system performance. The extra gain achieved when using more than two subgroups is marginal. Therefore, the authors recommend creating two subgroups for multicast group-oriented service delivery.

Current multi-rate applications over broadband multimedia technologies are based on OMA, such as time-division multiplexing (TDM) and frequency-division multiplexing (FDM) [28]. Nevertheless, in [76], the authors propose a new multiplexing access technology for resource allocation: the power-domain NOMA (P-NOMA) based on layer-division multiplexing (LDM) [77, 78]. The proposal is evaluated by applying the subgrouping technique for providing multicast services in LTE advanced over vehicular networks. P-NOMA based on LDM can be implemented to deliver simultaneously multiple services by assuming different power levels with a non-orthogonal resource allocation (i.e., neither in frequency, time, nor code domain). The main advantage of NOMA over OMA systems is the 100 % use of the radio frequency resources and 100 % of the time to transmit in a group-oriented multicast approach. In this context, the RRM assigns a different weighted power to each multiplexed service, performing layered multicast service delivery. In [63], the authors exploit the LDM concept, splitting the total available power to perform the multicast subgrouping. The different LDM layers are matched with the created

subgroups according to the users' reception conditions.

In [24], the authors propose using S-NOMA in the envisaged 5G environments, where different quality video services are delivered to a group of users interested in the same content. The results show the S-NOMA outperformance over S-OMA in specific use cases where group-oriented users experience orthogonal channel conditions. Moreover, in [79], the authors propose an advanced NOMA-based RRM schemes for broadcasting in 5G mmWave frequency bands. This article designs and evaluates different models that combine OMA (TDM) with NOMA for innovative applications in the 5G mmWave. In [80], the authors propose using NOMA for enabling the unicast/broadcast convergence in 5G networks. The authors in [25] present a resource allocation algorithm for layered multicast video streaming using NOMA.

In the last years, multiple works have been specifically oriented to compare the performance gain of NOMA over OMA, such as [31, 32, 81–83]. The authors in [81] provide a comparative study of different multiple access techniques. It is demonstrated that non-orthogonal approaches have a spectral-power efficiency advantage over orthogonal ones for delay-sensitive applications in fading environments. Ghosh *et al.* [32] demonstrate that NOMA outperforms the OMA-based solutions for mmWave mMIMO communications, providing significant improvements in spectrum and energy efficiency and outage probability. The authors affirm that NOMA empowers the networks with seamless connectivity and provides a secure transmission strategy for the industrial IoT. In [30], the authors prove that NOMA outstrips the OMA scheme regarding average sum rate and energy efficiency in a system where relay nodes aid the communication between the BSs and the cell-edge users. The results show that the NOMA scheme provides better average fairness in the evaluated system. Recent studies, such as [80, 84], prove that NOMA could improve service capacity over OMA technologies by delivering multi-rate broadcast services in the same radio resources. In [84], the authors propose adding NOMA to the technology toolbox for 5G and beyond MBS implementations and characterize the capacity benefits that it could bring.

In recent years, spatial splitting has gained momentum for multi-beams MIMO multicasting in highly directional communications over mmWave. The spatial splitting techniques use the users' SD to deliver multicast group-oriented services. The service is provided over multiple beams steered toward the right direction, dynamically adjusting the beamwidth and the number of beams and managing the beams' gain and power. In [85], the authors propose an incremental multicast grouping scheme for mmWave networks with directional antennas. The principal outcome of this research is proving the advantage of using adaptive multicast beamforming over fixed precomputed beams. In [86, 87], it is analyzed how the combination of multicast and mmWave presents many issues, starting with how the beams should be shaped regarding the users' SD and also the CQD. The results show how restricting the wireless links to be unicast only may be strongly suboptimal. In [88, 89], an efficient solution for managing multicast traffic in directional mmWave networks is

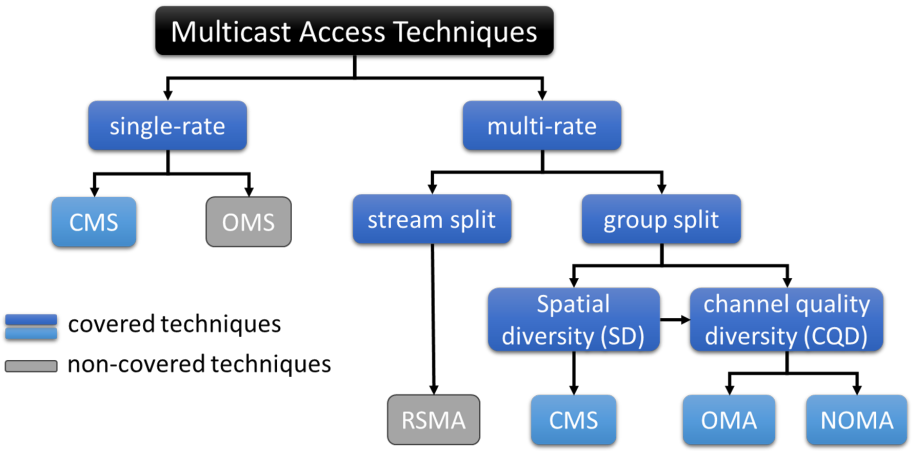


Figure 2.1: Covered multicast access techniques.

proposed. In particular, the authors provide a solution to perform the beam switching and propose an RRM policy to determine the number and width of the beams required to deliver the multicast content to all interested users. In [27], the authors present the use of ML techniques for optimal multi-beams multicasting over 5G small cells. Moreover, in [26], the authors propose a different strategy for exploiting the SD. They present a subgrouping method for the multicast users based on the similarity of their channel correlation matrices.

In Figure 2.1, we summarize the multiple concepts presented in this subsection. Figure 2.1 does not exclude other existing technologies and strategies that are not included in this review since they are considered out of the scope of the thesis. We highlight that our proposal focuses on the multicast access technologies CMS, and subgrouping based on OMA/NOMA (S-OMA, S-NOMA).

The above investigations focus on specific use cases and conditions oriented to prove the advantage of multicasting through multi-rate approaches based on OMA or NOMA over the traditional single-rate CMS or OMS. Nevertheless, none of the previous works realize an integral analysis to identify the specific conditions and applications suited for each one of these multicast access techniques. Moreover, the discussed solutions do not consider a dynamic multicast RRM approach profiting from the advantage of such different multicast strategies. None of these investigations characterize the interrelations of the multiple network variables that shape the performance of the multicast access techniques. Therefore, our proposal aims to complement and fill the above-analyzed gaps by accomplishing **SO-3** and **SO-4**. We address the effective coexistence of CMS and opportunistic S-OMA and S-NOMA in 5G and beyond networks. Moreover, we define the interrelations among the variables that shape the performance of the evaluated multicast access techniques and their dynamic behavior in terms of the users' reception conditions, multimedia service constraints, and network parameters. We provide conditions for an effective

dynamic selection of the multicast strategy for the specific network characteristics, handling the multicasting gain and multiuser diversity trade-off.

To the authors' knowledge, all the proposed solutions in the literature that aim to deliver multicast group-oriented services over multiple beams based on the users' SD perform the resource allocation employing CMS. These conservative approaches inherit the above-discussed weaknesses of CMS. Our proposal aims to fill this gap (**SO-4**) by applying a dynamic double subgrouping strategy. First, we take advantage of the users' SD and CQD and dynamically consider the multicast access techniques CMS, S-OMA and S-NOMA. Then, we assess the effectiveness of our contributions simulating a realistic 5G MBS use case operating in mmWave.

2.3 Computational Complexity of the Multicast Radio Resource Management

Future communication systems are becoming more sophisticated as they must meet growing user needs, increased data rates, massive connections, lower latencies, and extreme reliability requirements [36, 90]. In such context, the RRM is progressively increasing its critical role to deal with 3D ultra-dense HetNet, multiple RATs, and the ABC paradigm, as described in [35, 91, 92]. As defined in the survey [93], the RRM plays a pivotal function in enhancing spectrum utilization, load balancing, and network energy efficiency. As defined in [35], recent research about RRM has been chiefly oriented towards optimizing resource management, spectrum utilization, and interference mitigation. However, the associated complexities have been given minor attention. For example, in [93], the authors identify as RRM challenges the interference management, user association, resources, and power allocation. However, upcoming RRM solutions and use cases like 5G MBS significantly increase the network and communication complexity, adding new computational levels, constraints, and hardware needs to provide seamless connectivity and real-time response [35]. For example, during delay-sensitive service delivery, high induced latency cannot be tolerated during the RRM, making the CC critical during solutions design [36].

The MBS capability from the RRM point of view will evolve into a complex technology where the CC becomes critical and must be considered a KPI during the optimal solution-finding. In [90], the authors define the CC concept as "*the amount of processing required to acquire information, decide on resource allocation, and relay the results back to their intended users. It includes the difficulty of calculations involved when executing the resource allocation algorithms*".

As discussed in [91], ML-based solutions can relax the RRM-associated computational burden and provide an acceptable trade-off between network performance and CC. This survey analyzes how ML-based solutions can improve the performance of complex RRM tasks based on network' QoS parameters while maintaining an adequate CC. When considering URLLC requirements and the demand for massive

access to the medium, the RRM decision-making should be done by prioritizing the CC. Nevertheless, in such tight and constrained use cases, the mmWave propagation and the complex mobility patterns of UEs boost the RRM CC. Under such conditions, maintaining an adequate trade-off between performance and complexity becomes critical [94]. In [91], it is discussed how, despite deep learning (DL) solutions being highly efficient, they increase CC if employing multiple hidden layers to yield accurate results. For this reason, the distributed and decentralized solutions using mobile edge computing (MEC) architectures, deep reinforcement learning (DRL), and federated learning (FL)-based algorithms should be considered [15, 95]. Intelligent-native RRM conception will require flexible architectures such as O-RAN to add the virtualization, softwarization, and disaggregation expected in 6G [96].

As defined in [6, 97], O-RAN is a native ML framework that uses virtualized and disaggregated elements to conduct dynamic tasks. This architecture disaggregates the BS functionalities into the Open-RAN Control Unit (O-CU), Open-RAN Distributed Unit (O-DU), and Open-RAN Radio Unit (O-RU). The logical split allows these functional units to be flexibly deployed at different network locations and hardware platforms. This disaggregated architecture could be essential in reducing the CC of the RRM solution, adding flexibility during the optimal solutions finding. The O-RAN framework enables an effective ML closed-loop workflow to dynamically conduct several optimization actions directly impacting QoS, user perception, and the E2E complexity of the envisioned systems [6].

One of the critical challenges of multicast RRM is the CC of the subgrouping techniques. As discussed above, to simplify such complexity, [33, 74] prove that the subgrouping methods achieve the best results when creating just two subgroups. The proposed solutions achieve optimal subgrouping with reduced complexity considering each user's SINR and the corresponding CQI. In [24], the authors consider the CC as a KPI in the proposed RRM algorithm. They define the CC associated with the implemented solution for multicast subgrouping based on OMA and NOMA and analyze the implication of the resource blocks (RBs) and the injection level. Regarding the use of NOMA, in [82], the authors propose several methods to reduce the implementation complexity and delay of NOMA-based transmissions.

In [27], the authors consider ML techniques for group-oriented multi-beams multicasting, achieving an adequate trade-off between performance and CC. The results offer a CC-efficient solution for optimal multicast grouping in mmWave with directional multi-beams antennas. Moreover, [60] propose a low-complexity multicast resource allocation strategy based on beamforming that allocates the resources to different UE sub-groups.

To the best of the authors' knowledge, none of the previous investigations delve into the implications of fast variations in the reception conditions of the MG members in the dynamic multicast RRM strategies over 5G MBS use cases. Such variations in the channel quality conditions imply recalculating the multicast delivery solution to cope with the specific network conditions and service requirements.

Therefore, non-optimized multicast RRM could exponentially increase the CC and associated delay of the multicast sessions. In specific 5G MBS use cases with severe variations in the users' reception conditions due to the users' mobility behaviors or the mmWave propagation, an exponential increase in CC and induced delays could not be tolerated.

Our proposal aims to fill and complement the above-described gaps (**SO-5**) by analyzing and recreating specific conditions that increase CC due to the users' mobility behaviors or the mmWave propagation and evaluating their implications. Moreover, we propose multicast access techniques solutions to cope with specific network conditions, effectively handling the trade-off between optimal network performance and CC. We propose novel low-complexity multicast RRM strategies for dynamic access technique selection and resource allocation. We present sub-optimal solutions that maximize the QoS, paying particular attention to the CC as a KPIs. To assess the effectiveness of our approach, we perform a comprehensive simulation campaign highlighting the existing interrelation between the users' velocity, the propagation frequency, and the variations in channel conditions.

We consider as essential the convergence of ML-native multicast RRM solutions and extremely virtualized and disaggregated architectures, such as O-RAN, in effectively handling the trade-off between optimal network performance and CC. We analyze the insertions of ML-based multicasting RRM solutions in the O-RAN framework under a softwarized and intelligent vision (**SO-5**).

.

Chapter 3

System Model and Problem Formulation

This chapter defines the system model and problem formulation for addressing the dynamic multicast multiuser diversity. To advance into the research proposals, analysis, and conclusions, we first describe the main mathematical notations that will be used in the manuscript, as well as constraints and recreated network conditions. After setting the research conditions, we formulate the main research problems to be addressed. For better understanding, Table A.1 summarizes the main mathematical notations used throughout the document.

3.1 General System Model

In the system model, we assume a single cell 5G NR BS that provides a multicast multimedia service by employing either the benchmark single-rate CMS strategy or the multi-rate S-OMA or S-NOMA. We assume K users requesting the same multicast multimedia service, where the sub-index k identifies each user u_k in the MG, with $\{k = 1, \ldots, K\}$. Let $\mathcal{K}$ be the set of MG users, with $K = |\mathcal{K}|$. Let us define $\{XYZ\}_k$ as the spatial information of each u_k in the simulated area. $\mathcal{XYZ}$ is the set including the spatial information of the K users. We assume that the users periodically send updates regarding their Global Positioning System (GPS)-derived location to the network. These updates can be part of regular signaling or control procedures, depending on the network architecture.

We consider a BS with an effective channel bandwidth (BW) equal to B and a total number of RBs equal to R. We assume R_M as the number of RBs dedicated to the enabled multicast session, with $R_M \leq R$. One RB is the smallest frequency resource that the BS can allocate. An RB corresponds to 12 consecutive and equally spaced subcarriers [98]. The 5G NR standard defines multiples numerologies (μ) for different SCS values, according to $\Delta f = 15 \times 2^{\mu}$, expressed in kHz (symbol duration

μ	Δf[kHz]	$\Delta t[\mu s]$	B_0[kHz]	Cyclic prefix
0	15	66.7	180	Normal
1	30	66.7	360	Normal
2	60	16.7	720	Normal, Extended
3	120	8.33	1440	Normal
4	240	4.17	2880	Normal

Table 3.1: Numerology defined for 5G NR [98].

$1/\Delta f$). B_0 is the bandwidth of an RB, being $B_0 = 12 \times \Delta f$ and also equal to B/R [98]. Table 3.1 summarizes the main numerology-associated parameters of 5G NR[98].

The set of available RBs is managed by the RRM, which must execute the fast link adaptation procedures, selecting the most appropriate MCS to deliver the multicast service. The BS carries out such selection every transmission time interval (TTI) equal to 1 ms. Every TTI, the K users send as feedback to the BS, its experienced channel quality through the uplink channel state information (CSI), specifically throughout the CQI [98]. The CQI values range from 0 to 15, where 0 means that the user is out of the BS service area, 1 is for the lowest channel quality, and 15 is for the best. We assume the K users are in the BS service area. The reported CQI_k is directly related to the SINR experienced by each u_k. The $SINR_k$ experienced by u_k must be higher than or equal to the minimum SINR ($SINR_{min}$) required to correctly decode the MCS associated with the reported CQI_k. Each CQI and MCS has associated a specific code rate and efficiency (eff) value, expressed in bps/Hz, as summarized in Table 3.2.

In [99], it is presented an expression that relates the SINR, the eff associated to a specific MCS and the target bit error rate (BER) in the following way

$$SINR_{min} = (2^{eff} - 1) \times \frac{-\log(5 \times BER)}{1.5}. \tag{3.1}$$

Fig. 3.1 presents the relationship between the SINR, expressed in decibels (dB), and the eff for the additive white gaussian noise (AWGN) [100] channel (as presented in [42, 101]), the Jakes [102] channel, and two instances of (3.1). We use (3.1) in Chapter 5 (Section 5.1) to numerically determine the $SINR_{min}$ associated with each reported CQI. When the assessment of the proposals is based on the LLS, we considered the Jakes channel model as detailed in [42] to obtain the $SINR_{min}$ associated with each CQI.

The array of CQI values collected at the BS is $CQI_{\text{MG}} = \{CQI_{k=1}, \ldots, CQI_{k=K}\}$. We define M as the number of different CQI values reported in CQI_{MG} with $M \leq 15$. We define the users-per-CQI vector $U_{\text{MG}} = \{u_{CQI=1}, \ldots, u_{CQI=15}\}$ from the CQI_{MG} array, where $u_{CQI=x}$ is the number of users in $\mathcal{K}$ that report a CQI equal to x.

CQI index	modulation	code rate ×1024	efficiency (bps/Hz)
0	out of range		
1	QPSK	78	0.1523
2	QPSK	193	0.3770
3	QPSK	449	0.8770
4	16QAM	378	1.4766
5	16QAM	490	1.9141
6	16QAM	616	2.4063
7	64QAM	466	2.7305
8	64QAM	567	3.3223
9	64QAM	666	3.9023
10	64QAM	772	4.5234
11	64QAM	873	5.1152
12	256QAM	711	5.5547
13	256QAM	797	6.2266
14	256QAM	885	6.9141
15	256QAM	948	7.4063

Table 3.2: CQI indexes s.t. Table 5.2.2.1-3 in [98].

In our proposal, we assume perfect CQI estimation. The impact of imperfect CQI estimation is out of the scope of this research. Nevertheless, in [74], the authors conclude that the imperfect CQI estimation minimally impacts the results when adopting a subgrouping approach.

During the subgrouping creation for the multi-rate techniques, we assume that the MG is split into S subgroups, with $S = 2$ according to the results presented in [74, 75]. We use the terms $G1$ and $G2$ to identify subgroups 1 and 2, respectively. $G1$ is conformed by the low channel quality users u_{k1} with $\{k1 = 1, \ldots, K_{\mathrm{G1}}\}$, identified as the set $\mathcal{K}_{\mathrm{G1}}$ with K_{G1} users. The set $\mathcal{K}_{\mathrm{G2}}$ identifies the high channel quality users u_{k2} with $\{k2 = 1, \ldots, K_{\mathrm{G2}}\}$, grouped into $G2$, with K_{G2} users. $K = K_{\mathrm{G1}} + K_{\mathrm{G2}}$, $\mathcal{K} = \mathcal{K}_{\mathrm{G1}} \cup \mathcal{K}_{\mathrm{G2}}$, and $\mathcal{K}_{\mathrm{G1}} \cap \mathcal{K}_{\mathrm{G2}} = \varnothing$. Moreover, we define $CQI_{\mathrm{MG}}^{G1} = \{CQI_{k1=1}, \ldots, CQI_{k1=K_{G1}}\}$ and $CQI_{\mathrm{MG}}^{G2} = \{CQI_{k2=1}, \ldots, CQI_{k2=K_{G2}}\}$, as the CQI values of the MG, grouped into $G1$ and $G2$, respectively, with $CQI_{\mathrm{MG}} = CQI_{\mathrm{MG}}^{G1} \cup CQI_{\mathrm{MG}}^{G2}$.

Fig. 3.2 represents a glance at the recreated system model. It illustrates that K users can dynamically change their spatial distribution in the service area (i.e., they can be all very close or far to the BS, randomly distributed, or with a group-oriented distribution). The following distribution and mobility behavior will characterize their diversity regarding reception conditions.

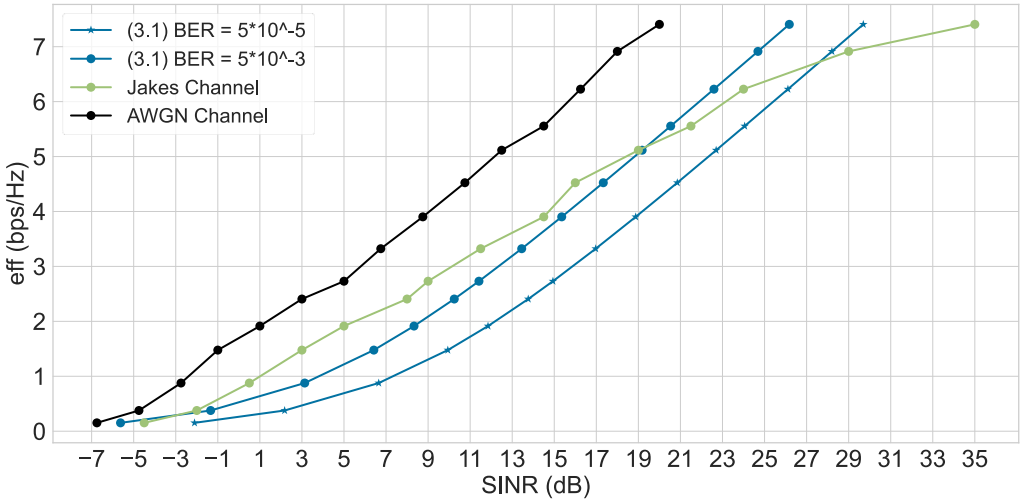


Figure 3.1: SINR and eff relationship.

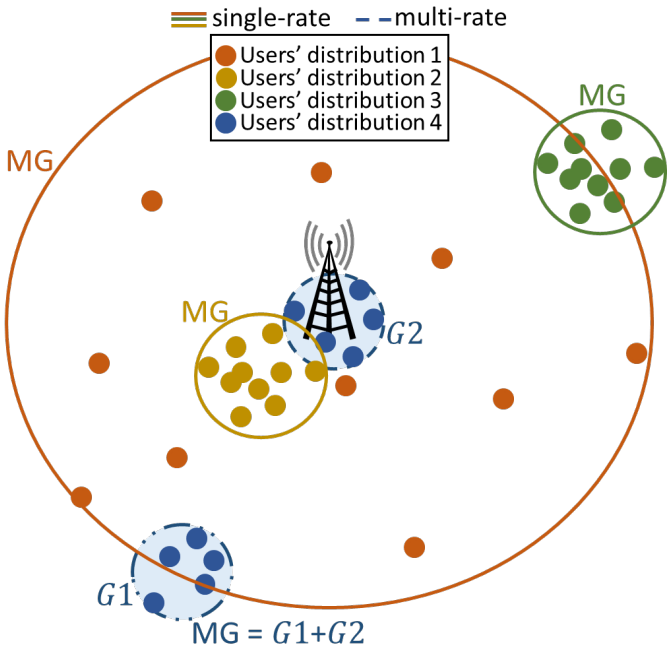


Figure 3.2: Toy system model of the recreated scenarios.

3.1.1 Resources, Requirements, and Metrics

We consider that the delivered multicast multimedia service has a requirement of minimum and maximum throughput (Th_{min} and Th_{max}), respectively. We define the effective number of allocated resources to accomplish the throughput service requirement as R^*_M, with $R^*_M \leq R_M \leq R$. As a constraint, the following resource

allocation strategy must guarantee to correctly deliver at least the Th_{min} to a predefined percent, P (e.g., 95 %, 98 %, 100 %), of the MG users. Then, the capacity (C) assigned to each u_k, expressed in bits per seconds (bps), is computed as

$$C_k = r_k \times B_0 \times eff_k, \tag{3.2}$$

where r_k is the number of RBs assigned to the u_k and eff_k is the efficiency associated to the CQI feedback of u_k. C_k must be higher than or equal to Th_{min} for each u_k in $\mathcal{K}$, to accomplish with $P = 100$ %. Specifically, the capacity of MG if CMS, S-OMA, or S-NOMA is applied, can be respectively computed as

$$C_{\mathrm{C}} = R^*_M \times B_0 \times eff_{\mathrm{min}}, \tag{3.3}$$

$$C_{\mathrm{O}} = r_{G1} \times B_0 \times eff^{G1}_{\mathrm{min}} + r_{G2} \times B_0 \times eff^{G2}_{\mathrm{min}}, \tag{3.4}$$

$$C_{\mathrm{N}} = r_{G1} \times B_0 \times eff^{\mathrm{UL,G1}}_{\mathrm{min}} + r_{G2} \times B_0 \times eff^{\mathrm{LL,G2}}_{\mathrm{min}}. \tag{3.5}$$

The eff_{min} in (3.3) is the minimum eff corresponding to the MCS assigned to the u_k that reports the lowest CQI_k (CQI_{min}) in the CQI_{MG}. Therefore, CMS assigns the same capacity C_{C} to the entire MG, based on the user u_k with the lowest channel quality. In the case of S-OMA, the eff^{G1}_{min} is the efficiency corresponding to the u_{k1} belonging to $G1$ that reports the lowest CQI (CQI^{G1}_{min}). The same applies for eff^{G2}_{min} in (3.4) regarding $G2$ with CQI^{G2}_{min}. Therefore, S-OMA assigns the same capacity $C_{\mathrm{O,G1}}$ and $C_{\mathrm{O,G2}}$ to all members in the MG grouped into $G1$ and $G2$, respectively.

In the case of S-NOMA, the $eff^{\mathrm{UL,G1}}_{\mathrm{min}}$ and $eff^{\mathrm{LL,G2}}_{\mathrm{min}}$ are the result of the SINR adaptation [77] in the LDM approach. We assume that the lowest channel quality users grouped into $G1$ are powered multiplexed into the upper layer (UL) and the highest channel quality users ($G2$) into the lower layer (LL). The SINR adaptation allows calculating from the real SINR (in dB) of each u_k ($SINR_k$) the corresponding SINR (in dB) for the UL ($SINR^{\mathrm{UL,G1}}_k$) or for the LL ($SINR^{\mathrm{LL,G2}}_k$), as defined in [77]. In other words, the SINR adaptation is the way to find from the actual $SINR_k$ of a u_k, which MCS could be demodulated under LDM conditions of reception in the UL or LL. We apply SINR adaptation based on the following expressions

$$SINR^{\mathrm{UL,G1}}_k = SINR_k + \log_{10}\left(\frac{1 + 10^{il/10}}{1 - 10^{\frac{SINR_k + il}{10}}}\right), \tag{3.6}$$

$$SINR^{\mathrm{LL,G2}}_k = SINR_k + \log_{10}(1 + 10^{il/10}). \tag{3.7}$$

We consider the injection level (IL) values between -25 and -5 at steps of 0.5 and between -5 and 0 at steps of 0.25, with l equal to 61. il belongs to the vector of IL denoted as $\mathcal{I} = \{il_i : i = 1, \ldots, l\}$ [24].

The $eff_{\min}^{\mathrm{UL,G1}}$ and $eff_{\min}^{\mathrm{LL,G2}}$ are always less than or equal to the $eff_{\min}^{\mathrm{G1}}$ and $eff_{\min}^{\mathrm{G2}}$, respectively. When the selected IL value tends to -25, the $eff_{\min}^{\mathrm{UL,G1}}$ tends to the $eff_{\min}^{\mathrm{G1}}$ and the $eff_{\min}^{\mathrm{LL,G2}}$ tends to the $eff_{\min}^{\mathrm{UL,G1}}$. The capacity of G2 ($C_{G2}^{\mathrm{LL}}$) tends to the capacity of G1 (C_{G1}^{UL}). In contrast, when the selected IL value tends to 0, the $eff_{\min}^{\mathrm{UL,G1}}$ tends to the minimum available efficiency ($eff_{\min}^{G1}$), and the $eff_{\min}^{\mathrm{LL,G2}}$ tends to the $eff_{\min}^{G2}$.

The NOMA signal based on LDM ($W(n)$) is described by [63]

$$W(n) = \sqrt{P_W \times \alpha_1} \times S_1(n) + \sqrt{P_W \times \alpha_2} \times S_2(n), \tag{3.8}$$

where P_W is the total power transmitted and $S_1(n)$, and $S_2(n)$ are the delivered signal in the UL ($G1$) and LL ($G2$), respectively. The factors α_1 and α_2 stand for the fractions of power assigned to each service with $\alpha_1 + \alpha_2 = 1$.

Equal power fractions for both layers mean that the selected IL is zero. Therefore, we can choose a more efficient MCS at S_2 at the expense of a higher impact of auto-interference from the LL over the UL. It implies that a less efficient MCS must be selected at S_1 to ensure the correct decoding. On the other hand, a lower fraction of power assigned to S_2 (i.e., when IL tends to -25) means that the LL is buried deeper down the UL. Then, we must select a more robust (i.e., less efficient in terms of capacity) MCS for S_2 to ensure the correct successive interference cancellation (SIC). From the receivers' point of view, the main characteristic of LDM is the SIC process for accessing the overlaid content [63].

To assess the performance of the benchmark single-rate CMS and the evaluated group-oriented multi-rates strategies S-OMA and S-NOMA, we consider the overall system aggregated data rate (ADR) metric [63, 74], defined as

$$ADR = \sum_{k=1}^{K} r_k \times B_0 \times eff_k, \tag{3.9}$$

The ADR is the sum of the C_k delivered to each u_k in the MG. For the sub-grouping approaches, it is the sum of the total capacity delivery of $G1$ and $G2$, respectively. This metric enables us to compare how efficiently the multicast access strategies allocate their resources, maximizing the sum of the total experienced throughput of the MG. Then, the ADR of the specific multicast access techniques CMS, S-OMA, and S-NOMA can be computed as

$$ADR_{\mathrm{C}} = K \times C_{\mathrm{C}}, \tag{3.10}$$

$$ADR_{\mathrm{O}} = K_{G1} \times C_{\mathrm{G1,O}} + K_{G2} \times C_{\mathrm{G2,O}}, \tag{3.11}$$

$$ADR_{\mathrm{N}} = K_{G1} \times C_{\mathrm{G1,N}} + K_{G2} \times C_{\mathrm{G2,N}}. \tag{3.12}$$

The drawback of the ADR metric is the lack of information about how fair the available multicast RBs are distributed among users regarding their CQI diversity [23]. The proportional fairness (PF) and minimum dissatisfaction index (MDI) are two well-known metrics to assess how fairly the available resources are allocated among the MG members [23]. In [103], the authors demonstrate that one unique fair allocation exists, and it is obtained by maximizing the sum of the logarithm of the data rate of each u_k. The MDI is computed as the sum of the ratio between the data rate achieved by each u_k and the maximum possible data rate value achieved if all RBs (R_M^*) are assigned to the u_k [63]. We define the PF and MDI as follow

$$PF = \sum_{k=1}^{K} \ln\left(r_k \times B_0 \times eff_k\right), \tag{3.13}$$

$$MDI = \sum_{k=1}^{K} \frac{r_k \times B_0 \times eff_k}{R_M^* \times B_0 \times eff_k^*}, \tag{3.14}$$

where eff_k^* is the maximum efficiency that each u_k can receive according to their reported CQI_k.

We use the metric Δ_R to measure how efficiently the multicast strategies use the available resources to accomplish the service constraint. Δ_R is equal to the number of RBs dedicated to enabled multicast session minus the effective number of allocated RBs, $\Delta_R = R_M - R_M^*$.

3.1.2 Channel Modeling

For the channel modeling and link level computation, we consider an urban scenario following the 3GPP urban microcell (UMi) Street Canyon path loss model defined in [104]. As the result of the path loss computation, the total received power at the end-user $P_{rx,pl}$ is computed as

$$P_{rx,pl} = P_{tx} G_{tx} G_{rx} PL^{-1}(d), \tag{3.15}$$

where G_{tx} and G_{rx} are the transmit (tx) and receive (rx) antenna gain, respectively. P_{tx} is the transmission power. $PL(d)$ is the path loss, where d is the 3D distance between tx and rx.

The final total received power at the u_k and the r-th sub-channel ($P_{rx,k,r}$) is determined as

$$P_{rx,k,r}|dB = (P_{rx,pl,k,r} + T_k + S_k + H_{k,r})|dB, \tag{3.16}$$

where T_k, S_k, and $H_{k,r}$ are the penetration attenuation component, shadowing fading, and fast fading, respectively, expressed in dB. In the penetration attenuation, we include the human blockage with an attenuation of 15 dB.

The link channel quality is evaluated in terms of the SINR measured over each sub-channel r corresponding to the dedicated RBs to the users:

$$SINR_{k,r} = \frac{P_{rx,k,r}}{I + (FN_0B_r)}, \tag{3.17}$$

where F, N_0, and B_r are the configured noise figure, the noise spectral density (with a default value of -174 dBm/Hz), and the bandwidth of the sub-channel. I is the sum of interference components. The SINR computation is the first step of the link to the system's mapping process throughout the CQI estimation and the block error rate (BLER) computation.

In our approach, we use the enhanced version of the traditional Exponential Effective SINR Metric (EESM) [105] function to map the $SINR_{k,r}$ to an effective SINR value ($SINR_{eff}$). This method is presented in [101] as

$$EESM_k = -\alpha \times \ln\left(\frac{1}{r_k}\sum_{r=1}^{r_k}\exp\left(\frac{-SINR_{k,r}}{\beta}\right)\right), \tag{3.18}$$

where r_k is the number of available sub-channels (number of RBs allocated to u_k). α and β are the adjustment factors to each specific MCS and the corresponding CQI value.

To find the optimal α and β, we follow an iterative process to minimize the difference between the real BLER ($BLER_{real}$) achieved with the $SINR_{k,r}$ array of the real channel and the equivalent BLER ($BLER_{eff}$) in the AWGN channel achieved with $SINR_{eff}$. Another approach could be to minimize the difference between the equivalent SINR in the AWGN channel ($SINR^*_{AWGN}$) that ensures the same $BLER_{real}$ and the computed $SINR_{eff}$ through (3.18). By minimizing such metrics, we ensure that the computed $SINR_{eff}$ mimics the real channel behavior as best as possible. Then, we could estimate the correct CQI using the curves of BLER versus SINR for the AWGN channel, with a CQI estimation error rate as low as possible. The target BLER used for selecting the CQI can be set during the initialization, typically defined as 0.1. The details about the finding process of α and β are presented in Appendix B. Additional information can be found in [42, 100, 101, 106, 107].

3.1.3 Antenna Modeling

The single cell BS is modeled as a three-sectorial antenna. We implement a cone antenna model for each sector, representing the radiation pattern as a conical zone. The angle θ coincides with the antenna array's potential half-power beamwidth (HPBW). It is determined in [108] by

$$\theta = 2|\theta_m - \theta_{3db}|, \tag{3.19}$$

where θ_{3db} is the angle at which the radiated power value is 3 dB below the maximum, and θ_m is the location of the maximum array. $\theta_m = arccos(\varphi/\pi)$, where φ is the phase excitation difference affecting the physical orientation of the array. We consider $\theta_m = \pi/2$ for $\varphi = 0$.

The antenna gain is computed as

$$G_{tx} = \frac{1}{\theta_{3db}^{+} - \theta_{3db}^{-}} \int_{\theta_{3db}^{-}}^{\theta_{3db}^{+}} \frac{sin(N\pi cos(\theta)/2)}{sin(\pi cos(\theta)/2)} \, d\theta \ , \tag{3.20}$$

where N is the number of antenna elements and $\theta_{3db}^{\pm} = arccos(\varphi \pm 2.782/N\pi)$ [108]. We consider a fixed value for G_{rx}.

We assume that L beams can be steered in different directions, splitting the total power P_{max} equally among the beams. The HPBW of the beams depend on the number of antenna elements and are limited by a lower bound [27]. Our approach considers fixed pre-computed beams following the above-defined equations. Let $\mathcal{L}$ be the set of available pre-computed beams, with $L = |\mathcal{L}|$ and l is the index to identify each beam. In the baseline case of $L = 1$, the antenna model follows the simplified antenna pattern given in ITU-R M.2135 [104].

3.2 Problem Formulation

As we analyzed in Subsection 1.1.3, addressing the trade-offs between multicasting gain and multiuser diversity and between optimal network performance and CC are the baselines of this research problem. Then, we present the problem formulation with a twofold approach. First, we characterize the multicast access techniques, their variables interrelation, and conditions for an effective coexistence. Second, we formulate the dynamic multicasting complexity and the implications of the channel quality variation.

3.2.1 Problem Space Formulation for Characterizing the Multicast Access Techniques

Addressing the coexistence of the traditional CMS with dynamic multi-rate multicasting implies identifying the conditions for dynamically selecting the best multicast strategy for specific network characteristics. Such conditions can be defined by the interrelation among the variables that shape the performance of CMS, S-OMA, and S-NOMA. To evaluate the relative ADR performance among the evaluated multicast access techniques and the benchmark CMS, we consider the following metrics [39, 41]

$$\Delta_C^O = 1 - \frac{ADR_{\mathrm{C}}}{ADR_{\mathrm{O}}}, \tag{3.21}$$

Equation	Outcome (≤ 0)	Outcome (> 0)
Δ_C^O	CMS	S-OMA
Δ_C^N	CMS	S-NOMA
Δ_O^N	S-OMA	S-NOMA

Table 3.3: Meaning of the possible outcomes with (3.21), (3.22), and (3.23).

$$\Delta_C^N = 1 - \frac{ADR_{\mathrm{C}}}{ADR_{\mathrm{N}}}, \tag{3.22}$$

$$\Delta_O^N = 1 - \frac{ADR_{\mathrm{O}}}{ADR_{\mathrm{N}}}. \tag{3.23}$$

If the values of (3.21) and (3.22) are lower than or equal to zero, it means that the single-rate CMS technique is the best multicast access technique regarding the system ADR metric for the specific evaluated network conditions; in other cases the multi-rate approaches S-OMA and S-NOMA are the best options. In (3.23), if the value is lower than or equal to zero, the S-OMA technique exceeds S-OMA; in other cases, S-NOMA can achieve a higher ADR.

A zero value means that the two evaluated strategies have the same performance. However, in these cases, we choose the simplest one regarding CC. If we define C_C, C_O and C_N as the CC of the RRM process associated to the multicast access techniques CMS, S-OMA and S-NOMA, respectively, we can define that $C_C < C_O < C_N$ as presented in [41]. The details about the CC associated with the multicast access techniques will be presented in the following Subsection 3.2.2. Table 3.3 summarizes the meaning of the possible outcomes with (3.21), (3.22), and (3.23).

To gain clarity into the variables that characterize the behavior of (3.21), (3.22), and (3.23), we mathematically transform these equations. The resulting expressions are

$$\Delta_C^O = 1 - \frac{(\kappa \times \eta)^{-1} \times eff_{\min}^{G1}}{eff_{\min}^{G1} + eff_{\min}^{G2}(1 + (\kappa \times \eta)^{-1} - \kappa^{-1} - \eta^{-1})}, \tag{3.24}$$

$$\Delta_C^N = 1 - \frac{\kappa^{-1} \times eff_{\min}^{G1}}{eff_{\min}^{\mathrm{UL,G1}} + eff_{\min}^{\mathrm{LL,G2}}(\kappa^{-1} - \eta^{-1})}, \tag{3.25}$$

$$\Delta_O^N = 1 - \frac{eff_{\min}^{G1} + eff_{\min}^{G2}(1 + (\kappa \times \eta)^{-1} - \kappa^{-1} - \eta^{-1})}{eff_{\min}^{\mathrm{UL,G1}} + eff_{\min}^{\mathrm{LL,G2}}(\kappa^{-1} - \eta^{-1})}, \tag{3.26}$$

where $\kappa = K_{G1}/K$ and $\eta = r_{G1}/R_M$. The possibility of expressing (3.24), (3.25), and (3.26) in terms of κ and η proves that the performance among the evaluated

Equation	Number of Combinations
Δ_C^O	$\frac{n(n-1)}{2} \times R_M \times K = 1200000$
Δ_C^N	$\frac{n(n-1)}{2} \times l \times K = 732000$
Δ_O^N	$\frac{n(n-1)}{2} \times R_M \times l \times K = 73200000$
$n = 15$, $R_M = 100$, $l = 61$, $K = 100$	

Table 3.4: Maximum number of combinations with (3.24), (3.25), and (3.26).

multicast techniques is independent of the total available resources (R_M) (i.e., the system bandwidth) and the total number of multicast users K. It depends on the corresponding ratios between $G1$ and $G2$. In terms of η, the equations only depend on the percent of RBs assigned to $G1$ (r_{G1}) regarding R_M (i.e., $R_M = r_{G1} + r_{G2}$). In the case of κ, it proves that the performance is also independent of the specific number of users in the MG; i.e., it only depends on the percent of users grouped into $G1$ (K_{G1}) regarding the total MG members (with $K = K_{G1} + K_{G2}$). The dependencies of η and κ show that the identified performance thresholds can be generalized to any available RBs dedicated to the multicast multimedia service (R_M) and MG members (K).

The problem space defined by the goal of finding the condition for adequate dynamic multicasting based on (3.24), (3.25), and (3.26) is multi-dimensional. It means that the users' reception conditions determine the CQI value of the users (i.e., the corresponding MCS and eff associated) and how the users are grouped into $G1$ and $G2$. Moreover, the multicast multimedia service constraints, such as Th_{min}, determine the optimal RBs allocation and IL selection for the S-OMA approach and S-NOMA.

The multiple variables that characterize (3.24), (3.25), and (3.26) boost the complexity of finding the conditions for effective dynamic multicasting and handling the MG channel quality diversity. Table 3.4 presents the maximum number of possible combinations with (3.21), (3.22), and (3.23) for $R_M = 100$, $K = 100$, and $l = 61$. The partial sum formula $(n(n+1))/2$, with n equal to 15, represents all the possible combinations of CQI_{min}^{G1} and CQI_{min}^{G2} (with $CQI_{min}^{G1} \leq CQI_{min}^{G2}$), when the number of the possible combinations of r_{G1} and r_{G2} is equal to R_M.

This high number of combinations evidences the broad problem space and complexity of finding the multicasting strategy that maximizes the system ADR with an optimal balance between the multicasting gain and multiuser diversity. In such a context, we aim to identify the specific network conditions or inflection points when each multicast access technique starts to outperform the other. It allows the definition of the interrelation among the variables that shape the performance of the benchmark and the evaluated multi-rate solutions in terms of the users' reception conditions, multimedia service constraints, and network parameters. For finding such inflection points, we define the following cost function (CF) as an optimiza-

tion problem

$$\Theta = \begin{cases} arg\min_{x_i} |\Delta_C^O|, & \text{if S-OMA vs. CMS} \\ arg\min_{x_i} |\Delta_C^N|, & \text{if S-NOMA vs. CMS} \\ arg\min_{x_i} |\Delta_O^N|, & \text{if S-NOMA vs. S-OMA} \end{cases} \tag{3.27}$$

where x_i is the argument that minimizes (as close to zero as possible) the module of (3.24), (3.25), and (3.26). x_i is composed of the variables that shape the performance of the evaluated multicast access techniques under evaluation.

3.2.2 Dynamic Multicasting Complexity

An effective 5G MBS RRM strategy aims to maximize the QoS of the MG members by efficiently allocating the available resources of the multicast sessions. The metrics to evaluate how efficiently the resources are allocated could be the above-mentioned (Subsection 3.1.1) or a weighted combination of them and additional complex metrics. Nevertheless, the envisaged ultra-dense HetNet with differentiated services and tight QoS requirements is a dynamic scenario, which makes managing and exploiting network resources even more complex [38].

In such a context, the necessity of a constant recalculation and the intrinsic convergence time of the best multicast access techniques and resource allocation (to maximize the CF) strategies could become a critical factor from the CC point of view. It increases the control plane delay, the total E2E delay, and induces a critical extra latency in the communication. Therefore, CC must be considered a KPI for our proposed multicast RRM solutions.

The dynamic selection of the best multicast access technique and resource allocation, from the CC perspective, can be divided into three main phases: (a) the selection of the best multicast access technique (i.e., MCS, S-OMA, or S-NOMA for our approach); (b) the MG subgrouping for the multi-rate solutions; (c) the selection of the optimal resource allocation that maximizes a specific CF.

Selecting the best multicast access technique through an exhaustive search strategy (ESS) implies evaluating the N available multicast access techniques (i.e., $N = 3$ in our analysis). This ESS-Multicast Access Technique Selection (ESS-MAT) has a CC equal to the sum of the CC of the N techniques, as presented below.

The CC of CMS (CC_C), when the Th_{min}, Th_{max} and P are considered as the constraints of the service, can be defined as $\mathcal{O}(M*R_M)$. M is the number of different CQI values reported by the MG in CQI_{MG} ($M \leq 15$), and R_M represents the available RBs to iterate over subject to the throughput constraints. If P is assumed equal to 100 %, the CMS algorithm only considers the CQI value corresponding to the lowest channel quality user (i.e., the CQI_{min} of the MG). Therefore, the M term does not impact the CC. Moreover, if Th_{max} is assumed as infinity, R_M^* is equal to R_M, and R_M has not impact in the complexity. For these specific conditions, the

CC of CMS can be assumed as insignificant. In the following, we will assume P_{users} as 100 % and Th_{max} as ∞ to simplify the analysis.

The subgrouping process for the multi-rate solutions has an essential contribution to the CC. In [74], the authors define that the subgrouping process solved through an ESS has a CC equal to $\mathcal{O}(M^S)$, where S is the number of considered subgroups. Specifically, we assume $S = 2$ with $\mathcal{O}(M^2)$. Following the assumption of P as 100 %, the CC can be redefined as $\mathcal{O}(M)$ because the CQI_{min}^{G1} is fixed to the CQI_{min} of the MG. Then, the algorithm must only iterate over the remaining $M-1$ combinations. The multi-rate approaches for each subgrouping alternative must evaluate the best resource allocation that maximizes the specific CF, returning the best combination overall. For S-OMA, the CC (CC_O), including the subgrouping process (for ESS), can be defined as $\mathcal{O}(M^S * R_M)$ and redefined after the assumptions as $\mathcal{O}(M * R_M)$. In the case of S-NOMA, the CC (CC_N), including the subgrouping process (for ESS), can be defined as $\mathcal{O}(M^S * R_M * l)$, and redefined after the assumptions as $\mathcal{O}(M*l)$. Regarding the assumptions, S-NOMA uses all the available R_M; therefore, this factor is not considered in the CC_N.

We can summarize the analysis by defining the CC of the dynamic selection of the best multicast access technique and resource allocation based on an ESS ($CC_{ESS-MAT}$) equal to $CC_C + CC_O + CC_N$:

$$CC_{ESS-MAT} = \mathcal{O}(M * R_M) + \mathcal{O}(M * l). \tag{3.28}$$

According to the above-mentioned assumptions, the CC_C is considered insignificant. The inefficiency of the ESS-MAT approach in terms of CC reaches its maximum expression each time CMS is selected as the best multicast access technique. It happens because, even when the CC of CMS is insignificant, the algorithm has to execute the other two mult-rate strategies, as reflected in (3.28). The major drawback of ESS-MAT is that $CC_{ESS-MAT}$ increases linearly with the number of M, R_M, and l, and is independent of the MG CQI distribution.

To assess the effectiveness of our proposals and the benchmark algorithms in terms of CC, we measure the central processing unit (CPU) execution time as Et, in seconds, of the evaluated algorithms using a predefined function of the Python library Time [109, 110]. The Et metric only measures the CPU's time executing the code without including the time spent waiting for input/output resources or sleeping time.

3.2.3 Implications of the Channel Quality Variations

The complexity analysis cannot be just faced from the perspective of the intrinsic CC associated with the dynamic multicast access technique selection and resource allocation process, as defined in the above Subsection. How often the RRM must recalculate these solutions can directly affect the network performance, increasing the control plane delay and overhead and affecting delay-sensitive traffic.

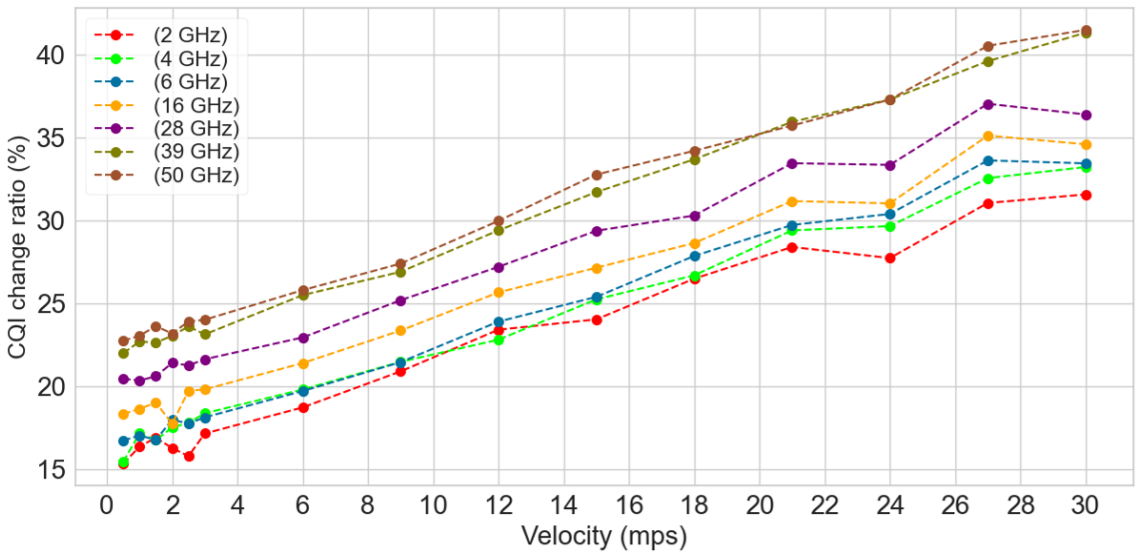


Figure 3.3: CQI change ratio regarding the propagation frequency and users' velocity.

As the frequency increases, the large and small-scale fading increases, with higher path loss attenuation, penetration loss, shadowing, and fast fading attenuation [111–113]. If this frequency dependence is essential at the traditional μWave band (below 6 GHz), it becomes more critical at mmWave [111]. Another factor directly correlated with the dynamics in the channel conditions is the mobility behavior of the users, and the dimensions of its impact are also highly frequency dependent.

Let us define the metric CQI changing ratio (CCR) as the percent of users in the MG that changes its reported CQI from the instant $t-1$ to the instant t. Fig. 3.3 shows how the CCR is affected by the increment in the propagation frequency and the users' velocity. In the Figure, we can see how the CCR increases linearly with the speed and the frequency. The results reveal an average CQI variation increase of 2.5 % as the evaluated frequencies go higher. Moreover, the speed increase adds an average extra 5.5 % of CQI variation. In such conditions, the necessity of constant recalculation and the intrinsic convergence time finding the best RRM strategies could become a critical factor from the complexity point of view. It increases the control plane and E2E delay, which induces a critical extra latency in communication. These simulations are carried out through our homemade LLS [42], for $K = 50$, 60 seconds of simulation with 100 ms of the resolution, and 20 simulation runs. Chapter 6 presents additional details about the simulation setup.

As the worst case, in terms of CC, let us assume a dynamic multicast access technique solution based on ESS-MAT that executes the algorithm at each simulation time-step with a long Et and independently of the user's CQI variation. For these ESS-MAT solutions, the mean Et over 60 seconds of simulation will remain almost

constant (s.t. the intrinsic CC of ESS-MAT) independent of the users' velocity and frequency. This optimal solution can be considered the superior bound in terms of CC for our specific problem. Let us consider a basic trigger that enables to run the ESS-MAT (basic trigger ESS-MAT (btESS-MAT)) algorithm only if more than 20 % of the MG members change their reported CQI from the time-step $t-1$ to t. This trigger will reduce the mean Et over the 60 seconds of simulation. This trigger solution based on the 20 % can imply overrunning the algorithm where the CQI changes are not in the low-channel quality users or letting users without service because its CQI change does not represent the 20 %. To address this challenge, we propose in Chapter 6 an MG-oriented trigger solution that reduces the mean Et and is less affected by users' velocity and propagation at high mmWave frequencies.

Chapter 4

Multicast Access Techniques Characterization

In this chapter, we identify the conditions and variables interrelation for an effective coexistence of CMS, S-OMA, and S-NOMA. We realize the analysis of the variables that characterize the performance of the evaluated multicast access techniques. The proposed solutions, analysis, and conclusions respond to (**SO-3**).

The main contributions can be summarized as follows: (*i*) We handle the multicasting gain and multiuser diversity trade-off throughout CMS and multi-rate subgrouping strategies for dynamic RRM over 5G MBS use cases; (*ii*) We define the interrelations among the variables that shape the performance of the evaluated multicast access techniques and their dynamic behavior in terms of the user's reception conditions, multimedia service constraints, and network parameters; (*iii*) We provide conditions for an effective dynamic selection of the multicast strategy that better suits the specific network characteristics; (*iv*) We assess the effectiveness of our proposal under a link-level simulated 5G MBS use case operating in mmWave, evaluating a wide range of network conditions.

4.1 Proposed Solution

Aimed to characterize and identify the conditions and variables interrelation for an effective coexistence of CMS, S-OMA, and S-NOMA, we develop three main steps:

- Design of a resource allocation algorithm for the benchmark traditional CMS, and S-OMA, and S-NOMA. The goal is to maximize the system ADR (ADR_{C}^{max}, ADR_{O}^{max}, and ADR_{N}^{max}) for any user distribution and multimedia service requirements.

- Design of a particle swarm optimization (PSO) based algorithm to identify the performance inflection points among the multicast access techniques, s.t.

Algorithm 1: CMS

Input: $Th_{min}, Th_{max}, P, CQI_{MG}, R_M, B_0$
Output: $ADR_{\mathrm{C}}^{max}, eff_{min}, R_M^*$
1: From CQI_{MG} determine CQI_{min} and its corresponding eff_{min} from Table 5.2.2.1-3 in [98].
2: Compute R_{min} and R_{max} from Th_{min} and Th_{max}:
$R_{min} = int(Th_{min}/(eff_{min} \times B_0))$
$R_{max} = int(Th_{max}/(eff_{min} \times B_0))$
4: Compute R_M^*:

$$\mathrm{R}_M^* = \begin{cases} 0, & \text{if } R_M \leq R_{min} \\ R_{max}, & \text{if } R_M \geq R_{max} \\ R_M, & \text{if } R_{min} \leq R_M \leq R_{max} \end{cases}$$

5: Compute the system ADR with (3.14)
if $R_M^* == 0$ **then**
  return: $ADR_{\mathrm{C}}^{max} = 0$
  Not available resources to accomplish with Th_{min}
else
  return: $ADR_{\mathrm{C}}^{max}, R_M^*, eff_{min}$
end

(3.27) and the above-defined problem space.

- Iteration process over the designed algorithms, and the whole range of variables that characterize the problem space defined by 3.24), (3.25), and (3.26. Processing of all results to identify the conditions and variables interrelation for an effective coexistence and dynamic selection of CMS, S-OMA, and S-NOMA.

4.1.1 Resource Allocation Algorithms

In the first step, we design a resource allocation algorithm for the traditional CMS, S-OMA, and S-NOMA. These algorithms aim to find the configuration that maximizes the metric ADR regarding the K users' reception conditions and for any multimedia service constraints (i.e., Th_{min}, Th_{max}, and P).

The RRM algorithm for the benchmark single-rate CMS strategy has as inputs the multimedia service constraints Th_{min}, Th_{max}, P and the MG CQI distribution CQI_{MG}. The pseudo-code for the proposed CMS solution is presented in Algorithm 1. As shown in the algorithm, we determine in step four if, regarding the service constraints and users CQI distribution, the multimedia service Th_{min} can be accomplished with the available RBs. If possible, the final step of the algorithm is computing the ADR_{C}^{max} with (3.14). The selected R_M^* (with $R_M^* \leq R_M$) is subject to $Th_{min} \leq C_C \leq Th_{max}$, allocating the maximum possible resources to maximize

Algorithm 2: S-OMA

Input: $Th_{min}, Th_{max}, P, CQI_{MG}, R_M, B_0$

Output: $ADR_{\mathrm{O}}^{max}, eff_{min}^{G1}, eff_{min}^{G2} R_M^*, r_{G1}, r_{G2}$

1: From CQI_{MG} apply the subgrouping process to determine CQI_{MG}^{G1} and CQI_{MG}^{G2}

2: From CQI_{MG}^{G1} and CQI_{MG}^{G2} determine CQI_{min}^{G1}, CQI_{min}^{G2} and its corresponding eff_{min}^{G1} and eff_{min}^{G2} from Table 5.2.2.1-3 in [98].

3: Compute R_{min} and R_{max} from Th_{min} and Th_{max}:

$R_{min} = int(Th_{min}/(eff_{min} \times B_0))$

$R_{max} = int(Th_{max}/(eff_{min} \times B_0))$

4: Compute $r_{G1,min}$, $r_{G2,max}, r^*_{G2,max}$:

$r_{G1,min} = int(Th_{min}/(eff_{min}^{G1} \times B_0))$

$r_{G2,max} = R_M - r_{G1,min}$

$r^*_{G2,max} = int(Th_{max}/(eff_{min}^{G2} \times B_0))$

5: Compute $C_{O,G1}$ with $r_{G1,min}$ and $C_{O,G2}$ with $r_{G2,max}$ s.t. (3.4)

6: Compute r_{G1}, r_{G2}:

$$\mathrm{r}^*_{G1} = \begin{cases} 0, & \text{if } R_M \leq r_{G1,min} \\ 0, & \text{if } C_{O,G1} \geq C_{O,G2} \\ r_{G1,min}, & otherwise \end{cases}$$

$$\mathrm{r}_{G2} = \begin{cases} 0, & \text{if } r^*_{G1} = 0 \\ r_{G2,max}, & \text{if } r_{G2,min} \leq r^*_{G2,max} \\ r^*_{G2,max}, & otherwise \end{cases}$$

then: $r^*_{G1} \leq r_{G1} \leq R_M - r_{G2}$**, s.t.** $C_{O,G1} < C_{O,G2}$

$R^*_M = r_{G1} + r_{G2}$

7: Compute the system ADR with (3.11)

if $R^*_M == 0$ **then**

  return: $ADR_{\mathrm{C}}^{max} = 0$

  Not available resources to accomplish with Th_{min}

else

  return: $ADR_{\mathrm{O}}^{max}, eff_{min}^{G1}, eff_{min}^{G2}, R^*_M, r_{G1}, r_{G2}$

end

the system ADR as the CF.

For the multi-rate strategy S-OMA, the RRM algorithm has the same inputs as CMS. The pseudo-code for the proposed S-OMA solution is presented in Algorithm 2. The main particularity of this algorithm is the subgrouping process applied in step one. The K users are divided into $G1$ (with K_{G1} users) and $G2$ (with K_{G2} users), regarding CQI_{MG}.

One of the main challenges of the group-oriented strategies is splitting the users into the S subgroups. Therefore, the following proposed algorithm was designed based on a low-complexity procedure. First, we determine the CQI_{min} and CQI_{max} in CQI_{MG} and we assume that a u_k belongs to the $G1$ if $(CQI_k - CQI_{min}) \leq (CQI_{max} - CQI_k)$, otherwise, it belongs to $G2$. This simple mechanism can be applied because we assume S equals two.

In the S-OMA approach, the resources R_M are split into $G1$ and $G2$, as r_{G1} and r_{G1}, with $r_{G1} + r_{G2} = R_M^* \leq R_M$. The eff_{min}^{G1} and eff_{min}^{G2} are the efficiency value of the MCS assigned to $G1$ and $G2$, according to the lowest channel quality user in $G1$ (CQI_{min}^{G1}) and $G2$ (CQI_{min}^{G2}), respectively. Let's highlight that $eff_{min} = eff_{min}^{G1}$ as well as $CQI_{min} = CQI_{min}^{G1}$ because the lowest channel quality user in the $G1$ is the lowest channel quality user in the MG. The final step of the algorithm is computing the ADR_{O}^{max} with (3.11). The selected r_{G1} and r_{G2} are subject to $Th_{min} \leq C_{G1,O} \leq C_{G2,O} \leq Th_{max}$ allocating the maximum possible resources to maximize the system ADR.

In the case of S-NOMA, the RRM algorithm has as an extra input the $\mathcal{I}$ to perform the SINR adaptation process. The pseudo-code for the proposed S-NOMA solution is presented in Algorithm 3. The algorithm starts with the subgrouping operation and the creation of the CQI_{MG}^{G1} and CQI_{MG}^{G2} following the same procedure analyzed for Algorithm 2. Then, the SINR adaptation process is applied in step three, iterating over the whole range of $\mathcal{I}$ to find the optimal IL value and the corresponding $eff_{min}^{UL,G1}$, $eff_{min}^{LL,G2}$. As we can see in steps four and five, the LDM approach can use all the available RBs in both layers, being a more spectrum efficiency-friendly technology. The final step of the algorithm is computing the ADR_{N}^{max} with (3.12). The selected R_M^* and IL are subject to $Th_{min} \leq C_{G1,N} \leq C_{G2,N} \leq Th_{max}$ allocating the maximum possible resources to maximize the system ADR.

The three designed algorithms allow maximizing the system ADR for any user distribution, network parameters, and multimedia service requirements.

4.1.2 Particle Swarm Optimization Algorithm

The PSO is an evolutionary technique that initializes the system with a population of random solutions called particles. These particles emulate the behavior of the animals' societies with no leader in their group or swarm [114, 115]. Algorithm 4 describes the following iterative process according to the procedure for implementing the global version of PSO presented in [115]. The algorithm's goal is the outperformance conditions identification (OCI) among the multicast access

Algorithm 3: S-NOMA

Input: $Th_{min}, Th_{max}, P, CQI_{MG}, R_M, \mathcal{I}, B_0$
Output: $ADR_{\mathrm{N}}^{max}, eff_{min}^{UL,G1}, eff_{min}^{LL,G2} R_M^*, il_i$
1: From CQI_{MG} apply the subgrouping process to determine CQI_{MG}^{G1} and CQI_{MG}^{G2}
2: From CQI_{MG}^{G1} and CQI_{MG}^{G2} determine CQI_{min}^{G1}, CQI_{min}^{G2} and its corresponding eff_{min}^{G1} and eff_{min}^{G2} from Table 5.2.2.1-3 in [98].
3: Apply SINR adaptation [77] iterating over $\mathcal{I}$ to find:
$eff_{min}^{UL,G1}$ and $eff_{min}^{LL,G2}$
s.t. $Th_{min} \leq C_{N,G1} \leq C_{N,G2}$, with (3.5) and R_M
if *foreach* $\mathcal{I}$: $C_{N,G1} \leq Th_{min}$ then
  return: $ADR_{\mathrm{C}}^{max} = 0$
  Not available resources to accomplish with Th_{min}
else if $Th_{min} \leq C_{N,G1} \leq C_{N,G2}$ then
  continue
end
4: Computed R_{max} from Th_{min} and Th_{max}:
$R_{min} = int(Th_{min}/(eff_{min}^{UL,G1} \times B_0))$
$R_{max} = int(Th_{max}/(eff_{min}^{UL,G1} \times B_0))$
4: Compute R_M^*:

$$R_M^* = \begin{cases} R_{max}, & \text{if } R_M \geq R_{max} \\ R_M, & \text{if } R_{min} \leq R_M \leq R_{max} \end{cases}$$

5: Compute the system ADR with (3.12)
return: $ADR_{\mathrm{N}}^{max}, eff_{min}^{UL,G1}, eff_{min}^{LL,G2}, R_M^*, il_i$
end

techniques regarding the CQI diversity, the network conditions, and the service constraints.

For the PSO conseption, we assume a population $\wp$ of 100 particles x_i (with $\{i = 1, 2, \ldots, Pop = 100\}$). Each particle x_i in the swarm is composed of four variables: $x_i = \{G1_{min}^{(i,j=1)}, G1_{max}^{(i,j=2)}, G2_{min}^{(i,j=3)}, G2_{max}^{(i,j=4)}\}$. These variables are defined in terms of CQI following the rules: $1 \leq G1_{min}^{(i,j=1)} \leq 14$, $G1_{min}^{(i,j=1)} \leq G1_{max}^{(i,j=2)} \leq 14$, $G1_{max}^{(i,j=2)} + 1 \leq G2_{min}^{(i,j=3)} \leq 15$, and $G2_{min}^{(i,j=3)} \leq G2_{max}^{(i,j=4)} \leq 15$. Each particle x_i defines the CQI boundaries of $G1$ and $G2$.

In the algorithm, each particle keeps track of its coordinates, best personal experience (p_{best}), and the global best experience (g_{best}) in the problem space defined by the (3.27). The p_{best} and g_{best} are the cognitive and social components of the algorithm, respectively [115]. At each time step, the particles change their velocity and position toward their p_{best} and g_{best} location. The velocity ($v_{ij}(t+1)$) and position ($x_{ij}(t+1)$) of the j variables of each particle x_i are based on the procedure defined

Algorithm 4: OCI-PSO

Input: $\wp$, Pop, t_{max}, K, Th_{min}, Th_{max}, P, CQI_{MG}, R_M, $\mathcal{I}$, B_0

Output: $x_{ij} = g_j(t = 50)$ and the corresponding best multicast resource allocation strategy

1: Initialize the population $\wp$ of 100 particles x_i (with $\{i = 1, 2, \ldots, Pop = 100\}$) and create their corresponding $CQI_{MG,i}$ using the LLS.

2: For each $CQI_{MG,i}$, apply the corresponding algorithms (Algorithms (1), (2), (3)) of the two multicast access techniques under evaluation.

3: Evaluate (3.27) (the CF) for the two multicast access techniques under evaluation.

4: Compare the CF evaluation of the particle with particle's p_{best}. If current value is better than p_{best}, set p_{best} value equal to the current value, and the p_{best} location equal to the current location in the j variables.

5: Compare the CF evaluation with the overall previous best (g_{best}). If the current value is better than g_{best}, then reset g_{best} to the current array index and value.

6: Change the velocity and position of the particles according to (4.1) and (4.2), respectively.

7: Loop to step 2 (with $t_{max} = 50$) until the algorithm find a $x_{ij} = g_j$ that satisfy CF.

end

in [115] and computed as follow

$$v_{ij}(t+1) = w \times v_{ij}(t) + r_1 \times c_1 \times (p_{ij}(t) - x_{ij}(t)) + r_2 \times c_2 \times (g_j(t) - x_{ij}(t)) \tag{4.1}$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1). \tag{4.2}$$

$p_{ij}(t)$ and $g_j(t)$ are the p_{best}, and g_{best} of each variable. We assume c_1 and c_2 equal to 1.49618 and w equal to 0.72984, as recommended in [115]; these are the acceleration and inertia constants, respectively. The values of r_1 and r_2 are randomly set between 0 and 1 [115].

In step one, the initial position of each particle ($x_{ij}(t = 0)$) is defined as a random value following the rules previously described. Moreover, the velocity of each particle at the initial stage of the algorithm ($v_{ij}(t = 0)$) is assumed equal to 0. After initializing the variables, we must run a link-level simulator (LLS) to obtain the users' CQI feedback, s.t. the rules for the CQI boundaries, and create $CQI_{MG,i}$

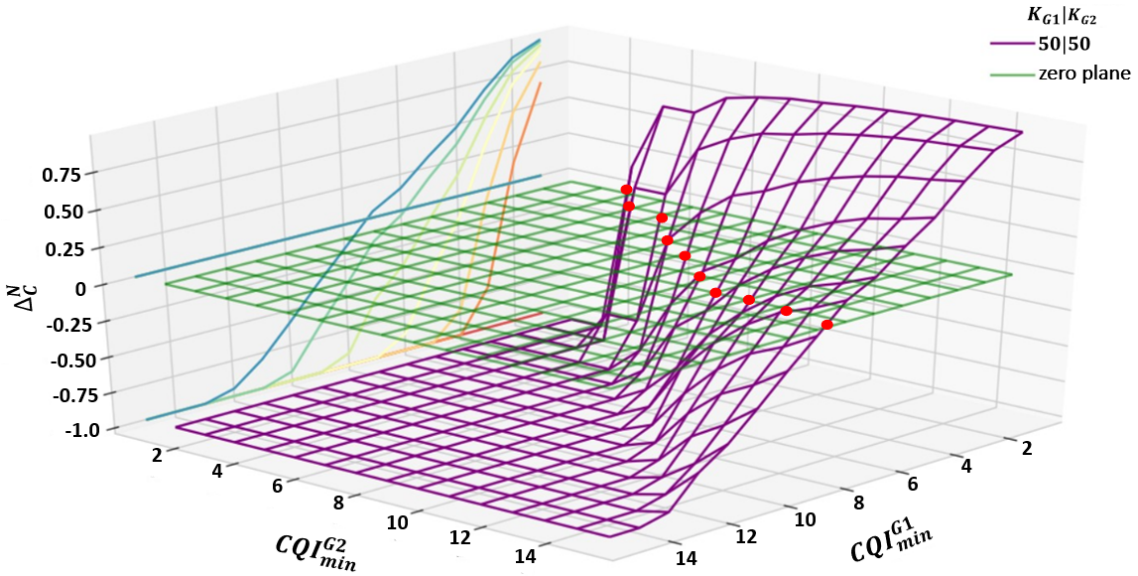


Figure 4.1: 3D surface of Δ_C^N for $il = 0$, and $K_{G1}|K_{G1}(\%) = 50|50$

of K users. Then, in step two, the algorithm applies the best resource allocation algorithms of the multicast techniques under evaluation for each $CQI_{MG,i}$. The resulting maximized values of ADR for the specific multimedia service constraints are used to evaluate the CF (3.27) in step three. Then, the algorithm runs 50 iterations ($t_{max} = 50$), finding a x_i and its corresponding $CQI_{MG,i}$ that satisfy (3.27) minimizing the module of the equation as close to 0 as possible.

To understand how the algorithm converges, Fig. 4.1 shows the 3D surface of Δ_C^N over the whole range of CQI_{min}^{G1} and CQI_{min}^{G2}. To generate the surface, we fix il equal to zero, which is the upper bound of the S-NOMA performance regarding the CMS technique. The negative values in the Δ_C^N axes are restricted to -1 for $\Delta_C^N \leq -1$ and the undefined values when $CQI_{min}^{G1} \geq CQI_{min}^{G2}$. In the figure, we can see how for each CQI_{min}^{G2} can be found a CQI_{min}^{G1} where the surface crosses the zero plane. These zero-crossing points are the target values of algorithm OCI-PSO, where (3.27) is satisfied.

4.2 Results and Discussion

4.2.1 Scenario Description

To assess the effectiveness of the proposed solution, we perform a simulation campaign based on the ad-hoc LLS [42] developed in Python. Table 4.1 summarizes the main simulation parameters.

In the recreated use case, we consider a UMi Street Canyon open area of 100

Parameter	*Value*
Scenario type	UMi Street Canyon
Frequency	28 GHz
Numerology, μ	2
RBs bandwidth, B_0	720 kHz
Available RBs, R_M	100
Noise spectral density	-174 dBm/Hz
BS/user height	10/1.5 m
Transmission power	10 dBm
BS/UE antenna Gain	10/0 dBi
Antenna	Sectorial (120°)
Large-scale fading models	[104]
Small-scale fading model	Jakes [102]
Dynamic line of sight	Yes
Mobility model	Random directional [116]
Users' speed	0.8 m/s

Table 4.1: Simulation parameters.

$\times$ 100 meters with a single cell 5G NR BS located at the center of the grid and operating at 28 GHz. We consider the NR numerology $\mu = 2$ for an SCS equal to 60 kHz and $B_0 = 720$ kHz. We assume that the number of available RBs for the multicast multimedia session is equal to 100 ($R_M = 100$). We consider 100 users ($K = 100$) group-oriented distributed in the service area, s.t. $x_i = \{G1_{min}^{(i,j=1)}, G1_{max}^{(i,j=2)}, G2_{min}^{(i,j=3)}, G2_{max}^{(i,j=4)}\}$ as described in the previous subsection. We simulate multiple multicast multimedia services constraints, evaluating Th_{min} from 1 Mbps to 50 Mbps at steps of 0.5 Mbps. For the simulations, we do not consider any constraint for Th_{max}, which means that the resource allocation algorithms must allocate all the available RBs ($R_M^* = R_M$). Moreover, we fix P equal to 100 %.

The simulation intends to recreate a realistic scenario that can be an open area museum where two subgroups of users have a group-oriented movement following a museum guide in the service area. To reduce the complexity of the simulation campaign, and without loss of generality, we only include in the analysis the following combinations of percent of the MG members into $G1$ and $G2$:

$$K_{G1}|K_{G2}\ (\%) = \{90|10, 70|30, 50|50, 30|70, 10|90\}. \tag{4.3}$$

This assumption helps to discretize all the possible distributions of the K users into $G1$ and $G2$ during the subgrouping process to these five combinations. Moreover, it facilitates presenting the results regarding these combinations that can be further extrapolated to other users' distributions.

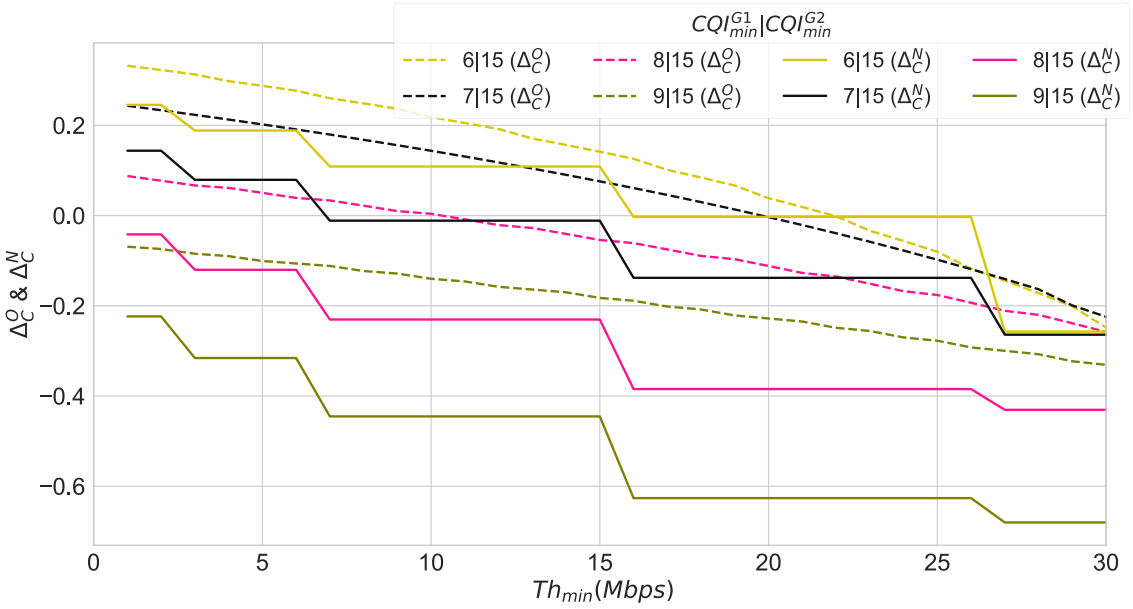


Figure 4.2: Δ^O_C and Δ^N_C versus Th_{min}, for $K_{G1}|K_{G2}(\%) = 50|50$ and $CQI^{G2}_{min} = 15$.

4.2.2 Performance Characterization

Let us start the analysis by comparing the performance of the subgrouping techniques against the conventional CMS approach. In Fig. 4.2, we plot the curves of Δ^O_C and Δ^N_C versus Th_{min} for the simulated network conditions where CQI^{G1}_{min} ranges from 6 to 9, fixing $CQI^{G2}_{min} = 15$, and for $K_{G1}|K_{G2}(\%) = 50|50$. As it was mentioned above, a positive value of Δ^O_C and Δ^N_C means that S-OMA and S-NOMA outperform the benchmark CMS; otherwise, the traditional CMS has the best performance.

The figure shows how, for each combination of $CQI^{G1}_{min}|CQI^{G2}_{min}$, the curves tend to the maximum value as long as the multimedia service constraint Th_{min} tends to 1 Mbps. In the case of Δ^O_C, it happens because as Th_{min} decrease, the S-OMA algorithm must allocate fewer r_{G1} to $G1$ to fulfill the service constraints. Therefore, the remainder RBs are assigned to $G2$, which improves the performance of the subgrouping approach, maximizing the $C_{G2,O}$ and the resulting ADR_O. Moreover, as well as Th_{min} increases, the S-OMA algorithm must reallocate more r_{G1} to $G1$, increasing the $C_{G1,O}$ (i.e., for $C_{G1,O} \geq Th_{min}$) at the expense of reducing the $C_{G2,O}$. It results in a progressive S-OMA performance degradation while the CMS performance remains constant.

In the case of Δ^N_C, for lower values of Th_{min}, the selected IL for the S-NOMA algorithm tends to be zero. The algorithm chooses the minimum possible $eff^{UL,G1}_{min}$ for $G1$ to fulfill the Th_{min}. Therefore, the $eff^{LL,G2}_{min}$ tends to be equal to the maximum eff^{G2}_{min} supported by $G2$ (without the SINR adaptation), and the allocated resources are close to the maximum capacity supported by $G2$, maximizing the

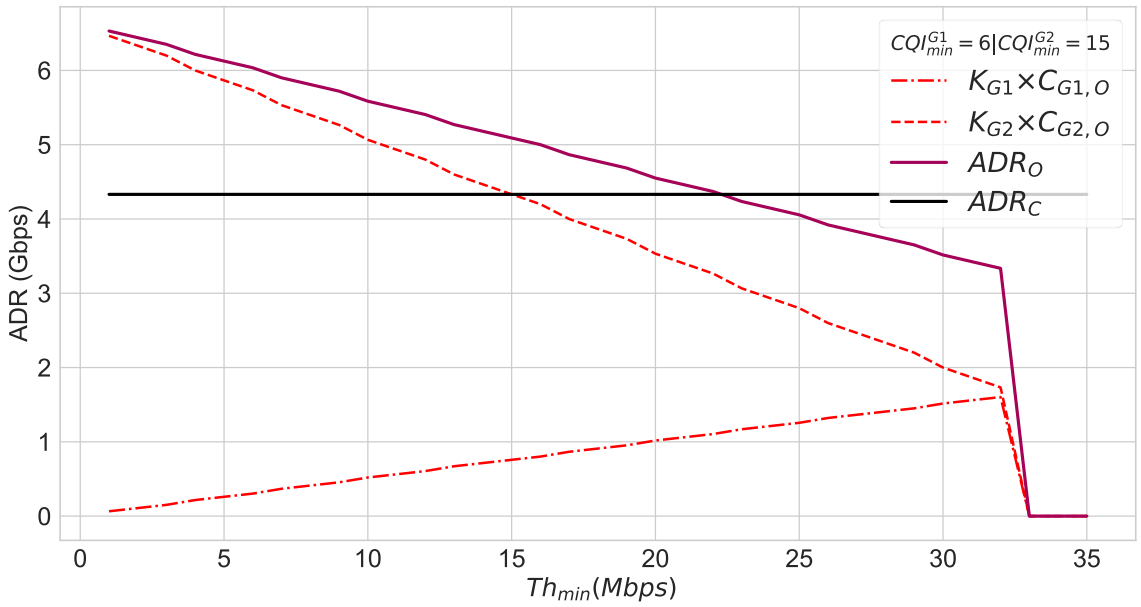


Figure 4.3: $C_{G1,O}$, $C_{G2,O}$, ADR_O and ADR_C variations according to Th_{min}, for $K_{G1}|K_{G2}(\%) = 50|50$, $CQI^{G1}_{min} = 6$ and $CQI^{G2}_{min} = 15$.

$C_{G2,N}$ and ADR_O. Nevertheless, for higher values of Th_{min}, the IL tends to -25, which means that the LL is progressively buried deeper down the UL. Hence the algorithm must select a lower $eff^{LL,G2}_{min}$ (a more robust MCS) for $G2$ to ensure the correct successive interference cancellation.

The stepped shape of the curves for Δ^N_C is because in NOMA, based on LDM, the resolution of the possible $C_{G1,N}$ and $C_{G2,N}$ have only fifteen levels defined by the available MCSs and the corresponding efficiency values. In contrast, in S-OMA, the resolution is determined by the fifteen available MCSs multiplied by the 99 possible RBs distributions (for $R_M = 100$).

In contrast to the behavior of S-OMA and S-NOMA, the CMS technique is independent of Th_{min}. The algorithm for CMS always allocates all the R_M to the entire MG as long as a Th_{max} is not considered. The ADR_C value only depends on the CQI_{min} defined by the lowest channel quality users in the MG and their corresponding MCS.

To complement this analysis, Fig. 4.3 shows the result of fixing the user distribution in terms of CQI and progressively increasing the Th_{min} constraint. As we can see, $C_{G1,O}$, $C_{G2,O}$, and ADR_O vary alongside with Th_{min} while the ADR_C remains constant. As Th_{min} increases, the multicast RRM reallocates RBs from $G2$ to $G1$ until the algorithm cannot reallocate more RBs without letting $C_{G2,O}$ under the service constraint Th_{min}. After this specific point, the S-OMA technique cannot accomplish the simulated service constraint. Moreover, we can see how, for Th_{min} greater than 22 Mbps, the ADR_C is greater than the ADR_O, which matches the

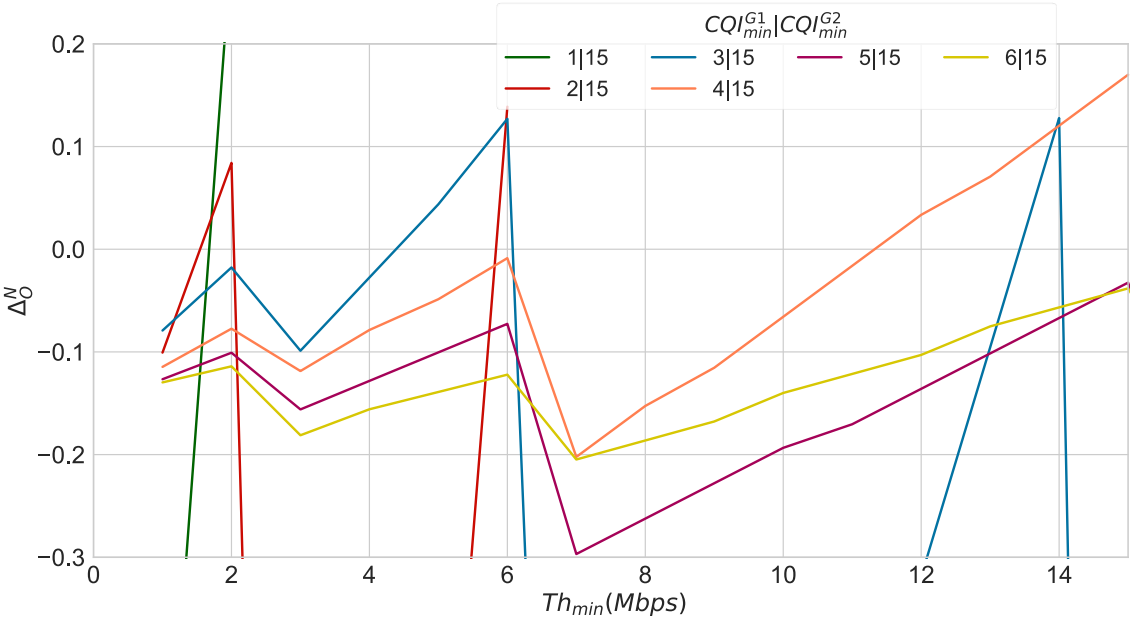


Figure 4.4: Δ^N_O versus Th_{min}, for $K_{G1}|K_{G2}(\%) = 50|50$ and $CQI^{G2}_{min} = 15$.

results in Fig. 4.2, at the point where the benchmark CMS starts exceeding (i.e., zero-crossing for $Th_{min} \approx 22$ Mbps) S-OMA for the curve of $CQI^{G1}_{min} = 5$.

In Fig. 4.2, we can see that the separation between CQI^{G1}_{min} and CQI^{G2}_{min} also influences the positive or negative values of Δ^O_C and Δ^N_C. When CQI^{G1}_{min} tends to be equal to one, the multi-rate technique outperforms the CMS technique. Nevertheless, when CQI^{G1}_{min} tends to be equal to CQI^{G2}_{min}, the curves tend to be all negative independently of the allocated resources.

Such analysis helps us to identify specific network conditions where for CQI^{G1}_{min} values greater than or equal to 9, the traditional CMS technique always outperforms S-OMA. For CQI^{G1}_{min} values greater than or equal to 8, the CMS technique consistently exceeds S-NOMA. Such conditions are valid for any $K_{G1}|K_{G2}(\%)$ distribution where K_{G1} is greater than or equal to 50 % of the MG members.

In Fig. 4.4, we plot the curves of Δ^N_O versus Th_{min} for CQI^{G1}_{min} from 1 to 6, fixing $CQI^{G2}_{min} = 15$, and for $K_{G1}|K_{G2}(\%) = 50|50$. As was analyzed before, a positive value of Δ^N_O means that the S-NOMA technique outperforms S-OMA; otherwise, S-OMA performs best. The results show how, for each curve, the multicast strategy S-NOMA tends to increase its performance regarding higher values of Th_{min}. Nevertheless, the curves for $CQI^{G1}_{min} > 4$ are mostly negative. The best performance of S-NOMA concerning S-OMA is for $CQI^{G1}_{min} \leq 3$.

Fig. 4.5 presents the final system ADR achieved with S-OMA, S-NOMA, and CMS for the same network conditions. This figure enables us to compare the behavior of the three analyzed multicasting strategies for group-oriented distributed

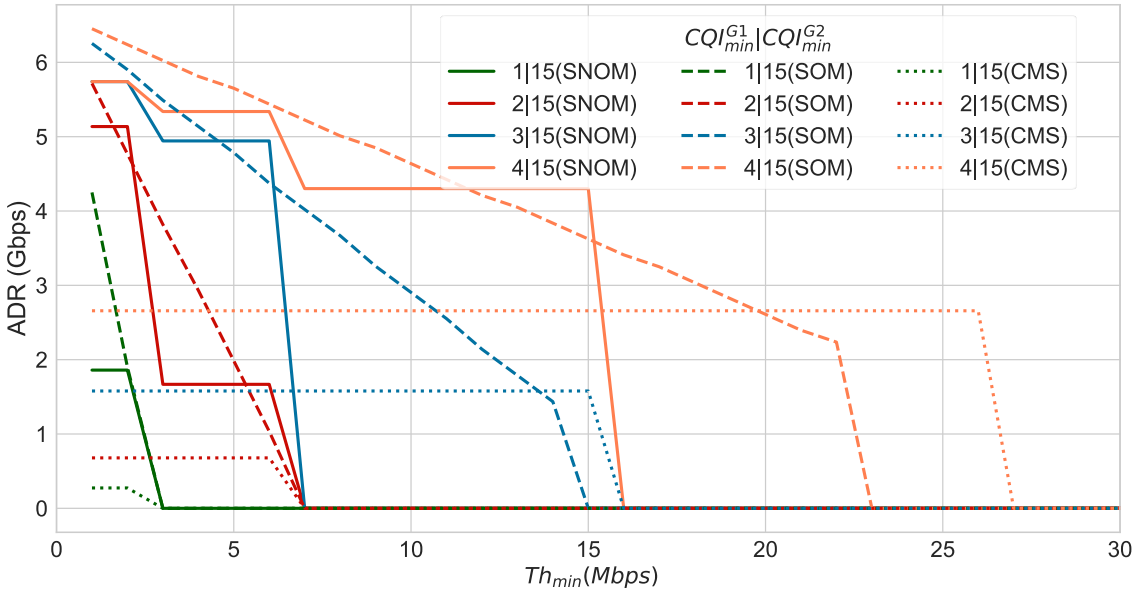


Figure 4.5: System ADR versus Th_{min} of Δ_C^N for $K_{G1}|K_{G2}(\%) = 50|50$ and $CQI_{min}^{G2} = 15$.

users concerning the system ADR. We can see how the interception points among ADR_N, ADR_O, ADR_C match with the zero crossing of the curves in Fig. 4.2 and 4.4, respectively. Moreover, the figure shows how, for network conditions where the CQI_{min}^{G1} is equal to 1 or 2, the multi-rate approaches always ensure the highest system ADR.

From Fig. 4.2 to 4.5, we analyze the interrelations among the variables that shape the performance of the three multicast techniques under evaluation. The presented analysis helps to understand how these multicast strategies behave under specific conditions regarding the service constraints and the users' CQI distribution. Nevertheless, the analysis presented is restricted to the cases where the CQI_{min}^{G2} is fixed to 15 and for the 50 % of the MG members into $G1$ ($K_{G1}|K_{G2}(\%) = 50|50$). To complement this analysis, Fig. 4.6 shows the 3D surfaces of Δ_C^N over the whole range of CQI_{min}^{G1} and CQI_{min}^{G2}, fixing IL equal to zero, and for $K_{G1}|K_{G2}(\%) = 90|10$ and $K_{G1}|K_{G2}(\%) = 10|90$. As discussed for Fig. 4.1, the negative values in the Δ_C^N axes are restricted to -1 for $\Delta_C^N \leq -1$ and the undefined values when $CQI_{min}^{G1} \geq CQI_{min}^{G2}$.

Fig. 4.6 recreates two extreme group-oriented users' distributions, helping to show the impact of this variable. In the case of $K_{G1}|K_{G2}(\%) = 10|90$, the subgrouping approach improves the performance because 90 % of the users are experiencing high channel quality conditions and receive a tailored MCS according to their specific reception conditions. In contrast, with the conservative CMS, 100 % of the users are treated as the 10 % with low channel quality conditions. The performance of the proposed multi-rate algorithms increases as long as the separation between CQI_{min}^{G1}

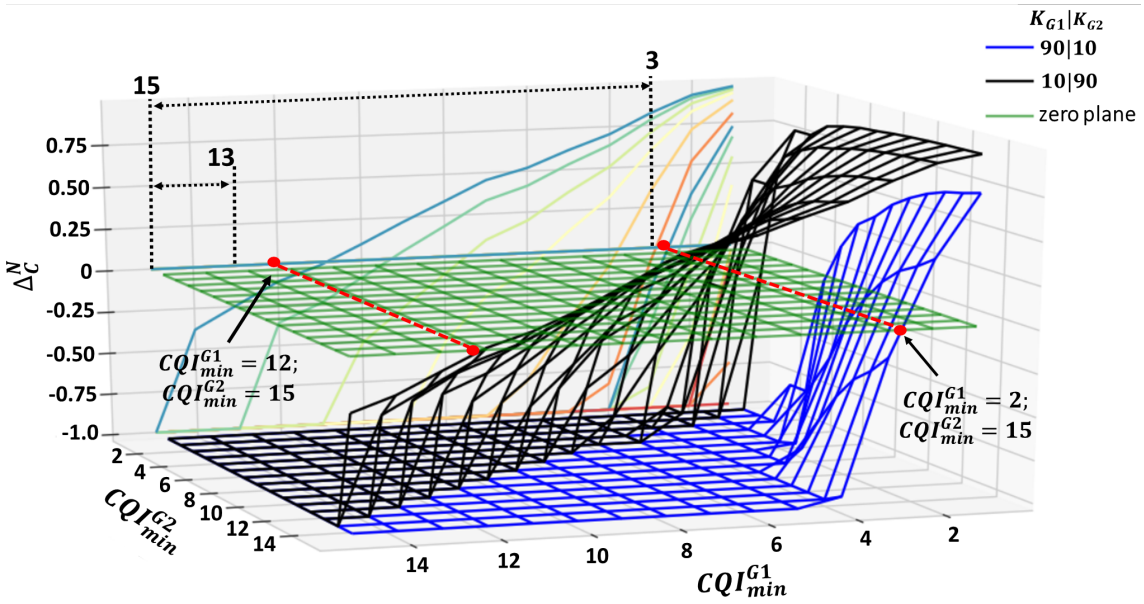


Figure 4.6: 3D surface of Δ_C^N for $il = 0$, and $K_{G1}|K_{G1}(\%) = 90|10$ and $K_{G1}|K_{G2}(\%) = 10|90$.

and CQI_{min}^{G2} increases. In the figure, we highlight the range for $CQI_{min}^{G1} \geq 13$, where the CMS technique consistently exceeds S-NOMA if there are only 10 % of the users in $G1$. In contrast, if there are 90 % of the users in $G1$, the range where CMS consistently outperforms S-NOMA is for any $CQI_{min}^{G1} \geq 3$, being the 3D surfaces primarily negative. The same analysis presented in Fig. 4.6 can be applied for Δ_C^O with identical behavior.

Table 4.2 summarizes the simulation results for CQI_{min}^{G1} from 1 to 15 and fixing $CQI_{min}^{G2} = 15$. We generalize the results for the combinations of the MG members into $G1$ and $G2$ defined in (4.3). We can appreciate how, as the percentage of users in $G1$ increases, the region of overperformance of CMS over the multi-rate approach increases. For example, if K_{G1} is equal to 90 % of the users, the CMS technique always exceeds the subgrouping techniques when the $CQI_{min}^{G1} \geq 3$. In contrast, if K_{G1} is equal to 10 % of the users, the CMS technique only exceeds both subgrouping techniques when the $CQI_{min}^{G1} \geq 14$. The major outcome from such analysis is how, for $CQI_{min}^{G1} = 1$ or $CQI_{min}^{G1} = 2$ with $CQI_{min}^{G2} = 15$, the proposed multi-rate algorithms always outperform the traditional CMS solution for any percent of the MG members into $G1$ less than or equal to 90 %.

The above-presented results are limited to conditions where $CQI_{min}^{G2} = 15$. Hence, Table 4.3 helps us to extend this analysis. This Table summarizes the simulation results regarding the best multicast access technique selection for network conditions where $CQI_{min}^{G1} = 1$ or $CQI_{min}^{G1} = 2$, evaluating CQI_{min}^{G2} from 1 to 15 s.t. (4.3). These results allow to define that for a $CQI_{min}^{G1} = 1$ with $CQI_{min}^{G2} \geq 8$ and for

	CQI^{G1}_{min}(CMS → C; SOM →O; SNOM → N)													
$K_{G1}\|K_{G2}$(%)	1	2	3	4	5	6	7	8	9	10	11	12	13	14
90|10	O/N		C											
70|30	O/N		O/N/C		O/C	C								
50|50	O/N		O/N/C			O/C			C					
30|70	O/N		O/N/C					O/C			C			
10|90	O/N		O/N/C			O/C								C

Table 4.2: Best multicast access technique for CQI^{G1}_{min} from 1 to 14, fixing $CQI^{G2}_{min} = 15$, s.t. (4.3).

		CQI^{G2}_{min} (CMS → C; SOM →O; SNOM → N)														
$K_{G1}\|K_{G2}$	CQI^{G1}_{min}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
90|10(%)	1	C				O/C			O/N							
90|10(%)	2	C								O/N/C			O/N			
70|30(%)	1	C		O/C			O		O/N							
70|30(%)	2	C			O/C							O/N				
50|50(%)	1	C		O/C			O		O/N							
50|50(%)	2	C		O/C								O/N				
30|70(%)	1	C		O/C			O		O/N							
30|70(%)	2	C		O/C								O/N				
90|10(%)	1	C		O/C		O/N										
90|10(%)	2	C		O/C						O/N /C		O/N				

Table 4.3: Best multicast access technique for CQI^{G2}_{min} from 1 to 15, fixing CQI^{G1}_{min} equal to 1 and 2, s.t. (4.3).

$CQI^{G1}_{min} = 2$ with $CQI^{G2}_{min} \geq 12$, the subgrouping approach always outperforms the CMS technique for any percent of the MG members into $G1$ less than or equal to 90 %.

Table 4.4 presents the whole set of ranges where CMS outperforms S-OMA and S-NOMA for any CQI group-oriented distribution subject to the CQI constraints defined in the table. The presented CQI ranges define in which conditions, in terms of users' channel quality distribution, we must not consider the subgrouping approach resource allocation strategies to provide the multicast multimedia service. For example, for Δ^N_C with $K_{G1}|K_{G2}(\%) = 50|50$, the first CQI range 9-15, includes the first major threshold previously presented in the Fig. 4.2 analysis. This range means that for $CQI^{G1}_{min} \geq 9$ and $9 \leq CQI^{G2}_{min} \leq 15$, the benchmark CMS technique always outperforms S-OMA independently of the multimedia service constraints.

For any $CQI_{min}^{G1} \geq x$ and $CQI_{min}^{G2} \leq y$ (with $CQI_{min}^{G1} \leq CQI_{min}^{G2}$) CMS outperforms S-OMA and S-NOMA techniques independently of the Th_{min}, subject to (4.3).

90|10(%)		70|30(%)		50|50(%)		30|70(%)		10|90(%)	
Δ_C^O	Δ_C^N	Δ_C^O	Δ_C^N	Δ_C^O	Δ_C^N	Δ_C^O	Δ_C^N	Δ_C^O	Δ_C^N
x-y	x-y	x-y	x-y	x-y	x-y	x-y	x-y	x-y	x-y
3-15	3-15	6-15	6-15	9-15	8-15	11-15	10-15	14-15	13-15
2-8	2-7	5-13	5-14	8-13	7-13	9-12	9-13	13-14	11-14
1-4		4-10	4-12	7-11	6-12	8-10	8-12	12-13	10-12
		3-7	3-8	6-10	5-10	7-9	6-10	11-12	9-11
		2-3	2-5	5-8	4-9	6-8	5-8		8-10
		1-2		4-7	3-6	5-7	4-7		7-9
				3-4	2-4	4-5	3-5		6-8
							2-4		5-7
									4-6
									3-5
									2-4

Table 4.4: Conditions where CMS outperforms S-OMA and S-NOMA independently of Th_{min}.

For CQI values out of these ranges, the relative performance depends on the specific multimedia service constraints or can be defined according to Tables 4.2 and 4.3 as it was previously analyzed.

Table 4.5 presents the conditions where CMS outperform S-OMA and S-NOMA regarding the Th_{min} constraint and assuming the number of available RBs $R_M = 100$. These results help to understand when to consider the subgrouping approach or just the CMS technique during multicast resource allocation.

Fig. 4.7 shows the implication of the multiuser channel quality diversity. In this simulation, we fix the users belonging to $G2$ with $CQI_{min}^{G2} = 15$ and evaluate the achieved system ADR of the evaluated multicast access techniques when moving the CQI_{min}^{G1} from 1 to 15. During the simulation runs, we fix the $K_{G1}|K_{G2}(\%)$ as 90|10, 50|50 and 10|90. In the figure, we can see how the ADR achieved with the benchmark CMS algorithm grows almost linearly, and it is independent of the percent of users in $K_{G1}|K_{G2}$. The performance of CMS only depends on the CQI of the lowest channel quality user in the MG. Nevertheless, the performance of the subgrouping approaches is highly influenced by the percent of users into $G1$ and $G2$ as it was above analyzed in Fig. 4.6. Moreover, the figure illustrates how the advantage of delivering a multicast service based on multi-rate subgrouping increases as the channel quality diversity between low and high-channel-quality users increases.

The presented results delimit the CQI application regions for each evaluated mul-

For any service constraint higher than or equal to Th_{min}, the CMS always outperforms S-OMA, subject to (4.3) and $R_M = 100$.

$K_{G1}\|K_{G2}(\%)$	90|10	70|30	50|50	30|70	10|90
$Th_{min}(Mbps)$	6.0	17.0	22.5	28.0	30.5

For any service constraint higher than or equal to Th_{min}, the CMS always outperforms S-NOMA, subject to (4.3) and $R_M = 100$.

$K_{G1}\|K_{G2}(\%)$	90|10	70|30	50|50	30|70	10|90
$Th_{min}(Mbps)$	6.5	15.5	26.5	26.5	26.5

Table 4.5: Conditions where CMS outperforms S-OMA and S-NOMA regarding Th_{min}.

ticast strategy and how they are modified by the percent of MG members classified into low and high-channel quality subgroups. An essential outcome of the simulation campaign is to prove that there is no superior multicast strategy. Therefore, the single-rate and multi-rate multicast strategies must coexist in the multicast to ensure an adequate service delivery over the envisioned 5G networks and beyond.

The followed research methodology, outcomes, and the corresponding analysis help to understand the interrelation among the multiple variables that shape the performance of the evaluated multicast access techniques. It represents an excellent framework for handling the trade-off between multicasting gain and multiuser diversity.

4.3 Conclusions

This research addresses the MG diversity through dynamic single-rate and multi-rate multicasting defining conditions for an effective coexistence of the evaluated multicast access techniques. We propose a tailored RRM algorithm aided by subgrouping and PSO. We provide conditions for an effective dynamic selection of the multicast strategy that better suits the specific network characteristics regarding the user's reception conditions, multimedia service constraints, and network parameters. We assess the proposal's effectiveness by recreating a realistic 5G MBS use case and evaluating a wide range of possible conditions.

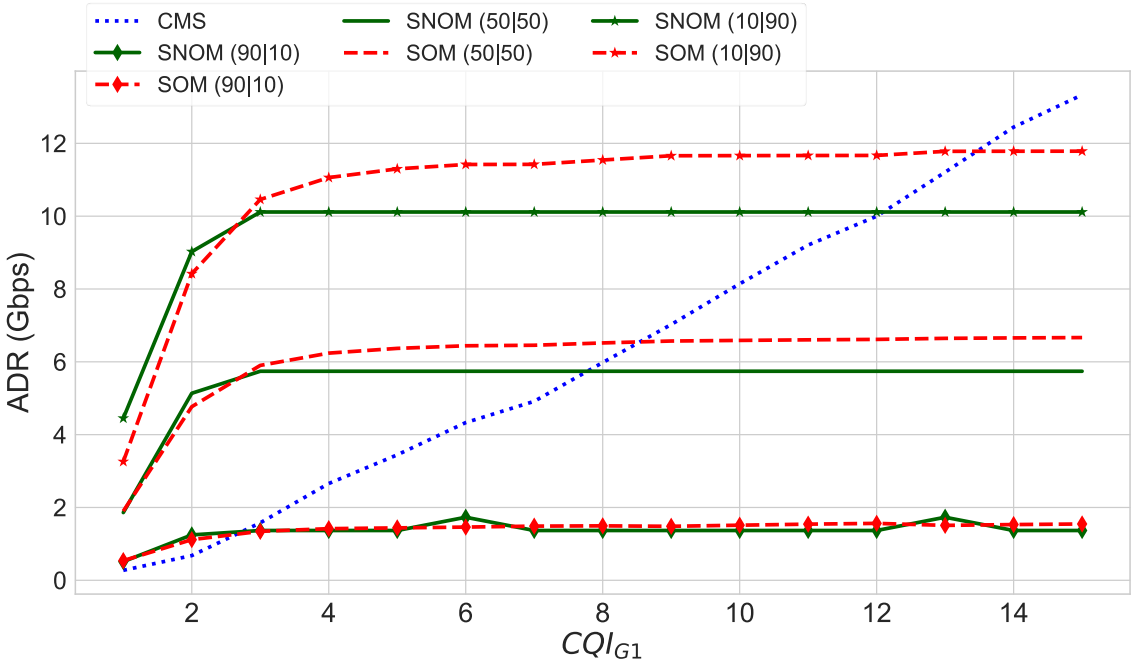


Figure 4.7: System ADR achieved with CMS, S-OMA and S-NOMA for $CQI_{min}^{G2} = 15$, different $K_{G1}|K_{G2}(\%)$, and CQI_{min}^{G1} from 1 to 15.

Chapter 5

Multicast Access Technique Selection and Resource Management

In this chapter, we propose heuristic-based algorithms to prove the QoS advantage of dynamically selecting among the traditional single-rate strategy CMS and the multi-rate subgrouping-based solutions. Moreover, we design an algorithm that combines multicasting over fixed pre-computed MIMO multi-beams and multi-rate subgrouping, taking advantage of the users' spatial and channel quality diversity. The proposed solutions, analysis, and conclusions respond to (**SO-4**).

The main contributions can be summarized as follows: (*i*) We provide a method for the dynamic multicast access technique selection and resource allocation based on the proposed performance equations; (*ii*) The proposal shows the advantages of dynamically selecting among CMS, S-OMA, and S-NOMA to effectively handle the trade-off between the available multicasting gain and the existing multiuser diversity; (*iii*) We propose a multicasting strategy based on fixed pre-computed MIMO multi-beams and S-NOMA, tackling specific throughput requirements for enabling XR applications to attend multiple users in a 5G MBS use case; (*iv*) The following approach allows us to dynamically take advantage of the users' spatial and channel quality diversity, maximizing specific QoS metrics; (*v*) The proposals are evaluated through numerical and link-level simulations recreating realistic network conditions for the application of the proposed algorithms.

Algorithm 5: OE-MAT

Input: $Th_{min}, Th_{max}, P, CQI_{MG}, R_M, \mathcal{I}, B_0$
Output: Selected multicasting strategy and corresponding resource allocation s.t. the Algorithm 1, Algorithm 2 or Algorithm 3.
1: From CQI_{MG} **apply the subgrouping process to determine** CQI_{MG}^{G1} **and** CQI_{MG}^{G2}
2: From CQI_{MG}^{G1} **and** CQI_{MG}^{G2} **determine** CQI_{min}^{G1}**,** CQI_{min}^{G2} **and its corresponding** eff_{min}^{G1} **and** eff_{min}^{G2} **from Table 5.2.2.1-3 in [98].**
3: Apply SINR adaptation [77] iterating over $\mathcal{I}$ **to find:**
$eff_{min}^{UL,G1}$ and $eff_{min}^{LL,G2}$
s.t. $Th_{min} \leq C_{N,G1} \leq C_{N,G2}$, with (3.5) and R_M
if ***foreach*** $\mathcal{I}$**:** $C_{N,G1} \leq Th_{min}$ **then**
  return: $ADR_{\mathrm{C}}^{max} = 0$
  Not available resources to accomplish with Th_{min}
else if $Th_{min} \leq C_{N,G1} \leq C_{N,G2}$ **then**
  continue
end
4: Compute r_{min}^{G1}**:**
$r_{min}^{G1} = Th_{min}/(B_0 \times eff_{min}^{G1})$
5: Compute κ **and** η**:**
$\kappa = K_{G1}/K$; $\eta = r_{min}^{G1}/R_M$
6: Compute 3.21, 3.22 and 3.23
if $\Delta_C^O \leq 0$ *and* $\Delta_C^N \leq 0$ **then**
  return: CMS (Algorithm 1)
else if $\Delta_C^O \geq 0$ *and* $\Delta_O^N \leq 0$ **then**
  return: S-OMA (Algorithm 2)
else if $\Delta_C^N \geq 0$ *and* $\Delta_O^N \geq 0$ **then**
  return: S-NOMA (Algorithm 3)

5.1 Dynamic Multicast Access Technique Selection

Durin the problem formulation, specifically in Subsection 3.2.1, we defined the equations (3.24), (3.25), and (3.26) to evaluate the relative ADR performance among CMS, S-OMA, and S-NOMA. Starting from these equations, we can determine which multicast access technique performs best regarding the criteria system ADR and the specific network conditions. Using these equations as a baseline, we proved in Chapter 4 that the traditional single-rate CMS and the multi-rate strategies S-OMA and S-NOMA must dynamically coexist to maximize the network performance.

Aimed to evaluate the advantage of dynamically selecting among these multicast strategies, we propose a dynamic multicast access technique selection algorithm.

We identify this algorithm as Outperformance Equation-Multicast Access Technique Selection (OE-MAT). Algorithm 5 presents the pseudocode that describes the designed OE-MAT. The inputs of the algorithms are the inputs needed to run the specific algorithms for CMS, S-OMA, and S-NOMA. The main particularity of this solution is the use of (3.24), (3.25), and (3.26) in step six to select the best multicast access technique dynamically. Therefore, the output of OE-MAT is the selection of the multicast access technique and the corresponding resource allocation subject to Algorithm 1, 2 or 3. Hence, OE-MAT maximizes the system ADR according to the multimedia service constraints and the users' reception conditions.

5.1.1 Scenario Description

We used a numerical simulation campaign to evaluate the effectiveness of the proposal, recreating a single-cell 5G NR considering multiple multimedia service constraints, system resources, and users' distribution. To assess the behavior of the proposed algorithm and analyze the corresponding results, we use the performance metrics ADR, PF, and MDI. We included in the assessment the parameter Δ_R to measure how efficiently the evaluated multicast strategies use the available resources to accomplish the service constraint.

For the numerical simulation campaign, we assume $\mu = 0$ (oriented to eMBB applications) for an SCS equal to 15 kHz and $B_0 = 180$ kHz. We consider R_M equal to 50, 100, and 200 RBs. From equation 3.1, we compute the $SINR_{min}$ needed to correctly demodulate the MCS corresponding to the CQI feedback. To develop the numerical simulations, we consider the following SINR distributions:

- K users with an $SINR_k$ randomly distributed in the ranges: 0-15 dB, 15-30 dB, or 0-30 dB.
- K users group-oriented distributed according to their $SINR_k$ values regarding the ranges ($SINR_{k1}|SINR_{k2}$): 0-5|25-30 dB, or 5-10|20-50 dB.

We consider $K_{G1}|K_{G2}$ (%) equals to 90|10 , 50|50, and 10|90 for a total of 100 users (K). For the multimedia service constraints, we assume multiple pairs of Th_{min} and Th_{max} ($Th_{min}|Th_{max}$) that help to recreate a wide range of possible multicast applications. Moreover, we fix P equal to 100 %. Table 5.1 summarizes the evaluated parameters. When Th_{max} is considered infinity, it implies that the algorithms will use all the available RBs to ensure the maximum possible capacity to maximize the system ADR.

For each evaluated algorithm, all the variables' combinations are simulated for a total of 1215 outputs of ADR, PF, MDI, and Δ_R. In this case, we consider CMS, S-OMA, and S-NOMA as benchmarks for assessing the algorithm OE-MAT. The outputs of the simulation process result from averaging 20 simulation runs, ensuring a 95 % confidence value [117].

Parameter	*Value*
Numerology, μ	0
RBs bandwidth, B_0	180 kHz
Available RBs, R_M	50, 100, 200
$SINR_{k1}\|SINR_{k2}$ (dB)	0-5|25-30, 5-10|20-50, 0-15, 15-30, 0-30
$K_{G1}\|K_{G2}$ (%)	90|10, 50|50, 10|90
$Th_{min}\|Th_{max}$	1|4, 2|8, 3|12, 4|16, 5|20, 8|32, 10|40, 15|60, 20|80 1|10, 2|12, 3|13, 4|14, 5|15, 8|18, 10|20, 15|25, 20|30 1|∞, 2|∞, 3|∞, 4|∞, 5|∞, 8|∞, 10|∞, 15|∞, 20|∞

Table 5.1: Numerical simulation variables.

5.1.2 Simulation Results

This numerical simulation campaign aims to assess the advantages of a dynamic multicasting strategy based on OE-MAT regarding a fixed approach. As presented in Chapter 4, algorithms 1 (CMS), 2 (S-OMA), and 3 (S-NOMA) are optimized to maximize the system ADR metric. For this reason, the ADR is the core metric during the evaluation process aided by PF, MDI, and Δ_R to complement the analysis. In all covered simulation results, when the metrics ADR, PF, and MDI go to zero, it means that the algorithms do not have enough resources to accomplish the Th_{min} constraint.

Fig. 5.1 shows the achieved ADR with the algorithms CMS, S-OMA, S-NOMA, and OE-MAT. The figures show the results for the first case of $Th_{min}|Th_{max}$ in Table 5.1 where the Th_{max} constraints are four times the Th_{min} (a.1, b.1, c.1) and for $Th_{max} = \infty$ (a.2, b.2, c.2). In the figures (a, b, c), we evaluate $SINR_{k1}|SINR_{k2} = 0-5|25-30$, $SINR_{k1}|SINR_{k2} = 5-10|20-25$ and $SINR_{k1}|SINR_{k2} = 15-30$, respectively. In every case R_M is fixed to 200 and $K_{G1}|K_{G2}$ (%) to 50|50.

In Fig. (a.1) and (b.2), the users follow a group-oriented distribution where 50 % are experiencing lower channel quality conditions than the other 50 %. Under these conditions, the multi-rate approaches outperform the traditional CMS. As the users are experiencing orthogonal reception conditions based on their group-oriented distribution, S-OMA and S-NOMA take advantage by managing the users' diversity. In contrast, where all the users have an SINR between 15 and 30 dB (c.1), the CMS solution tends to exceed the multi-rate strategies because the entire MG is experiencing good reception conditions. The exact general behavior can be found in Fig. (a.2), (b.2), and (c.2), where the only difference is $Th_{max} = \infty$. In these cases, the algorithms allocate all the available RBs, maximizing the performance of each case's most efficient resource allocation strategy. Such results follow the analysis presented in Chapter 4, where it was proved that the regions of outperformance of the multi-rate solutions against CMS are mostly when the users are experiencing group-oriented reception conditions with high levels of diversity.

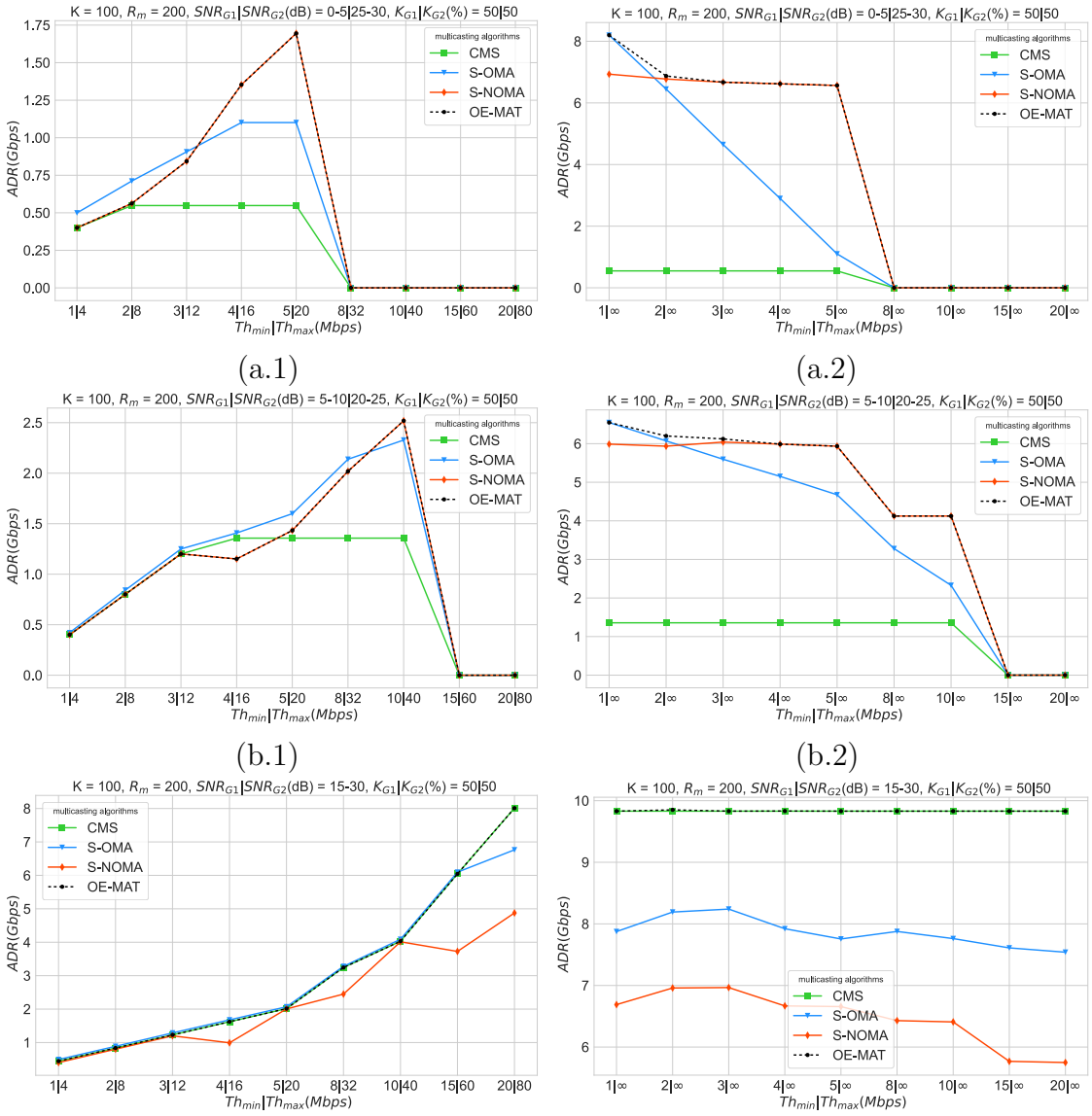


Figure 5.1: System ADR with CMS, S-OMA, S-NOMA, and OE-MAT for $R_M = 200$, $K_{G1}|K_{G2}$ (%) $= 50|50$ with (a) $SINR_{k1}|SINR_{k2} = 0-5|25-30$, (b) $SINR_{k1}|SINR_{k2} = 5-10|20-25$, and (c) $SINR_{k1}|SINR_{k2} = 15-30$; (a.2), (b.2) and (c.2) for $Th_{max} = \infty$.

The primary outcome of Fig. 5.1 is to see how the proposed OE-MAT algorithm always tends to follow the multicast access techniques that maximize the system ADR, proving the advantage of dynamic multicasting. Nevertheless, we can see how, in Fig. (a.1) and (a.2), OE-MAT always selects S-NOMA even when S-OMA achieves a higher SINR. This happened because OE-MAT always selects the strategy that ensures the maximum system ADR according to the total available R_M by evaluating (3.24), (3.25), and (3.26), and not according to the actual allocated R_M^* for each strategy. This is why for (a.2), (b.2), and (c.2) where $R_M^* = R_M$, OE-MAT always follows the multicast access solution that maximizes the ADR for an equal number of RBs. It means that OE-MAT always follows the most efficient technique in terms of resources, and this is why, in Fig. (a.1) and (a.2), the proposed algorithms always select S-NOMA instead of S-OMA.

To complement this analysis, Fig. 5.2 shows the PF, MDI, and Δ_R for the same conditions analyzed in Fig. 5.1 (b). The PF curves (Fig. 5.2 (a)) follow the same tendency that the system ADR curves which do not provide additional information. However, through Fig. 5.2 (b) and (c), we can see how the OE-MAT algorithm allocates the resources following the multicast strategy that provides the higher MDI and Δ_R values. These results show how the proposed dynamic multicast strategy based on the performance equations selects the optimal multicast access technology that ensures the most efficient resource utilization.

The figure shows that the Δ_R curves are equal to zero when $Th_{max} = \infty$. This happens because all the resources are used during the service delivery $R_M^* = R_M$. Consequently, OE-MAT always selects the technique that maximizes the system ADR. Moreover, in Fig. 5.2 (c.1), we prove the advantage of S-NOMA in terms of resource efficient utilization, achieving a considerably higher Δ_R regarding its counterparts. In the case of $Th_{min}|Th_{max}$ equal to 4|16 Mbps, S-NOMA can accomplish with the service constraints using 50 % less RBs than CMS and S-OMA. This result validates why OE-MAT selected the S-NOMA solution in Fig. 5.1 (b.1) even when S-OMA achieve a higher ADR.

An essential outcome of the simulation campaign is proving that there is no superior multicast strategy, following the conclusions and remarks presented in Chapter 4. The results show that the most spectral-efficient multicast strategy depends on the specific users' distributions regarding channel conditions and the multicast service constraints. Fig. 5.3 helps to complement this analysis. For each considered multicast strategy, we can see the percentage of the time (concerning the whole dataset) they ensure the highest performance criteria. Specifically, for all the simulated network conditions, the benchmark MCS technique provides the maximum system ADR 24 % of the time. In contrast, the multi-rate subgrouping based on OMA at 47 %, and subgrouping based on NOMA at 29 % of the time.

The results of these extensive numerical simulations prove the advantage of dynamically selecting the multicast access technique and the corresponding resource allocation, tailoring the specific network conditions and users' dynamics. The proposed algorithm enables the effective coexistence of the traditional single-rate CMS

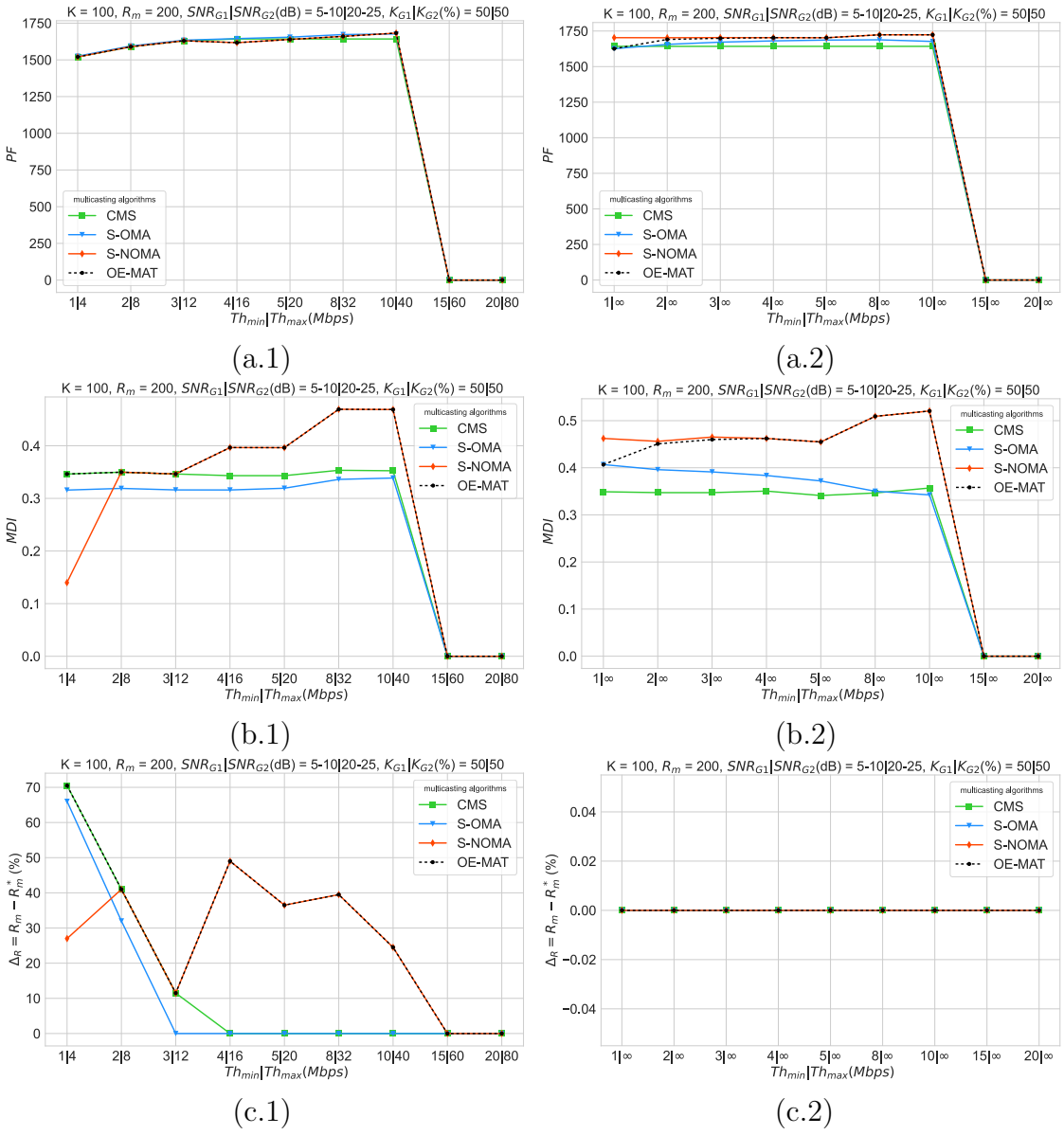


Figure 5.2: (a) PF (b) MDI and (c) Δ_R, with CMS, S-OMA, S-NOMA, and OE-MAT for $R_M = 200$, $K_{G1}|K_{G1}$ (%) $= 50|50$ and $SINR_{k1}|SINR_{k2} = 5-10|20-25$; (a.2), (b.2) and (c.2) for $Th_{max} = \infty$.

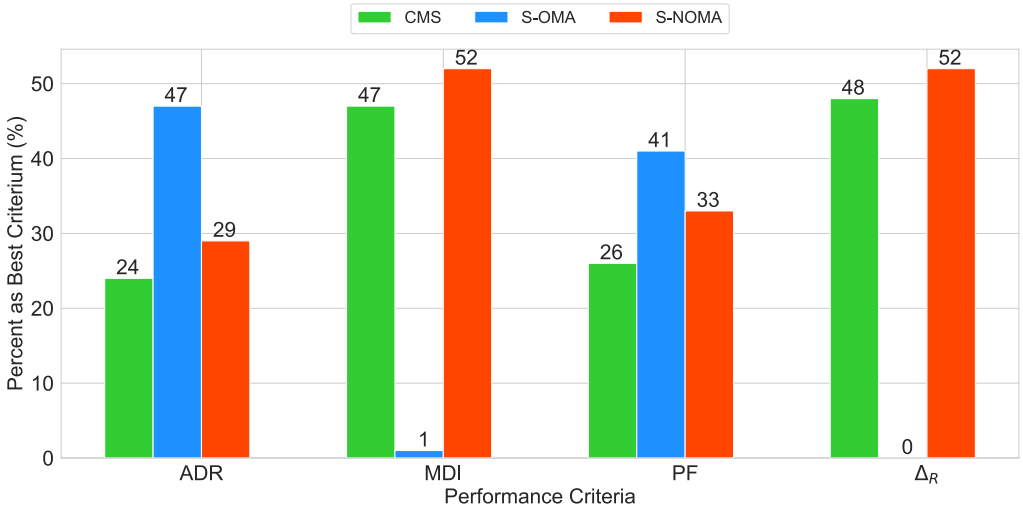


Figure 5.3: Percent of time concerning the whole dataset of CMS, S-OMA, and S-NOMA as best performance criterium.

and the multi-rate strategies S-OMA and S-NOMA, addressing the multiuser diversity and maximizing the network performance.

5.1.3 Conclusions

In this section, we propose a dynamic multicast access technique selection algorithm (OE-MAT) based on the equations (3.24), (3.25), and (3.26) to evaluate the ADR performance among CMS, S-OMA, and S-NOMA. The proposal and the validation outcomes show the advantages of dynamically selecting among different multicast strategies to effectively handle the trade-off between the available multicasting gain and the existing multi-user diversity. The specific multicast multimedia service constraints and the reception conditions of the users determine the dynamic selection followed by OE-MAT. The proposal's effectiveness is validated through an extensive numerical simulation campaign, recreating a single-cell 5G NR BS.

5.2 Fixed MIMO Multi-Beams Multi-rate Multicasting

Recent papers have dealt with RRM for optimal multicast grouping based on the above mentioned technologies. Chukhno *et al.* [27] present an optimization solution in 5G NR systems with directional multi-beam antennas. In [118], the authors introduce a learning-based approach to jointly design multicast/unicast beamformers with imperfect CSI. Iradier *et al.* [79] evaluate different RRM models that combine TDM with NOMA in the 5G mmWave frequency bands.

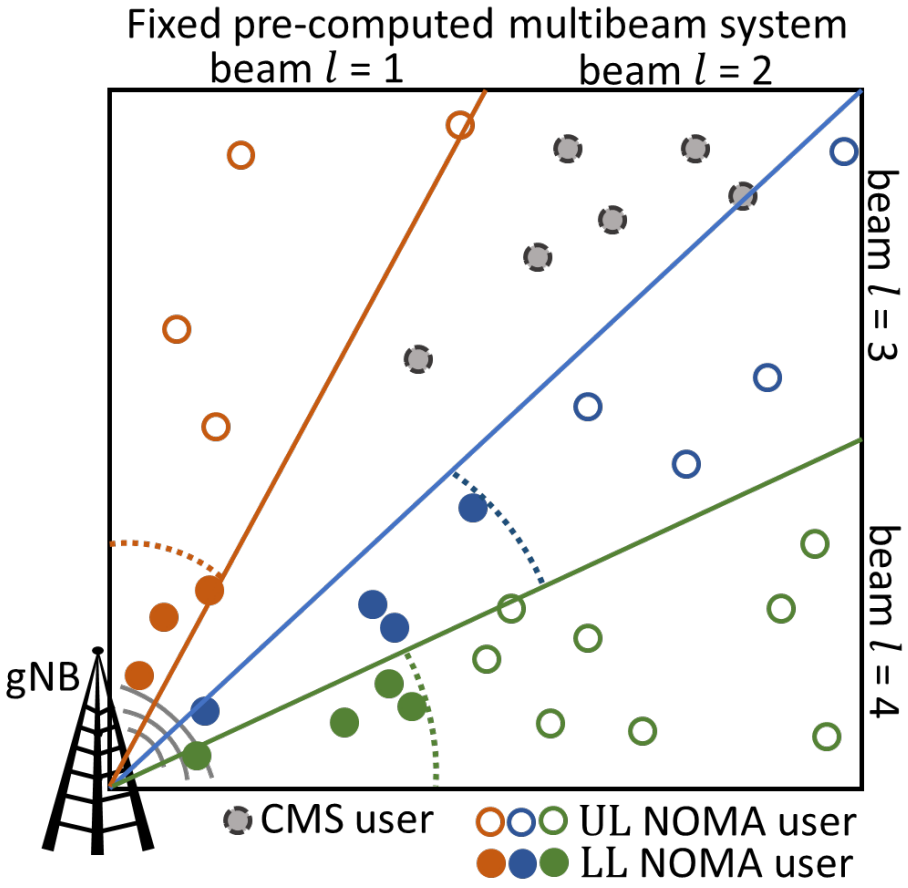


Figure 5.4: Considered deployment scenario.

Bearing this previous works as the benchmark, we propose a multicasting strategy based on fixed pre-computed MIMO multi-beams and S-NOMA, tackling specific throughput requirements for enabling XR applications to attend multiple users in 5G and beyond environments. The following approach allows us to dynamically take advantage of the users' spatial and channel quality diversity, maximizing specific QoS metrics. The proposed solution is suitable for XR applications in pedestrian mobility scenarios such as museums, exhibitions, and fairs. We evaluate our proposal through link-level simulations of an AR/VR multicast delivery use case. We recreate a restrained outdoor scenario with a single-cell 5G NR BS operating in mmWave and multiple pedestrian users requesting the same multicast services.

The goal is to maximize the system ADR of the multicast users regarding a traditional CMS strategy based on a single beam to cover the whole MG. Algorithm 6 presents the pseudocode of the proposed dynamic multicasting over fixed pre-computed multi-beams (DMFM) strategy. The solution combines fixed pre-computed multicast multi-beams with opportunistic S-NOMA based on specific spatial and channel quality diversity of the MG, as illustrated in Fig. 5.4. The

Algorithm 6: DMFM

Input: Users' spatial information, grid, and network parameters: $\mathcal{K}$, $\mathcal{XYZ}$, B_0, R_M, $\mathcal{I}$, Th_{min}, Th_{max}, $\mathcal{L}$

Output: Best multicast resource allocation strategy per beam

1: From the spatial and grid information performs the users' clustering s.t. the available pre-computed beams:

$\mathcal{K} = \mathcal{K}_{l=1} \cup \mathcal{K}_{l=2}, \ldots, \cup\, \mathcal{K}_{l=L}$

2: CQI feedback collection for each user regarding its corresponding beam:

$CQI_{\mathrm{MG}'} = CQI_{l=1} \cup CQI_{l=2}, \ldots, \cup\, CQI_{l=L}$

3: Identify the best resource allocation strategy that maximizes the ADR for each CQI subset CQI_l and allocate the resources to each $\mathcal{K}_l$ subset in $\mathcal{K}$:

foreach l ***in*** L **do**
  if $M_l > 1$ **then**
    Split $\mathcal{K}_l$ into $G1$ and $G2$ s.t. CQI_l to find the optimal: CQI_l^{G1}, CQI_l^{G2}.
    Execute the CMS and S-NOMA strategy based on [23] to find the resource allocation that maximizes the system ADR: $ADR_{max,l}^{CMS}$, $ADR_{max,l}^{S-NOMA}$, respectively.
    return: The best resource allocation for the $\mathcal{K}_l$ based on the highest achieved ADR value.
  end
  else if $M_l = 1$ **then**
    return: The best resource allocation to the $\mathcal{K}_l$ based on CMS.
  end
end

following multi-beam antenna modeling is based on Section 3.1.3.

The first step is the spatial users' clustering based on the $\mathcal{XYZ}$ and the HPBW of the available pre-computed beams. The beams are computed to cover the 90 ° of the service area (e.g., two beams of 45 °) symmetrically. After computing the angle of each user regarding the BS, we split the users according to the available beam angle and direction. After clustering the users into independent beams, we collect the CQI feedback from the link computation, s.t. the beam l parameters and the users in $\mathcal{K}_l$, with $\mathcal{K} = \mathcal{K}_{l=1} \cup \mathcal{K}_{l=2} \cup \cdots \cup \mathcal{K}_{l=L}$ and $CQI_{\mathrm{MG}'} = CQI_{l=1} \cup CQI_{l=2} \cup \cdots \cup CQI_{l=L}$.

The third step is to identify the best resource allocation strategy that maximizes the system ADR for each $\mathcal{K}_l$ subset in $\mathcal{K}$. The outcome of this step is s.t. the subset channel quality distribution CQI_l in $CQI_{\mathrm{MG}'}$ and the number of different CQI values reported in CQI_l defined as M_l. The following multicast resource allocation strategies and the proposed CMS and S-NOMA algorithms are based on Algorithms 1 and 3, respectively, optimized to maximize the system ADR.

Parameter	*Value*
Scenario type	UMi Street Canyon
Frequency	28 GHz
Numerology, μ	3
RBs bandwidth	1.44 MHz
Available RBs (multicast delivery), R_M	350
Noise spectral density	-174 dBm/Hz
BS/user height	10/1.5 m
Blocker radius, b_r	0.4 m
Blocker density	0.2 bl./m^2
Blockage attenuation	15 dB
Transmission power	20 dBm
BS antenna array per pre-computed beam	{1, 2, 4}x4
HPBW per pre-computed beam	{90, 45, 22.5} °
Gain per pre-computed beam	{0, 2.73, 5.57} dBi
User antenna array	1x4
User antenna gain	0 dBi
Large-scale fading models	[104]
Small-scale fading model	Jakes [102]
Dynamic line of sight	Yes
Mobility Model	Random directional [116]
Users' speed	0.8 m/s
AR constraints, $Th_{min}\|Th_{max}$	50|100 Mbps
VR constraints, $Th_{min}\|Th_{max}$	63|340 Mbps

Table 5.2: Simulation parameters.

5.2.1 Scenario description

This section presents the proposal assessment through simulation results with the ad-hoc 5G NR LLS [42] developed in Python. We consider a UMi Street Canyon open area with a single gNodeB (gNB) in the corner of a 70x70 meters grid operating at 28 GHz. We assume a sector cellular deployment of 90° covering the recreated grid with 350 RBs dedicated to delivering a multicast session. We assume 50 pedestrian users randomly distributed in the service area requesting the same multicast multimedia service for 60 seconds of simulation. The corresponding simulation parameters are adjusted to ensure a CQI feedback greater than or equal to 1 for all users. Table 5.2 summarizes the main simulation parameters.

The simulations include results for two XR multicast services: AR and VR. The throughput constraints are defined in Table 5.2. As the baseline, we consider a multicast RRM strategy based on the classic CMS with only one pre-computed beam (CMS-1x4) for the service delivery. Then, to evaluate the advantage of using

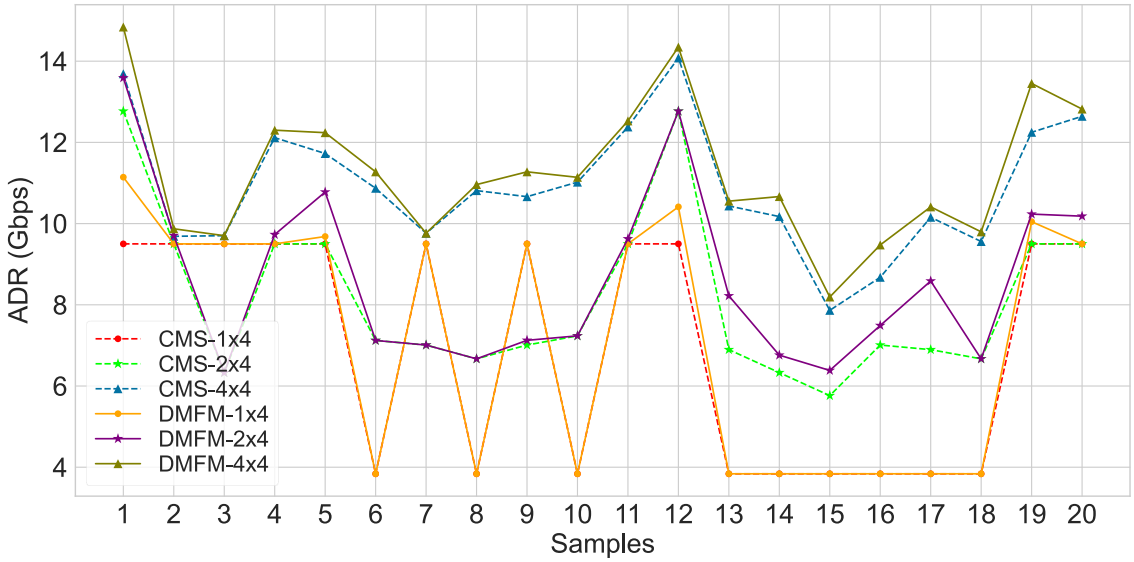


Figure 5.5: System ADR of the MG receiving the VR service over different samples.

fixed pre-computed beams, we analyze two and four beams with CMS as RRM strategy, respectively. Finally, we assess the advantage of combining the fixed pre-computed beams with a dynamic multicast strategy using opportunistic CMS or S-NOMA inside each beam (i.e., DMFM-1x4, DMFM-2x4, DMFM-4x4), maximizing the system ADR. Table 3.1 includes the HPBW and gain of the considered pre-computed beams according to equations 3.19 and 3.20.

5.2.2 Simulation Results

We quantify the performance of the evaluated multicast RRM strategies based on the system ADR, MDI, and PF as defined in [23]. The ADR allows us to compare how efficiently the resources are allocated regarding the total experienced capacity of the MG. Moreover, the PF and MDI metrics allow us to evaluate the fairness of the following resource allocation strategy. The outputs of the simulation process result from averaging 20 simulation runs, ensuring a 95 % confidence value.

Fig. 5.5 displays the achieved system ADR for each evaluated multicast RRM strategy over 20 different users' distributions in the grid requesting VR service. The results show how the highest ADR value is always achieved with the proposed DMFM-4x4, validating the improvement that supposes mixing pre-computed MIMO beams with S-NOMA to handle the users' spatial and channel quality diversity. As we can see, the second highest value is achieved with CMS-4x4. This result means that, for the evaluated conditions, the highest impact on the proposal's overall performance regarding CMS-1x4 lies in using four beams. Moreover, this simulation's dynamic subgrouping capability adds an extra ADR gain of around 3 %.

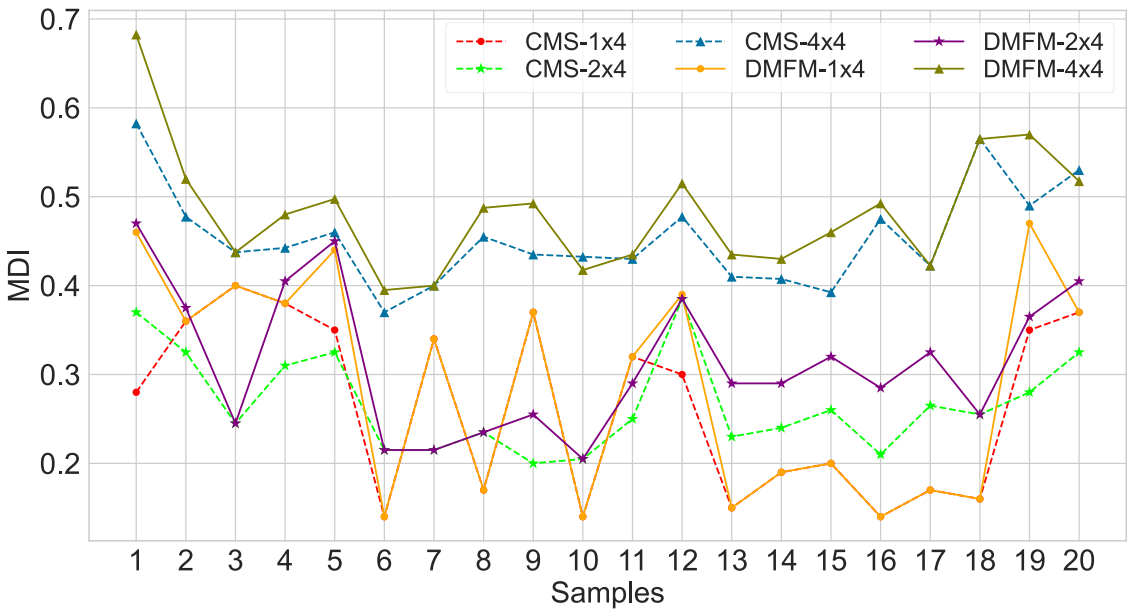


Figure 5.6: MDI of the MG receiving the VR service over different samples.

To assess the proposal's effectiveness from the resource allocation fairness point of view in delivering the VR service, we present Fig. 5.6 and 5.7. The results show a similar behavior regarding the ADR, where the highest MDI and PF values are always achieved with DMFM-4x4 and CMS-4x4. In all cases, the lowest results are achieved for the solutions that only consider a single multicast beam because they do not take advantage of the spatial diversity of the users and the beamforming gain. Let us highlight that the evaluated algorithms are optimized to maximize the system ADR, and a different approach based on multi-attribute decision-making, including these fairness metrics, can improve the performance. This strategy will be considered in further studies.

An equivalent analysis could be presented for the service AR, where the outcomes follow a similar behavior. In all the cases, the improvement added by the dynamic S-NOMA depends on the specific users' distribution regarding channel quality conditions. Moreover, the relation between the service Th constraints and the available RBs, defines the impact of using NOMA. It happens because if the multicast session has enough RBs to deliver the Th_{max} to all the users with CMS in each beam, it does not consider the NOMA approach.

To summarize the results for the two considered multicast services in terms of ADR, we present Fig. 5.8. The figure shows the ADR improvement of all the evaluated strategies regarding the traditional CMS-1x4. We include a service with a Th_{min} equal to 1 Mbps and no Th_{max} constraint to complement the above-presented results. We can see how the improvement of the proposal is higher as Th_{max} in-

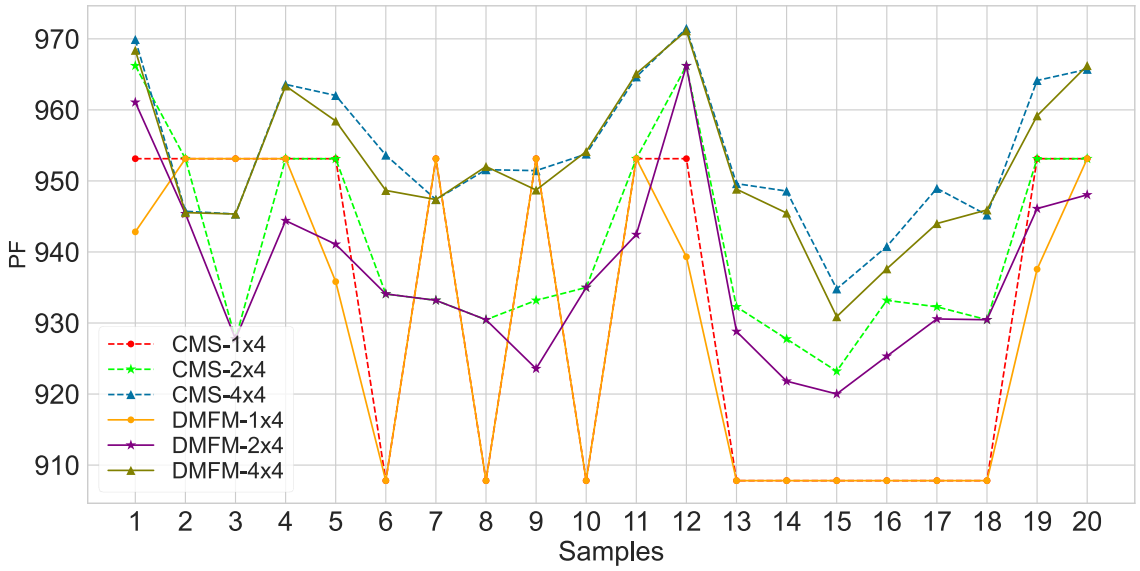


Figure 5.7: PF of the MG receiving the VR service over different samples.

creases for a fixed number of available RBs. The most significant impact on the performance increase is the capability of using dynamic S-NOMA. Even when the highest performance is always achieved with DMFM-4x4, it does not mean that increasing the number of beams and using NOMA is always the best solution. The final results will depend on the specific recreated conditions, and the best approach is to consider a dynamic multicast RRM tailored for such conditions and particular constraints.

5.2.3 Conclusions

In this section, we propose a multicasting strategy based on fixed pre-computed MIMO multi-beams and subgrouping based on NOMA for tackling specific multicast XR applications delivery in a restrained 5G and beyond environment with challenging throughput requirements. The proposed approach allows us to dynamically take advantage of the intrinsic users' spatial and channel quality diversity, maximizing specific QoS metrics. We evaluate our proposal through link-level simulations recreating a realistic 5G MBS use case. The assessment includes the critical metrics ADR, MDI, and PF. The results show the improvement added by combining pre-computed multi-beams with a dynamic multicasting strategy with at least 10 % and 17 % of ADR gain for AR and VR multicast service delivery, respectively. The extra gain added by the opportunistic use of subgrouping based on NOMA is around 3 %.

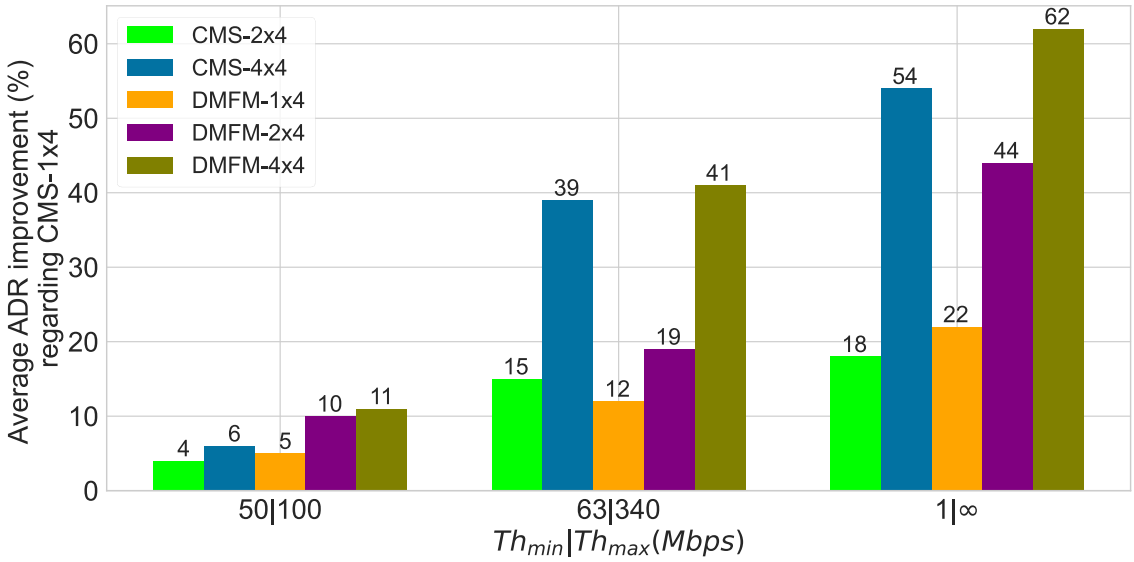


Figure 5.8: ADR improvement for three XR services.

Chapter 6

Machine Learning-based Low-Complexity Multicasting

In this chapter, we propose novel low-complexity multicast RRM strategies for dynamic access technique selection and resource allocation. We propose sub-optimal solutions that maximize the QoS but pay special attention to the CC as a KPI. The chapter delves into the complex implications of fast variations in the reception conditions of the MG members due to the users' mobility behaviors and the impact of mmWave propagation. The proposals are evaluated through numerical and link-level simulations recreating various network conditions. Moreover, it analyses the insertions in the O-RAN framework of ML-based multicasting RRM tasks. The proposed solutions, analysis, and conclusions respond to (**SO-5**).

The main contributions can be summarized as follows: (*i*) We address the CC associated with the dynamic multicast RRM strategies in 5G MBS use cases and highlight the implications of fast variations in the MG members' reception conditions; (*ii*) We evaluate the advantage of the dynamic selection of the multicast access technique based on Algorithm 5, and propose two alternative solutions based on multiclass classification ML algorithms with MLP and ETC. The proposed solutions allow for tailoring the radio resource allocation regarding users' distributions, multimedia service constraints, and network parameters; (*iii*) We propose the use of a K-Means clustering unsupervised ML approach for detecting and splitting group-oriented MGs based on the CQI values collected at the BS; (*iv*) We propose a multicast oriented trigger to avoid overrunning the entire algorithm subject to the temporal variations of the MG CQI distribution, reducing the induced latency over the time slot lattice; (*v*) Our proposed approaches allow addressing the trade-off between optimal network performance and CC by maximizing specific QoS parameters through non-optimal-solutions considerably reducing the CC of conventional exhaustive mechanism for specific 5G MBS use cases; (*vi*) We characterize the ML-based multicasting RRM insertion in the O-RAN framework under softwarized and intelligent vision.

6.1 Machine Learning based Radio Resource Management Soution

Multicasting to users with different channel conditions without a tailored resource allocation strategy can degrade the QoS of the whole MG and produce an unfair resource allocation [23]. In recent years, works such as [23–27] have addressed this challenging trade-off in the 5G context, with solutions mainly based on grouping the MG members according to their specific reception conditions and using a multi-rate MCS to deliver the service. These solutions take advantage of OMA and NOMA techniques, the spatial diversity of the users, and single/multi-antenna approaches. However, the previous works do not delve into the implications of fast variations in the reception conditions of the MG member for the CC of the RRM strategies. Significant variations in the channel quality conditions of the MG members imply the recalculation of the delivery solution. A non-optimized multicast RRM could exponentially increase CC and associated delay with a constant recalculation toward an optimal solution, which could not be tolerated in 5G MBS scenarios. Thus, novel multicast RRM solutions should consider the CC as a critical element in network performance [35].

Considering the previous analysis, this paper proposes novel low-complexity multicast RRM strategies for dynamic access technique selection and resource allocation subject to the 5G MBS paradigm. Our proposal is oriented to highlight and address the complexity associated with the multicast resource allocation process regarding the change ratio in the reception conditions of the MG members.

Our proposal can be divided into two main components: (i) a classifier for the selection of the best multicast access technique that better suits the network conditions, (ii) a K-means clustering for detecting and splitting group-oriented user distributions and a trigger algorithm that allows executing the entire algorithm dynamically.

6.1.1 Classifier for Multicast Access Technique Selection

To take advantage of the existing correlation among the specific service constraints, the network conditions, and the MG members' CQI distribution by selecting the best multicast access technique, we face the problem as a supervised ML multiclass classification [119]. Supervised ML classification algorithms are oriented to develop a learning model from a labeled dataset [120].

In our specific problem, we have a training dataset ($\mathcal{D}_{train}$) of S_{train} samples of the form (X_f, y), where $f \in \{1, 2, \ldots, F\}$ and $y \in \{1, 2, \ldots, C\}$, with F features and one label y of C classes. The goal of the training process is to obtain a learning model $\mathcal{H}$ such that $\mathcal{H}(X_f) = y$ for unseen samples of a testing dataset ($\mathcal{D}_{test}$) with S_{test} samples [119]. In this multiclass classification problem, we have a C equal to three classes, i.e., CMS, S-OMA, and S-NOMA, labeled 1, 2, and 3, respectively.

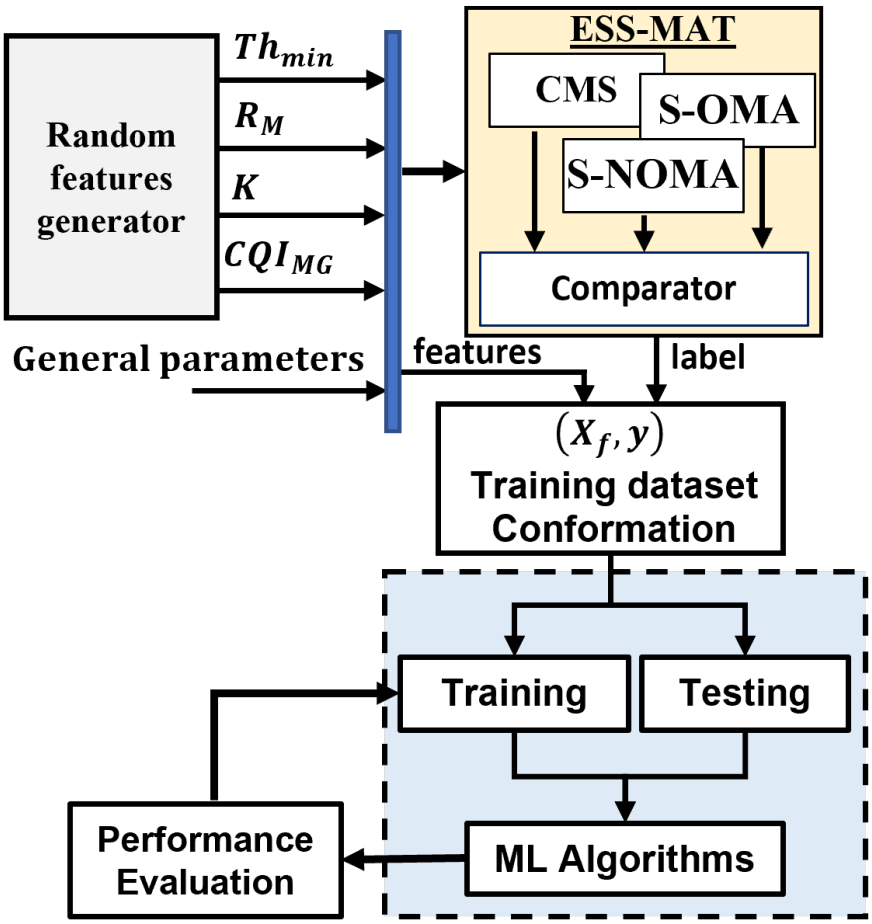


Figure 6.1: Diagram of the dataset creation and ML algorithm training process.

To obtain a set of features that represent the actual state of the network, s.t. our specific problem, we consider the users' CQI distribution vector U_{MG} divided by the K multicast group members U_{MG}^{K}, such as $\{u_{CQI=1}/K, u_{CQI=2}/K, \ldots, u_{CQI=15}/K\}$. We also included the features K, R_M and Th_{min}, for $F = 18$.

In Fig. 6.1, we present a diagram with the main components of the dataset creation and ML algorithms training. The complete process is carried out in a Python environment. For dataset creation, at each iteration, we generate a random value of K between 25 and 100 users, a Th_{min} of the multicast multimedia service between 1 Mbps and 60 Mbps, and a R_M value between 100 and 250 RBs (with a resolution of 50) enabled for the multicast session. Different Th_{min} values allow a wide range of eMBB services to be recreated, as seen in [47]. For the available RBs, we consider the 5G NR numerology $\mu = 2$ for an SCS equal to 60 kHz and $B_0 = 720$ kHz, as recommended in [121] for μWave and mmWave applications.

To generate the CQI_{MG}, we define four variables subject to the following bounds: $1 \leq min_{G1} \leq 15$, $min_{G1} \leq max_{G1} \leq 15$, $max_{G1} \leq min_{G2} \leq 15$ and $min_{G2} \leq$

$max_{G2} \leq 15$. We randomly generate these variables as the CQI limits of possible group-oriented CQI distributions. Then, we randomly generate from 30 % to 70 % of the K CQI values, between min_{G1} and max_{G1}, and the remainder between min_{G2} and max_{G2}. This approach enables us to recreate group-oriented CQI distribution with different percentages of the users as low and high-channel quality. Under these constraints, we generate a dataset of 20000 samples.

We opt for a numerically generated dataset because it enables us to obtain extensive data, covering a wide state space and directly oriented to our specific problem. Each sample created by the random feature generator plus other extra network parameters are used to run the ESS-MAT algorithm for the multicast access technique selection and to add the corresponding label to each feature set X_f in the dataset. The resulting dataset has 9680 samples of S-NOMA (3), 4545 samples of S-OMA (2), and 4438 samples of CMS (1). The remaining samples got the label 0 and were deleted from the dataset because it means that, for these samples, the randomly generated CQI distribution and R_M values can not satisfy the Th_{min} constraint, with any of the available access techniques. To lead with the unbalanced dataset, we use the Python library Synthetic Minority Oversampling Technique (SMOTE) [122] that oversamples the minority classes in the dataset.

The training process of the ML algorithms includes data normalization, data train/test split, grid-search/cross-validation, and, finally, the evaluation of the algorithms through specific error metrics [123]. We apply the Min-Max scaling method for normalization, transforming all features into the range $[0, 1]$ [123, 124]. The dataset is split into 80 % and 20 % for the training and testing, respectively. We apply Grid-search [125] and k-fold (with $k = 5$) cross-validation to evaluate multiple combinations of the hyperparameters associated with each ML algorithm and use the F1 score [123] as the tuning criterion.

To solve the problem, we evaluate in the above-described training process the performance of several scikit-learn native multiclass classifiers [123]. These estimators have multi-learning support as described in [119]. After multiple iterations, the best multiclass classifiers learning models $\mathcal{H}$ are obtained with MLP [126] and ETC [127] as $\mathcal{H}_{MLP}$ and $\mathcal{H}_{ETC}$, respectively. The ETC algorithm builds an ensemble of unpruned decision or regression trees according to the classical top-down procedure. After the tuning process, the number of estimators to form the ensemble is set to 450, with 100 as the maximum depth of the trees. The number of features to consider for the best split is set as the square root of the total number of features in the dataset. In the case of the artificial neural network (ANN) algorithm MLP, after the tuning process, the selected activation function is Rectified linear activation function (ReLU) [128], Adaptive Moment Estimation (Adam) [129] as the optimization algorithm, a learning rate of 0.03, and one hidden layer of 150 neurons.

To assess the effectiveness of the proposed classifiers, we consider the metrics accuracy and F1 score, including a 3x3 confusion matrix analysis [123]. The accuracy

Algorithm 7: MLP-MAT—ETC-MAT

Input: $\mathcal{H}_{MLP}|\mathcal{H}_{ETC}, CQI_{MG}, R_M, Th_{min}, K$
Output: Multicast access technique selection
1: From CQI_{MG} **and** K **determine:**
$U_{MG}^{K} = \{u_{CQI=1}/K, u_{CQI=2}/K, \ldots, u_{CQI=15}/K\}$
2: Conform X_f **as** $\{U_{MG}^{K}, R_M, Th_{min}, K\}$
3: Predict y **with** $\mathcal{H}_{MLP}|\mathcal{H}_{ETC}$ **and** X_f**:**
$y = \mathcal{H}_{MLP}.predict(X_f)|\mathcal{H}_{ETC}.predict(X_f)$
5: Select the multicast access technique
if $y = 1$ **then**
  return: CMS
else if $y = 2$ **then**
  return: S-OMA
else if $y = 3$ **then**
  return: S-NOMA
end

of predicting each one of the classes can be defined as

$$Accuracy = \frac{CP}{T}, \tag{6.1}$$

where $CP = TP + TN$ is the percent of correct classified predictions, including the true positive (TP) and true negative (TN). $T = TP + TN + FP + FN$ is the total number of predictions, including the false positive (FP) and false negative (TN). The TP, FP, TN, and FN are computed based on the correct and incorrect multicast access technique selection for a specific feature set X_f regarding the actual y label obtained with the optimal ESS-MAT algorithm. The model's accuracy is the mean of the accuracy of each class.

The F1 score is the harmonic mean of the metrics precision and recall, giving a balanced measurement of these metrics [123]. This expression is preferable to evaluate unbalanced datasets with an individual evaluation of multiple classes. The precision, recall, and F1 score metrics are defined in [123].

Algorithm 7 presents the description of the proposed algorithms MLP-Multicast Access Technique Selection (MLP-MAT) and ETC-Multicast Access Technique Selection (ETC-MAT). The first two steps aim to conform the specific feature set X_f regarding the network conditions. Then, in step three, is executed the multiclass classification inference.

Regarding the CC analysis presented in Section 3.2, let us define the CC of the proposed dynamic selection algorithms as CC_{DSA}. Then we can define the CC of

the proposed solutions, $CC_{OE-MAT}, CC_{MLP-MAT}$ and $CC_{ETC-MAT}$ as

$$CC = \begin{cases} CC_{DSA}, & \text{if } CMS \\ CC_{DSA} + \mathcal{O}(M * R_M), & \text{if } S-OMA \\ CC_{DSA} + \mathcal{O}(M * l), & \text{if } S-NOMA \end{cases} \tag{6.2}$$

where CC_{DSA} equals the CC of the proposed solutions to find the best multicast access technique. In the case of OE-MAT, we can define CC_{DSA} as $\mathcal{O}(l)$ because the solutions have to iterate over the l IL values to evaluate the performance equations in Algorithm 5. Hence, using OE-MAT, suppose a CC equal to $\mathcal{O}(l)$ plus the specific CC of the selected technique. As can be seen, the improvement in terms of CC of OE-MAT concerning ESS-MAT in (3.28) reaches its maximum value when the selected access technique is CMS because ESS-MAT still has to compute the three multicast access techniques.

6.1.2 Subgrouping and Trigger based on K-Means Clustering

As presented in the above Section 3.2, the MG subgrouping process significantly contributes to the CC of these techniques.

Aimed to split CQI_{MG}, we based our solution on the ML K-Means [130] algorithm. The K-Means is a simple unsupervised ML algorithm capable of clustering unlabeled datasets in a few iterations [123]. Our proposed solution applies a partitional clustering of CQI_{MG}, dividing the CQI values into two non-overlapping groups of CQI values labeled as 0 for $G1$ and 1 for $G2$. If all the CQI values are labeled as 0, all the reported CQI in the MG have the same value. Let us assume our K-Means learning model as $\mathcal{H}_{\text{km}}$, with $\mathcal{H}_{\text{km}}(X_f^{\text{km}}) = y_{\text{km}}$, where X_f^{km} will be the $M \leq 15$ unique CQI values in CQI_{MG}, and y_{km} the M labels (0 or 1) for each unique CQI in CQI_{MG}. Algorithm 8 shows a pseudocode that describes the functions of the proposed clustering strategy.

From the K-Means clustering algorithm, we conceive an MG-oriented trigger intending only to run the algorithm OE-MAT, MLP-MAT, or ETC-MAT and recompute the resource allocation strategy if either $CQI_{min} = CQI_{min}^{G1}$ or CQI_{min}^{G2} change from the RRM time slot $t-1$ to t. The trigger can be defined as

$$trigger_t = \begin{cases} 1, & \text{if } flag_t \neq flag_{t-1} \\ 1, & \text{if } CQI_{min,t} \neq CQI_{min,t-1} \\ 1, & \text{if } CQI_{min,t}^{G2} \neq CQI_{min,t-1}^{G2} \\ 0, & \text{otherwise} \end{cases} \tag{6.3}$$

where $trigger = 1$ means that RRM must recompute the multicast access technique selection and the resource allocation; otherwise, it will not. In Fig. 6.2, we present a

Algorithm 8: K-Means clustering

Input: CQI_{MG}
Output: $flag, CQI_{min} = \{CQI_{min}^{G1}, CQI_{min}^{G2}\}, CQI_{\mathrm{MG}}^{G1}, CQI_{\mathrm{MG}}^{G2}$
1: Conform $X_f^{\mathbf{km}}$ as a sorted array with just the unique values in CQI_{MG}:
$X_f^{\mathrm{km}} = sort(unique(CQI_{MG}))$
2: Predict $y_{\mathbf{km}}$ with $\mathcal{H}_{\mathbf{km}}$ and $X_f^{\mathbf{km}}$:
$y_{\mathrm{km}} = \mathcal{H}_{\mathrm{km}}.predict(X_f^{\mathrm{km}})$
3: If all the $y_{\mathbf{km}}$ are equal 0, $flag = 0$:
if $y_{km} = 0$ **then**
  return: $flag = 0, CQI_{min} = min(CQI_{MG})$
else if $y_{km} \neq 0$ **then**
  then: $flag = 1$
  4: Find CQI_{min}^{G2} as the minimum CQI value in $X_f^{\mathbf{km}}$ labeled as 1:
  $CQI_{min}^{G2} = min(X_f^{\mathrm{km}}|_{y_{\mathrm{km}}=1})$
  5: Conform $CQI_{\mathbf{MG}}^{G1}$ and $CQI_{\mathbf{MG}}^{G2}$ from $CQI_{MG}, CQI_{\mathbf{min}}^{G1}, CQI_{\mathbf{min}}^{G2}$:
  return: $flag, CQI_{min}^{G1}, CQI_{min}^{G2}, CQI_{\mathrm{MG}}^{G1}, CQI_{\mathrm{MG}}^{G2}$
end

diagram with the interconnection among the proposed K-Means clustering, the trigger, and the dynamic multicast access technique solutions: OE-MAT, MLP-MAT, or ETC-MAT.

6.2 Results and Discussion

The evaluation is conducted through two scenarios based on link-level simulations and an extensive simulation set covering a wide range of configurations. Metrics to assess the ML classification task, the QoS, and the CC are the common thread of the results and discussion of this section. In the following, we will refer to the proposed solutions using the flow diagram presented in Fig. 6.2 as OE-MAT, MLP-MAT, and OE-MAT. Only to highlight the advantage of the proposed trigger, we include the index "kt" to differentiate the solution with the trigger (ktOE-MAT, ktMLP-MAT, and ktETC-MAT) and without the trigger (OE-MAT, MLP-MAT, and ETC-MAT).

6.2.1 Scenarios' Description

The two recreated validation scenarios are based on an ad-hoc LLS developed in Python [42], oriented to recreate 5G-MBS use cases. The RRM strategy must ensure delivering the service correctly to 100 % of the users. We consider the 5G NR numerology $\mu = 2$ for an SCS equal to 60 kHz and $B_0 = 720$ kHz as recommended

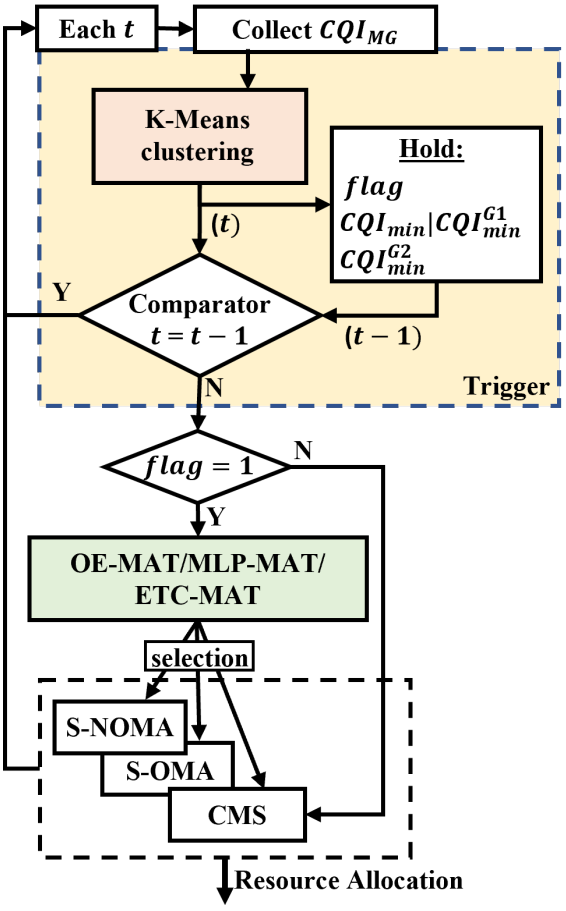


Figure 6.2: Flow diagram of the proposed K-Means clustering, the trigger, and the dynamic multicast access technique solutions.

in [121] compatible with μWave and mmWave applications. Table 6.1 summarizes the main simulation parameters.

Scenario A) We consider a UMi Street Canyon open area with a single gNB in the middle of a grid of 60x60 meters operating at 28 GHz. The goal of this scenario is to generate 1000 samples of uncorrelated network configurations where, for each sample, we generate a random number of stationary K users between 25 and 100, a Th_{min} of the multicast multimedia service between 1 Mbps and 60 Mbps, and a R_M value between 100 and 250 RBs (with a resolution of 50). We consider each sample a network snapshot where the simulator's output will be the instantaneous CQI_{MG} and network parameters. With the simulator's outputs, we generate a $\mathcal{D}_{test}$ with 1000 samples of unseen data for the proposed dynamic multicast access technique solutions.

Scenario B) We consider a UMi Street Canyon open area with a single gNB in the middle of a grid. The BS provides multicast multimedia service with a Th_{min} of

Parameter	*Scenario A*	*Scenario B*
Scenario type	UMi Street Canyon	UMi Street Canyon
Frequency (GHz)	28	2, 4, 6, 16, 28, 39, 50
Numerology μ	2	2
RBs bandwidth (kHz)	720	720
Available RBs, R_M	100-250	200
BS/user height (m)	10/1.5	10/1.5
Transmission power (dBm)	10	10
BS/user antenna gain (dB)	10/0	10/0
Large-scale fading models	[131]	[131]
Small-scale fading model	Jakes [102]	Jakes [102]
Dynamic line of sight	Yes	Yes
Antenna	Sectorial (120°)	Sectorial (120°)
Mobility Models	Stationary	Random Directional

Table 6.1: Simulation parameters.

50 Mbps and 200 available RBs. For each simulation, we consider 50 MG members in the service area of the BS, with a random directional mobility model [132], at the same constant speed during 60 seconds of simulation (with 100 ms of resolution). In this scenario, we test different users' speeds from 0.5 to 3 mps at a resolution of 0.5 and from 3 to 30 mps at a resolution of 3 for 15 different speeds. The selected speed range is oriented to evaluate typical pedestrian and vehicular mobilities. We also consider the scenario for different μWave and mmWave frequencies 2, 4, 6, 16, 28, 39, and 50 GHz. We adjust the grid size for each frequency to ensure collecting CQI values from 1 to 15 at each run. For each combination of velocity and frequency 15×7, we generate 20 random simulation runs.

6.2.2 Multiclass Classifier Assessment

To validate the performance of OE-MAT, MLP-MAT, and ETC-MAT and select the multicast access technique that better suits the specific conditions of the network, we use the dataset $\mathcal{D}_{test}$ from *Scenario A*. We employ the optimal solution found by ESS as a reference result. The validation is based on the accuracy, F1 score, precision, and recall metrics. Table 6.2 summarizes the evaluation results.

As shown in Table 6.2, the best multiclass classification performance in terms of

OE-MAT/MLP-MAT/ETC-MAT				
Class	precision (%)	recall (%)	F1 (%)	support
CMS	93/95/95	100/97/95	97/96/95	373
S-OMA	98/99/98	100/96/93	99/97/95	356
S-NOMA	100/92/94	87/94/94	93/93/91	271
mean F1(%)		96.4/95.7/94.1		
Accuracy(%)		96.0/96.0/94.0		

Table 6.2: Multiclass classification task evaluation.

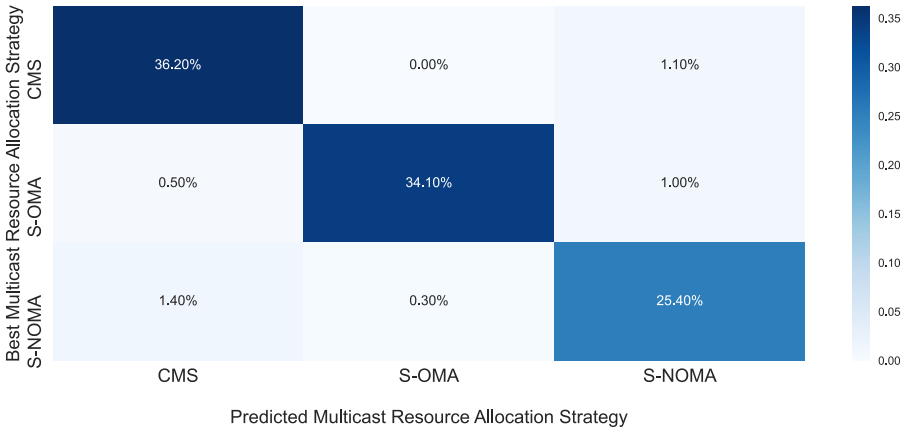


Figure 6.3: Confusion Matrix of the MLP multiclass classification model.

mean F1 score and accuracy is achieved with the heuristic solution OE-MAT. The worst result with OE-MAT is achieved for the S-NOMA samples with an F1 score of 93 %. This degradation is due to the 87 % of recall, meaning S-NOMA samples are misclassified. In contrast, the precision classifying S-NOMA is 100 %, which means that all the samples classified as S-NOMA are, in fact, S-NOMA. The best ML-based solution is achieved with MLP-MAT with the same global accuracy of 96 % that OE-MAT and a difference of less than 1 % of the mean F1 score.

For the three solutions, the worst performance is classifying S-NOMA samples. In the case of the ML solutions, one of the reasons for this behavior could be the unbalanced dataset during the training process, where S-OMA doubled the number of samples concerning S-NOMA. To put light on the meaning of the values presented in Table 6.2, in Fig. 6.3, we present the confusion matrix of the MLP multiclass classification model. If we analyze its confusion matrix, the 92 % of precision classifying S-NOMA samples is because 1.4 and 0.3 % of the S-NOMA samples were wrongly classified as CMS and S-OMA, respectively. Moreover, 1.1 and 1 % of the CMS and S-OMA samples were classified as S-NOMA, reducing the recall to 94 percent.

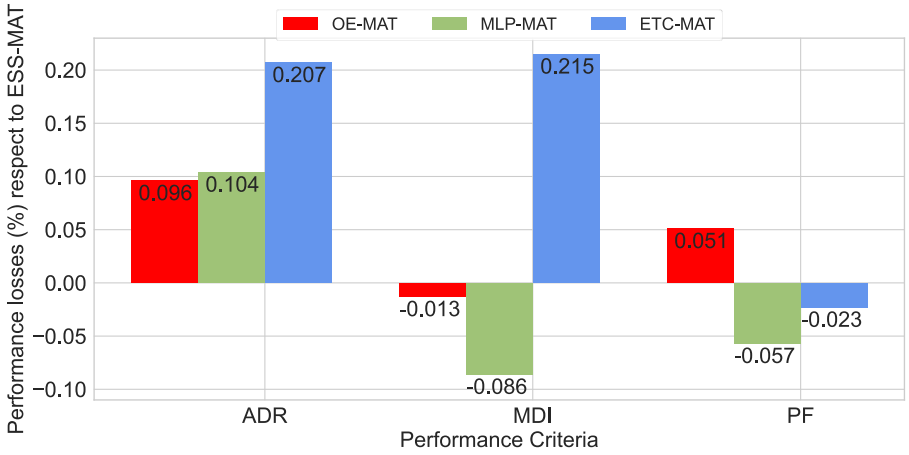


Figure 6.4: QoS performance losses for ESS-MAT.

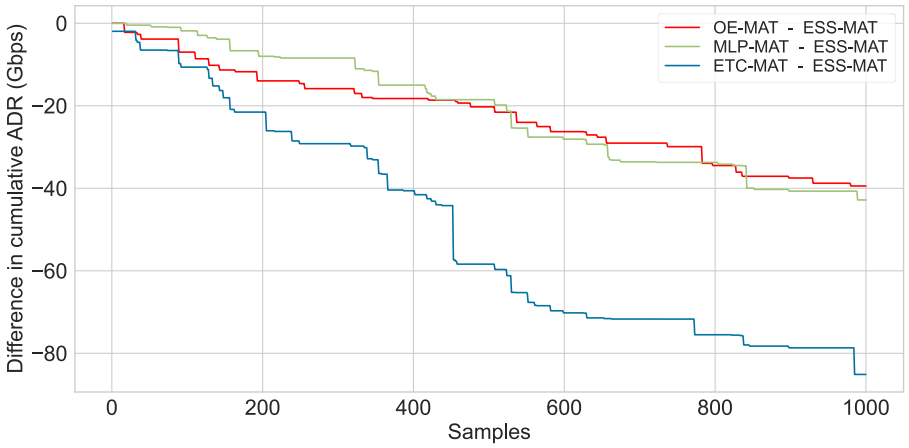


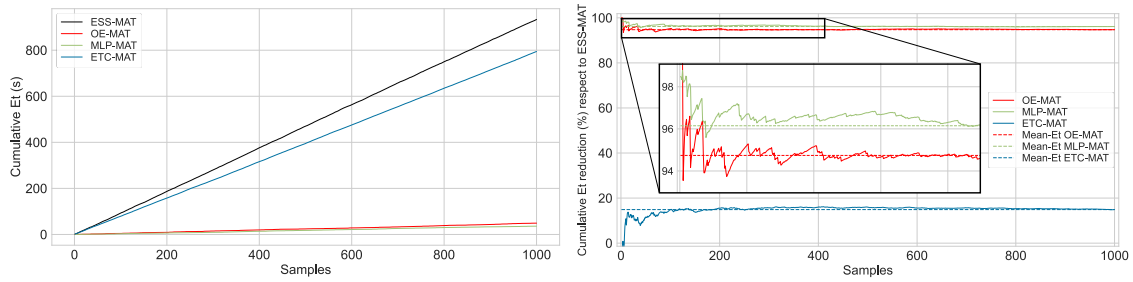
Figure 6.5: Difference in cumulative ADR of the proposed solutions and ESS-MAT.

In Fig. 6.3, the 36.20 %, 34.10 %, and 25.4 % represent the percent of samples from the evaluated dataset (1000 unseen samples) that were correctly classified as CMS, S-OMA, and S-NOMA, respectively. The remainder of the values represent the percentage of misclassifications.

The multiclass classification assessment shows that the three proposed solutions are considerably effective, with accuracy and F1 score values higher than 94 %. In the case of the ML solutions, the results could be improved by increasing the dataset and balancing the classes during dataset creation.

6.2.3 QoS Assessment

The performance of the proposed sub-optimal OE-MAT, MLP-MAT, and ETC-MAT is evaluated regarding the metrics system ADR, MDI, and PF. In every case, their



(a) (b)

Figure 6.6: (a) Cumulative *Et* and (b) cumulative *Et* reduction in percentage regarding ESS-MAT.

performance is compared with the optimal solution ESS-MAT. The validation is based on *Scenario A*, measuring the average performance losses of the proposals concerning the ESS-MAT for the 1000 samples. The results are shown in Fig. 6.4, evidencing that the QoS performance of the proposed solutions and ESS-MAT is almost the same. Specifically, the performance losses of the proposed solutions for ESS-MAT for the three QoS metrics are lower than 0.21 %. The ESS-MAT solution optimizes the system ADR, and OE-MAT and MLP-MAT have a minor ADR performance degradation on the order of 0.1 %. Moreover, OE-MAT and MLP-MAT outperform the ESS-MAT regarding MDI minimally (i.e., a negative value in Fig. 6.4). It happens because even when the resource allocation of the proposed algorithms is optimized for the ADR, the miss-classification during the selection of the best multicast access technique tends to favor fairer multicast strategies than ESS-MAT. The same happens for the metric PF with MLP-MAT and ETC-MAT. To help visualize the evolution of these metrics along the 1000 samples, Fig. 6.5 shows the difference in cumulative ADR between the proposed solutions and ESS-MAT. In this figure, for the worst case, the cumulative difference over 1000 samples equals 85.1 Gbps. This value, with ETC-MAT, represents a mean system ADR difference of 85.1 Mbps (85.1 Gbps / 1000), the 0.207 % respect to the mean ADR of ESS-MAT over the 1000 samples, equal to 41.138 Gbps. An equivalent analysis can be made for OE-MAT and MLP-MAT, where the mean ADR difference over the 1000 samples equals 39.44 Mbps and 42.8 Mbps.

6.2.4 Computational Complexity Assessment

After validating the performance of the proposed solutions in terms of QoS, let us evaluate their CC improvements. As defined above, we assess the CC in terms of the *Et* metric. The simulations are launched in a server with 1 x Barebone Asus ESC4000 G4 1+1 1600 W RPSU 4 x GPU 2 x Processor Intel Xeon Gold 6238 2.1 GHz 22C 140 W 8 x Samsung Memory DDR4 2933 MHz 32 GB.

Continuing with *Scenario A*, Fig. 6.6 (a) presents the cumulative *Et* of ESS-MAT,

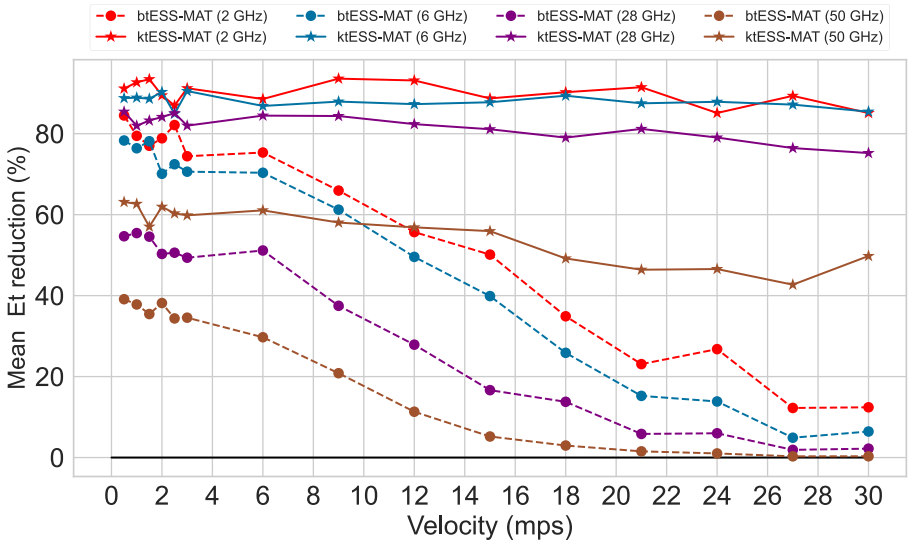


Figure 6.7: Mean *Et* reduction regarding ESS-MAT of btESS-MAT, ktESS-MAT.

OE-MAT, ETC-MAT, and MLP-MAT over the 1000 samples. As we can see, the lower *Et* and, consequently, the lower cumulative *Et* are attained with MLP-MAT and OE-MAT. Fig. 6.6 (b) helps to visualize the cumulative *Et* reduction of the analyzed solutions regarding ESS-MAT. The MLP-MAT reduces the mean *Et* concerning ESS-MAT over 1000 samples by more than 96 %. The solution based on the ETC learning model takes longer to converge, with an improvement of around 15 %. From this result, we can validate the advantage of using a non-optimal solution for the MG splitting and the multicast access technique selection regarding the CC. The proposed algorithms reduce the intrinsic CC associated with this dynamic process in 5G MBS use cases.

To assess the performance of the proposed MG-oriented trigger, we use *Scenario B* to evaluate the impact of the CQI change ratio increment due to the propagation frequency and user velocity. First, we add the proposed trigger to the ESS-MAT and compare it with the performance of btESS-MAT (ESS-MAT with the 20 % trigger) presented in Section 3.2. In this simulation, we find the mean *Et* of each solution over 60 seconds and 20 simulation runs for each possible combination of users' speed and frequency.

Fig. 6.7 illustrates the extra *Et* reduction added by the proposed MG-oriented trigger with ESS-MAT (ktESS-MAT) concerning ESS-MAT and btESS-MAT. For the mmWave cases, the improvement is always bigger than 20 %. This improvement increases with the velocity in all cases. The most important outcome of this simulation is the change in the behavior introduced by the proposed trigger regarding the velocity. We can see how the results for ktESS-MAT have a minimal variation with the velocity increase, which means that the proposed strategy for the trigger is less affected by the increment in the general CQI changing ratio of the MG.

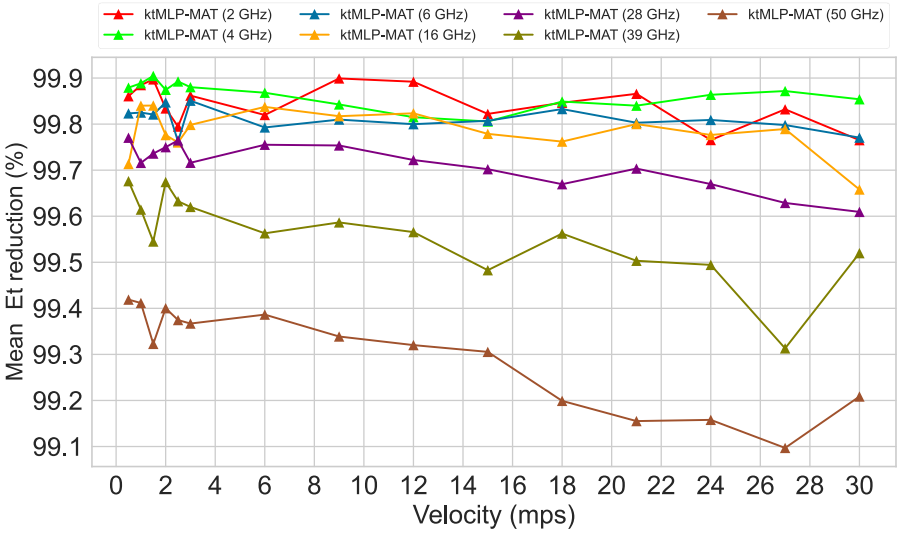


Figure 6.8: Mean *Et* reduction regarding ESS-MAT of ktMLP-MAT.

Fig. 6.8 shows the *Et* reduction of the entire proposal, including the trigger, the K-Means clustering, and the MLP-based multicast access technique solution. The improvement concerning ESS-MAT is around 99 % for all cases. We can see how the *Et* reduction increases at around 3 % concerning the result in Fig. 6.6. Such outcomes validate the performance of the proposed solution in terms of CC with an approach that is less sensitive to high CQI changing ratios. This is important in handling high mobility use cases and mmWave scenarios. Moreover, the performance of the solution in terms of QoS is almost the same as the optimal ESS-MAT solution. The validation results show a worthy trade-off between a non-optimal QoS performance less than 0.25 % concerning the optimal solution and an improvement of 99 % of CC reduction.

6.2.5 Conclusions

This Section is oriented to contextualize and address the complexity associated with the multicast RRM process regarding the fast variations in the reception conditions of the MG members subject to the 5G MBS paradigm. We propose a dynamic multicast access technique selection and resource allocation strategy based on an MG-oriented trigger, a K-Means clustering for detecting and splitting group-oriented user distributions, and a classifier for selecting the best multicast access technique. For the dynamic selection of the multicast access technique that better fits the specific reception conditions of the MG members, we propose a novel approach based on ML-multiclass classification with MLP and ETC algorithms. The proposal results show a worthy trade-off between a non-optimal QoS performance less than 0.25 % concerning the optimal solution and an improvement of 99 % of CC reduction. The

proposed dynamic multicast access technique solution and the following strategies can contribute to the envisaged 5G MBS use cases. It helps reduce the complexity of RRM and the induced latency in communication. Our approach effectively handled the trade-off between the multicasting gain and the multiuser diversity, highlighting the implications of the existing correlation between the users' velocity, the propagation frequency, and the variations in the channel conditions. The proposed solutions allow tailoring the multicast access technique and radio resource allocation regarding users' distributions, multimedia service constraints, and network parameters.

6.3 O-RAN Insertion of Machine Learning-Based Multicasting Solutions

Based on the above-presented results and remarks, and the analysis presented in Chapters 1 and 2, we can affirm that the conception of the future networks (i.e., 6G) must incorporate intelligence as an endogenous characteristic to manage ultra-dense HetNet where scalability is critical [133]. In such context, the O-RAN framework enables an effective ML closed-loop workflow to dynamically conduct several optimization actions directly impacting QoS, user perception, and the E2E complexity of such envisioned systems [6].

Recent studies surveyed O-RAN [6, 134, 135], the RRM, and multicast use cases [8, 136] separately. However, little attention has been devoted to discussing the intersection of MBS, RRM solutions, and O-RAN in the 6G context. We address this gap and propose the insertion of ML-based multicasting RRM solutions in the 6G O-RAN under a softwarized and intelligent vision. The major contributions of this Section can be summarized as follows: *i)* A taxonomy of the distributed O-RAN over the envisioned 6G networks; *ii)* Characterization of ML-based multicasting RRM tasks as potential MBS use cases and their insertion in the 6G O-RAN framework.

6.3.1 The 6G O-RAN Framework

O-RAN is an ML native framework that uses virtualized and disaggregated elements to conduct dynamic tasks [134]. The slicing paradigm aids this architecture, mapping multiple service types (e.g., services 1-L) into numerous NSs (e.g., slice 1-M) to manage differentiated traffic and ensure the defined SLA. As presented in Section 2.3, O-RAN disaggregates the BS functionalities into O-CU,O-DU, and O-RU. The logical split allows these functional units to be flexibly deployed at different network locations and hardware platforms.

Two critical elements in the O-RAN framework are the Non-Real Time RAN Intelligent Controller (Non-RT RIC) and the Near-Real Time RAN Intelligent Controller (Near-RT RIC). The former is a Service Management and Orchestration (SMO) element. It implements optimization actions through microservices termed rApps on a time scale superior to 1 s. It trains and updates ML models that will be

executed by structures nearer to the end-user (e.g., Near-RT RIC). Moreover, the Non-RT RIC realizes long-term monitoring via the *O1* interface and sends policies to the Near-RT RIC to drive E2E SLA assurance. The latter executes ML tasks through microservices termed xApps in control loops between 10 ms-1 s. Additionally, it conducts real-time monitoring tasks (*E2* interface) to detect whether the performance is out of the target KPIs.

6G involves an ultra-dense heterogeneous environment with massive data. The input/output ML design and how often this data is collected depend on specific optimization problems. Data preparation, including pre-processing, cleaning, and transformations, is an efficient mechanism to homogenize the ML algorithm's inputs [137]. Following the O-RAN specifications to prevent poor RAN performance or outages, the ML model that has not been previously trained and validated offline cannot be deployed [6]. Once these phases are successfully finished, the resulting trained model is published in the ML Catalog. The ML Catalog also includes under which specific conditions the ML-trained model delivers the best performance. Then, the ML model is deployed into the inference host and fed with data from the environment to execute specific tasks online. During the implementation phase, the ML model previously trained can be fine-tuned and updated based on architectural changes or inefficiencies detected. Continuous operation in the 6G dynamic environment is crucial to improve the previously trained ML models online.

The O-RAN alliance defines five scenarios detailing the allocation of ML training and inference entities [6, 97]:

- *Scenario 1.1*: Non-RT RIC performs ML training and inference functions.
- *Scenario 1.2*: Non-RT RIC performs ML training, and Near-RT RIC conducts the ML inference.
- *Scenario 1.3*: SMO performs ML training, and Non-RT RIC executes ML inference functionalities.
- *Scenario 1.4*: ML training is performed due to the collaboration between Non-RT RIC and Near-RT RIC. The ML inference task is executed by the Near-RT RIC.
- *Scenario 1.5*: Non-RT RIC performs ML training, and O-CUs/O-DUs assume ML inference (for further studies).

As the 6G requirements comprise very low latency, high speeds, and massive heterogeneous data, there are certain cases where the ML execution cannot be carried out within the timescale supported by the RAN Intelligent Controllers (RICs). Collecting sensitive data through *O1* and *E2* can negatively impact latency, overhead, and privacy. Then, elements nearer to the end-user must be capable of executing various ML inference tasks in a distributed manner at a timescale below 10 ms (e.g., *Scenario 1.5*). Moreover, as future extensions of the O-RAN specifications, the

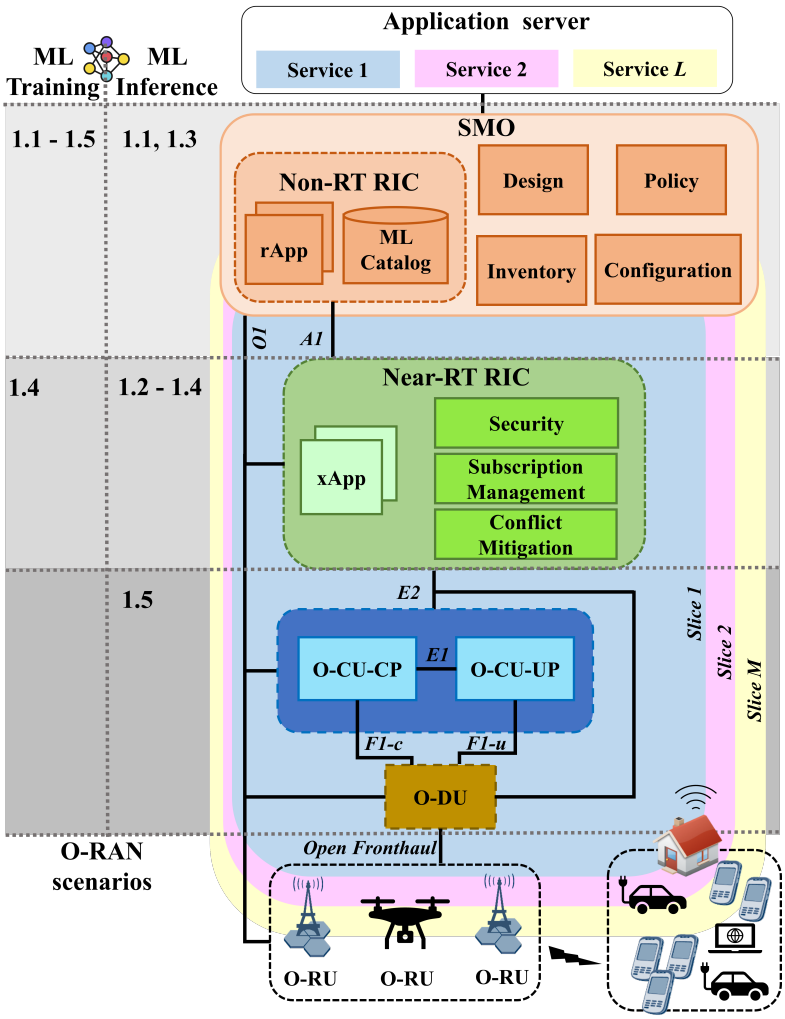


Figure 6.9: The O-RAN framework (control plane (CP); user plane (UP)).

Near-RT RIC could assume ML training in cases where extensive data arrives from the *E2* interface, which may not be available via *O1*. Thus, new scenarios enabling the ML training and inference in O-CUs/O-DUs will be fundamental for the future 6G deployment, being identified as one of the primary challenges of O-RAN [134, 138].

Figure 6.9 details the described O-RAN components, their interfaces, and scenarios deployment. Although not currently supported in O-RAN, the ML inference must also be considered in the O-RU for assignments like beamforming. Consequently, the nearer entities to the end user must be dotted with enough resources and open interfaces to efficiently assume ML tasks from different vendors. The power consumption, security, computational complexity, and storage capacity are critical concerns for future 6G ML solutions, sensibly impacting deployment costs and algorithm performance.

ML task	ML type	Scenario	Input	Output
MG CQI-based subgrouping for multi-rate multicasting [41]	Supervised, unsupervised	1.2, 1.4	Users' channel information, service requirements, number of layers	MG clustering for group-oriented multicasting
Dynamic multicast access technique selection [41]	Supervised, DRL	1.2, 1.4, 1.5	Users' channel information, service requirements, BS parameters	Optimal multicast access technique and resource allocation
Power allocation for multi-layer NOMA multicasting [139]	Supervised, DRL	1.5	Multicast subgroups information, BS parameters, service requirements, injection levels	Power assignment for each layer
MG spatial subgrouping for multibeam multicasting [27]	Supervised, unsupervised	1.2, 1.5	Users' spatial information, number of users, BS parameters, cell geometry, service requirements	MG clustering and number of beams
Group-oriented multibeam management [118]	Supervised, DRL	1.5	Users' channel/spatial information, BS parameters, cell geometry, number of beams, service requirements	Beams direction, width, and power allocation
MGs association under cell-free beamforming [37]	Supervised, DRL	1.5	Users' channel/spatial information, multi-cell geometries, network load, service requirements	BS clustering for MG service delivery
MGs-oriented RIS deployment [34]	Supervised	1.2, 1.4, 1.5	Users' channel/spatial information, number of MGs, set of possible RIS locations/orientations	RIS position and orientation per MG
D2DM cluster header selection [54]	Supervised, DRL	1.5	Users' channel/spatial and remaining battery information, social interests, BS parameters, service requirements	Selected cluster headers in the MG
BS/NS selection for shared MBS traffic delivery [61, 140]	Supervised, DRL	1.4	Users' channel and SLA information, BS/NS parameters, network load, service petitions and requirements	Optimal BS/NS selection
Slicing load balancing for shared MBS traffic delivery [61]	DRL	1.5	Users' channel and SLA information, BS/NS parameters, network load, service requirements	Slices resources adjustment

Table 6.3: ML-based multicasting RRM deployment in the O-RAN framework.

6.3.2 Multicasting into the O-RAN Framework

The expected MBS solutions can benefit from the O-RAN framework by exploiting enriched environment knowledge and interactions among the architectural elements. O-RAN introduces multiple potentialities for heterogeneous systems, adapting to various use cases, actions, and state spaces. Through O-RAN, multiple multicast tasks can be dynamically disaggregated and executed based on ML and the NS paradigm. ML-aided multicasting RRM can handle complex environments with high mobility, CQI variation, and challenging service constraints where resource optimization is critical with an adequate CC balance. Additionally, the ML life-cycle management guarantees constant monitoring and evaluation of the environment, updating the algorithm if necessary to preserve the long-term system performance.

Table 6.3 identifies some primary tasks required for optimum multicasting data delivery and efficient RRM in mmWave and sub-THz frequency bands [15]. The different functions have been related to the possible ML types [141] for performing these tasks, general inputs/outputs data, and the corresponding O-RAN scenario as defined in [97]. This division and task identification is critical for successfully integrating MBS services in an O-RAN framework for the future 6G networks. The

proposed considerations do not exclude other solutions where some tasks are divided into multiple subtasks. The final selection depends on specific service providers' concerns: latency, CC, data movement cost, and privacy issues.

The MG CQI-based subgrouping for multi-rate multicasting and the dynamic multicast access technique selection and resource allocation are the ML tasks covered in Section 6.1. As presented in Table 6.3, the task CQI-based subgrouping for multi-rate multicasting can be solved through supervised and unsupervised ML algorithms. As presented above, our specific subgrouping solution was based on the unsupervised K-Means algorithm. We visualize this task inserted into the O-RAN architecture as a xApp in the Near-RT RIC (*scenario 1.2* or *1.4*). This specific task can tolerate control loops between 10 ms-1 s.

In the case of the dynamic multicast access technique selection, this task can be solved through supervised and DRL solutions. As we present in Section 6.1, we design this task based on supervised MLP and ETC. However, features extensions of this task considering a more complex scenario with multiple BSs and shared multicast/unicast service delivery could be solved through DRL with the FL paradigm enabling multiple local agents to cooperate in building an ML global model based on individual experiences collected [141]. We visualize this specific task as inserted into *scenario 1.2*, *1.4*, or *1.5*. Nevertheless, selecting the specific scenario depends on the network entities' use case characteristics, security, and computational resources. For example, industrial automation, remote surgery, and XR applications are extremely sensitive to latency [8]. Thus, *scenario 1.5* or novel future scenarios as discussed in [15, 138] could be recommended for the above use cases to meet tight QoS requirements with control loops below 10 ms. In every case, the solution must consider the users' channel/spatial information (i.e., CQI report, mobility behavior), SLA, and BS parameters.

6.3.3 Conclusions

This Section discusses the insertion of ML-based multicasting RRM solutions into the envisioned disaggregated 6G O-RAN framework. We analyze specific MBS tasks and the importance of a native decentralized, softwarized, and intelligent conception. Moreover, we discuss the possible scenarios to insert such trending multicast ML solutions into the O-RAN framework.

It becomes crucial for the research community, industry, and network/service operators to work on effectively integrating multicasting ML-RRM into the envisioned O-RAN. These technologies are essential to comply with the stringent requirements of future 6G networks. However, they present several challenges regarding computational complexity, power consumption, security, and storage/processing resources that must be handled.

Chapter 7

Conclusions and Future Works

7.1 Conclusions and Remarks

The envisaged 5G and beyond networks represent a paradigm shift for global communications, offering unprecedented breakthroughs in media service delivery. Addressing the critical research verticals and challenges for the IMT-2030 framework will require a compelling mix of enabling RATs and native softwarized, disaggregated, and intelligent conceptions. Integrating the MBS capability is an appealing feature to overcome the ever-growing traffic demands, disruptive multimedia services, massive connectivity, and low-latency applications, offering considerable capacity gain and cost-effective delivery mechanisms. Nevertheless, integrating MBS into the upcoming 5G and beyond networks requires embracing well-established PTM and novel top-notch technologies with seamless PTP communications convergence.

Despite the enormous benefits of MBS, this capability will evolve into a complex fabric where multiple challenges must be addressed in the pass-through. In this research, we contextualize and address three primary challenges under the MBS research umbrella: the necessity of ensuring stringent QoS/QoE requirements, embracing the mmWave and sub-THz propagation, and handling complex mobility behaviors. We cover these critical challenges by addressing the trade-off between multicasting gain and multiuser diversity, along with the trade-off between optimal network performance and CC.

In this research, we cover essential aspects at the intersection of MBS, RRM solutions, ML, and the O-RAN framework. We characterize and address the dynamic multicast multiuser diversity through low-complexity RRM solutions aided by ML and OMA/NOMA in 5G MBS and beyond networks. We define the interrelations among the variables that shape the performance of the multicast access techniques: CMS, S-OMA, and S-NOMA. We characterize their dynamic behavior regarding the user's reception conditions, multimedia service constraints, and network parameters. We provide conditions for an adequate dynamic selection of the multicast strategy that better suits the specific network characteristics.

We provide a heuristic method for dynamically selecting the multicast access technique and resource allocation based on the proposed performance equations. The proposed approach effectively handles the trade-off between the available multicasting gain and the existing multiuser diversity. Moreover, we propose a multicasting strategy based on fixed pre-computed MIMO multi-beams and S-NOMA, tackling specific throughput requirements for enabling XR applications to attend multiple users in a 5G MBS use case. The proposed strategy allows us to dynamically take advantage of the users' spatial and channel quality diversity, maximizing specific QoS metrics.

After proposing baseline heuristic algorithms, we address the CC associated with the dynamic multicast RRM strategies and highlight the implications of fast variations in the MG members' reception conditions. We propose two novel solutions for the dynamic multicast access technique selection aided multiclass classification ML algorithms with MLP and ETC. We propose the use of a K-Means clustering unsupervised ML approach for detecting and splitting group-oriented MGs based on the CQI values at the BS. We present a multicast-oriented trigger to avoid overrunning the entire algorithm subject to the temporal CQI variations of the MGs. Our proposed approaches allow addressing the trade-off between optimal network performance and CC by maximizing specific QoS parameters through non-optimal solutions, considerably reducing the CC of conventional exhaustive mechanisms.

As a step forward in addressing the complexity that characterizes the envisioned MBS capability, we discuss the insertion of ML-based multicasting RRM solutions into the envisioned disaggregated 6G O-RAN framework. We analyze specific MBS tasks and the importance of a native decentralized, softwarized, and intelligent conception.

We assess the effectiveness of our proposal under multiple numerical and link-level simulations of recreated 5G MBS use cases operating in μWave and mmWave, evaluating a wide range of network conditions and users mobility behaviors. The carried link-level simulations are based on an implemented ad-hoc simulator oriented to facilitate the studies on 5G MBS. The simulator combines unicast/multicast/broadcast capabilities over terrestrial and airborne network deployments.

7.2 Future Works

After finishing this manuscript, multiple identified possible improvements, unsolved tasks, and future works must be conducted to manage the identified research problems and critical challenges properly. In the following, we identify our primary future works under discussion.

- Let us depart from the premise that the envisioned future networks need to encompass the principle of human-centric networks targeting users' QoE satisfaction aside from network-centric QoS optimization [19]. We must include

the QoE as a KPI during our dynamic multicasting RRM solution-finding process. We must add the capability of quantifying through objective methods the users' QoE in the recreated and evaluated MBS use cases. Embracing QoE metrics will enrich the scope of the above-covered solutions in our future works.

- In Chapter 5, we presented the design of an algorithm that combines multicasting over fixed pre-computed MIMO multi-beams and multi-rate subgrouping, taking advantage of the users' spatial and channel quality diversity. We need to extend this solution by adding adaptive multicast multi-beam beamforming capability. Our solution must be able to dynamically adjust the number of multicast beams and their parameters concerning the MG spatial and channel quality distribution, the service requirements, and the network characteristics. This approach will enable new degrees of freedom regarding efficient resource utilization and users' QoS/QoE improvements.

- The proposed solutions and addressed MBS use cases in Chapters 4, 5, and 6 were based on a single-cell NR BS, a restringed coverage area and a single multicast multimedia service delivery to multiple users interested in the same content. Nevertheless, the envisioned MBS future applications will be embedded into ultra-dense HetNets with a 3D coverage, multiple RATs, shared unicast/multicast service delivery, and the ABC paradigm [35, 91, 92]. In such context, we must extend our analysis considering multiple geographically distributed BSs in cell-free massive MIMO mode, allowing extra spatial diversity and co-processing gain by simultaneously and coherently delivering multiple unicast/multicast services [2, 8]. Moreover, our solution must consider the integration and cooperation of TNs/NTNs.

- As the future networks evolve into a unifying connectivity fabric, new capabilities will emerge, enabling and enhancing disruptive services such as the novel paradigm of Mobility as a Service (MaaS) [142–145]. MaaS will demand broadband communication with strict latency requirements, improved network capacity, coverage, and mobility support for a potentially high number of concurrent users with diverse traffic requirements. In such context, we must apply the proposed solutions enriched by the above presented future works to specific MaaS use cases such as context awareness, sensors data sharing, and emergency information dissemination, providing resilient, flexible, and resource-efficient delivery mechanisms to multiple end-users. The envisioned MBS solutions will foster the effective transition toward the MaaS paradigm with a direct impact on users' perception and transport system.

- As presented in Chapters 4, 5, and 6, all the recreated use cases and conducted validations were based on an implemented ad-hoc link-level simulator [41]. We must extend the capabilities of this simulator to embrace the above-presented

future work demands. Among the features we must add to our solution are enabling satellite link computation, RIS based link computation, dynamic mobility of unmanned aerial vehicles (UAVs) acting as aerial-BSs, and dynamic link-computation subject to the adaptive multi-beam beamforming. We must improve the simulator project framework to make it more usable and scalable.

Appendices

Appendix A

Main Mathematical Notations

Table A.1: Main mathematical notations.

Notation	Definition
R, R_M	Total number of RBs of the BS and assigned to the multicast session with $R_M \leq R$
R_M^*	Effective number of RBs used to deliver the multicast service with $R_M^* \leq R_M \leq R$
r_k	Number of RBs assigned to user k
B, B_0	Total bandwidth of the BS, and bandwidth of the RBs
Th_{min}, Th_{max}	Minimum and maximum throughput of the service
P	Percent of users to be served
$\mathcal{K}$, K	Set of users in the MG and total number of users
u_k	User k belonging to $\mathcal{K}$
$\mathcal{XYZ}, \{XYZ\}_k$	Set including the spatial information of the K users, and spatial information of each u_k
CQI_{MG}, CQI_k	Set of CQI values reported by the K users (CQI_k) in the MG
CQI_{min}	Minimum CQI in CQI_{MG}
$UM_{\text{MG}}, u_{cqi=x}$	Users' CQI distribution vector, and number of users in $\mathcal{K}$ that report a CQI equal to x
M	Number of different CQI values in CQI_{MG}
$SINR_{\text{min}}$	Minimum SINR required to correctly decode an specific MCS
eff	Efficiency value associated to an specific MCS
S	Number of subgroups, $S = 2$
$G1$, $G2$	Group 1 and Group 2
$\mathcal{K}_{\text{G1}}$, K_{G1}	Set of users belonging to $G1$, number of users of $G1$
$\mathcal{K}_{\text{G2}}$, K_{G2}	Set of users belonging to $G2$, number of users of $G2$

Notation	Definition
$CQI_{min}^{G1}, CQI_{min}^{G2}$	Minimum CQI in $G1$ and $G2$
C_k	Capacity of user k
$C_{\mathrm{C}}, C_{\mathrm{O}}, C_{\mathrm{N}}$	Capacity of CMS, S-OMA, and S-NOMA
eff	Efficiency value of the MCS
$eff_{\min} = eff_{\min}^{G1}$	Minimum eff in the MG and $G1$
$SINR_k^{\mathrm{UL,G1}}$	SINR of the UL after SINR adaptation
$SINR_k^{\mathrm{LL,G2}}$	SINR of the LL after SINR adaptation
$eff_{\min}^{\mathrm{UL,G1}}$	Minimum eff in $G1$ after SINR adaptation
$eff_{\min}^{\mathrm{LL,G2}}$	Minimum eff in $G2$ after SINR adaptation
$\mathcal{I}, il$	Set of IL, and specific il value
ADR_C, ADR_O, ADR_N	System ADR achieved with CMS, S-OMA, and S-NOMA
MDI_C, MDI_O, MDI_N	MDI achieved with CMS, S-OMA, and S-NOMA
PF_C, PF_O, PF_N	PF achieved with CMS, S-OMA, and S-NOMA
$\triangle_R$	Efficiency of the multicast strategies using the available resources to accomplish the service constraint, with $\Delta_R = R_M - R_M^*$
P_{tx}	Transmission power of the BS
G_{tx}, G_{rx}	Antenna gain for the transmitter (tx) and receiver (rx)
PL	Propagation path loss
$P_{rx,pl}$	Total received power at the end-user after path loss computation,PL
$P_{rx,k,r}$	Total received power at the user u_k and the r-th sub-channel
T_k, S_k, H_k	Penetration attenuation component, shadowing fading, and the fast-fading
$SINR_{k,r}$	SINR on the user u_k measured over each sub-channel r
F, N_0, B_r	noise figure, the noise spectral density (with a default value of -174 dBm/Hz), and the bandwidth of the sub-channel
$EESM_k$	EESM of u_k
θ	Antenna array's potential HPBW
θ_{3db}, θ_m	Angle at which the radiated power value is 3 dB below the maximum, angle of the maximum radiated power
L	Number of available beams
$\mathcal{L}$	Set of available pre-computed beams
Δ_C^O, Δ_C^N, Δ_0^N	Outperformace equations among CMS (C), S-OMA (O), S-NOMA (N)
CC_C, CC_O, CC_N	CC of CMS, S-OMA, S-NOMA

Appendix B

EESM Optimization

To find the optimal α and β that ensure better compression from the real $SINR_{u,r}$ to the $SINR_{eff}$, we use the PSO algorithm [115]. The PSO is a meta-heuristic algorithm inspired by the behavior of natural swarms, such as flocks of birds or fishes. This algorithm finds the global maximum or minimum in a non-convex optimization problem defined for a specific cost function. For our proposal, we use the PSO to minimize the following cost function (C_f):

$$\arg\min_{(\alpha,\beta)} \frac{1}{NcNs} \sum_{c=1}^{Nc} \sum_{s=1}^{Ns} (SINR_{AWGN}^{*,s,c} - SINR_{eff}^{s,c})^2, \tag{B.1}$$

where Nc is the number of channel realizations and Ns is the number of $SINR_{u,r}$ arrays evaluated for one specific CQI and one channel realization. As presented in [107], the critical range for the CQI estimation is defined from the SINR value corresponding to a BLER of 0.9 ($SINR_{0.9}$) to the SINR value corresponding to a BLER of 0.001 ($SINR_{0.001}$) for a target BLER of 0.1. Therefore, this is the critical training range for α and β. Then, Ns are the total $SINR_{AWGN}^{s,c}$ values into the limits $SINR_{0.9}^{CQI}$ and $SINR_{0.001}^{CQI}$. The process for finding the optimal calibration of α and β and effectively estimating the CQI is shown in algorithm 9

Algorithm 9: Optimal calibration of α and β through PSO

Input: C_f: CQI, $SINR_{0.9}^{CQI}$, $SINR_{0.001}^{CQI}$, Nc
Input: **PSO**: Particles, Iter, Bounds, options
Output: $(\alpha^{'},\beta^{'})$
for $c \leftarrow 1$ **to** Nc **do**
  for $SINR_{AWGN} \leftarrow (-10 : 0.2 : 40)$ **do**
    $s = 0$
    $S = ComputeShadowingAttenuation$
    for $r \leftarrow 1$ **to** Rb **do**
      $H_r = ComputeFastFadingAttenuation$
      $SINR_r = SINR_{AWGN} + S + H_r$
    end
    $BLER_{real} = GetBlerReal(SINR_r, CQI)$
    $SINR_{AWGN}^{*} = GetSinrAwgn(BLER_{real}, CQI)$
    if $SINR_{0.9}^{CQI} \leq SINR_{AWGN}^{*} \geq SINR_{0.01}^{CQI}$ **then**
      $s = s + 1$
      $SINR_{eff}^{s,c} = ComputeEESM(SINR_r, \alpha, \beta)$
      $SINR_{AWGN}^{*,s,c} = SINR_{AWGN}^{*}$
    else
  end
end
$C_f = \frac{1}{NcNs} \sum_{c=1}^{Nc} \sum_{s=1}^{Ns} (SINR_{AWGN}^{*,s,c} - SINR_{eff}^{s,c})^2$
$(\alpha^{'}, \beta^{'}) = ExecutePSO(C_f)$

Appendix C

Publications

C.1 Directly Related to the Thesis

- **E. F. Pupo**, C. C. González, L. Atzori and M. Murroni, "Thresholds of outperformance among Broadcast/Multicast access techniques in 5G networks," 2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Chengdu, China, 2021, pp. 1-6, doi: 10.1109/BMSB53066.2021.9547169 [39].

- **E. F. Pupo**, C. C. González, L. Atzori and M. Murroni, "Dynamic Multicast Access Technique in SC-PTM 5G Networks: Subgrouping with OM/NOM," 2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Bilbao, Spain, 2022, pp. 1-6, doi: 10.1109/BMSB55706.2022.9828674 [23].

- **E. F. Pupo**, C. C. González, E. Iradier, J. Montalban, and M. Murroni, "5G Link-Level Simulator for Multicast/Broadcast Services," 2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Beijing, China, 2023, pp. 1-6, doi: 10.1109/BMSB58369.2023.10211507 [42].

- **E. F. Pupo**, C. C. González, E. Iradier, J. Montalban, P. Angueira, and M. Murroni (2023). Dynamic Single/Multi-Rate Multicasting Aided NOMA for Addressing the Multiuser Diversity in 5G Networks. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.24312049.v1 [40]. (Under revision: IEEE Transactions on Network and Service Management)

- **E. F. Pupo**, González, C. C., Montalban, J., Angueira, P., Murroni, M., & Iradier, E. (2023). Artificial Intelligence Aided Low Complexity RRM Algorithms for 5G-MBS. IEEE Transactions on Broadcasting [41].

- **E. F. Pupo**, C. C. González, V. Popescu, D. Giusto and M. Murroni, "Beyond 5G Multicast for XR Communications aided by Pre-computed Multi-beams and NOMA," 2023 IEEE Global Communications Conference (GLOBECOM), Kuala Lumpur, Malaysia, 2023, pp. 1-6, (accepted paper)

- **E. F. Pupo**, C. C. González, E. Iradier, J. Montalban, P. Angueira, and M. Murroni (2023). Machine Learning-based Multicasting Radio Resource Management over 6G O-RAN Framework. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.24408250.v1 [15]. (Under revision: IEEE Communications Magazine)

C.2 Other publications

- C. C. González, **E. F. Pupo**, D. P. Ruisánchez, D. Plets and M. Murroni, "Three-stages concatenated Machine Learning model for SFN prediction," 2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Chengdu, China, 2021, pp. 1-6, doi: 10.1109/BMSB53066.2021.9547146.

- C. C. González, **E. F. Pupo**, L. Atzori and M. Murroni, "Dynamic access control and slice allocation algorithm for diverse traffic demand over 5G heterogeneous networks," 2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Chengdu, China, 2021, pp. 1-6, doi: 10.1109/BMSB53066.2021.9547129.

- D. P. Ruisánchez, D. G. Mirabal, **E. F. Pupo**, C. C. González, D. Pérez-Adán and F. A. Cesar, "Prediction of Signal Quality and SFN Interference Metrics Using Machine Learning Models," 2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Chengdu, China, 2021, pp. 1-6, doi: 10.1109/BMSB53066.2021.9547111.

- C. C. González, **E. F. Pupo**, D. P. Ruisánchez, D. Plets and M. Murroni, "From MFN to SFN: Performance Prediction Through Machine Learning," in IEEE Transactions on Broadcasting, vol. 68, no. 1, pp. 180-190, March 2022, doi: 10.1109/TBC.2021.3132804.

- C. C. González, **E. F. Pupo**, L. Atzori and M. Murroni, "Dynamic Radio Access Selection and Slice Allocation for Differentiated Traffic Management on Future Mobile Networks," in IEEE Transactions on Network and Service Management, vol. 19, no. 3, pp. 1965-1981, Sept. 2022, doi: 10.1109/TNSM.2022.3150978.

- C. C. González, **E. F. Pupo**, D. Pereira-Ruisánchez, L. Atzori and M. Murroni, "Deep Reinforcement Learning for Dynamic Radio Access Selection over

Future Wireless Networks," 2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Bilbao, Spain, 2022, pp. 1-6, doi: 10.1109/BMSB55706.2022.9828746.

- Alonso, R. M., Plets, D., Martens, L., Joseph, W., **E. F. Pupo**., Nieto, G. G. (2023). White spaces pattern finding and inference based on machine learning for multi-frequency spectrum footprints. Computer Networks, 109871.

- C. C. González, **E. F. Pupo**, J. Montalban, S. Pizzi, E. Iradier and M. Murroni, "Hybrid Terrestrial-Airborne Connectivity for Unicast/Broadcast Services Beyond 5G," 2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Beijing, China, 2023, pp. 1-6, doi: 10.1109/BMSB58369.2023.10211608.

- G. Brancati, **E. F. Pupo**, O. Chukhno, N. Chukhno, M. Murroni and G. Araniti, "Reconfigurable Intelligent Surface Deployment and Orientation in Beyond 5G Multicast Networks," 2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Beijing, China, 2023, pp. 1-6, doi: 10.1109/BMSB58369.2023.10211255.

Appendix D

Bio

Ernesto Fontes Pupo (Student Member, IEEE) (e.fontespupo@studenti.unica.it) received a B.Sc. degree in Telecommunications and Electronics engineering and an M.Sc. degree in Digital Systems from the Havana University of Technologies (CUJAE) in 2014 and 2018, respectively. He is a Ph.D. student at the Department of Electrical and Electronic Engineering (DIEE/UdR CNIT), University of Cagliari, 09123, Sardinia, Italy. He was an assistant researcher with LACETEL, R&D Telecommunications Institute (2014-2020). His research interests include 5G wireless networks and beyond, multicast/broadcast services, QoS, and artificial intelligence.

Appendix E

Acronyms

2D two-dimensional

3D three-dimensional

3GPP 3rd Generation Partnership Project

5G fifth-generation

5GC 5G core network

6G sixth-generation

ABC always best-connected

Adam Adaptive Moment Estimation

ADR aggregated data rate

AI artificial intelligence

ANN artificial neural network

AR augmented reality

AWGN additive white gaussian noise

BER bit error rate

BF beamforming

BLER block error rate

BS base station

btESS-MAT basic trigger ESS-MAT

BW bandwidth

BWP bandwidth part

CC computational complexity

CCR CQI changing ratio

CF cost function

CMS conventional multicast scheme

CQD channel quality diversity

CQI channel quality indicator

CSI channel state information

CP control plane

CPU central processing unit

D2D Device-to-Device

D2DM D2D underlaid multicasting

DL deep learning

DMFM dynamic multicasting over fixed pre-computed multi-beams

DRL deep reinforcement learning

E2E end-to-end

EESM Exponential Effective SINR Metric

eMBB enhanced Mobile Broadband

eMBMS evolved multimedia broadcast multicast service

EN-TV Enhanced Television

ESS exhaustive search strategy

ESS-MAT ESS-Multicast Access Technique Selection

ETC extra tree classifier

ETC-MAT ETC-Multicast Access Technique Selection

FDM frequency-division multiplexing

FeMBMS further evolved multimedia broadcast multicast service

FL federated learning

FoV field-of-view

gNB gNodeB

GPS Global Positioning System

HetNet ultra-dense heterogeneous network

HPBW half-power beamwidth

IL injection level

IMT International Mobile Telecommunications

IoT internet of things

KPI key performance indicator

LDM layer-division multiplexing

LL lower layer

LLS link-level simulator

LTE Long Term Evolution

MaaS Mobility as a Service

MBS multicast/broadcast service

MBSFN multicast-broadcast SFN

MCS modulation and coding scheme

MDI minimum dissatisfaction index

MEC mobile edge computing

MG multicast group

MIMO multiple-input multiple-output

ML machine learning

MLP multi-layer perceptron

MLP-MAT MLP-Multicast Access Technique Selection

mMIMO masive MIMO

mMTC massive Machine Type Communications

mmWave millimeter-wave

Near-RT RIC Near-Real Time RAN Intelligent Controller

NG-RAN next generation radio access network

NOMA non-orthogonal multiple access

Non-RT RIC Non-Real Time RAN Intelligent Controller

NR new radio

NS network slice

NTN non-terrestrial network

OCI outperformance conditions identification

OE-MAT Outperformance Equation-Multicast Access Technique Selection

OFDM orthogonal frequency-division multiplexing

OMA orthogonal multiple access

OMS opportunistic multicasting scheme

O-CU Open-RAN Control Unit

O-DU Open-RAN Distributed Unit

O-RAN Open RAN

O-RU Open-RAN Radio Unit

PF proportional fairness

P-NOMA power-domain NOMA

PSO particle swarm optimization

PTM point-to-multipoint

PTP point-to-point

QoE quality of experience

QoPE quality of physical experience

QoS quality of service

RAN radio access network

RAT radio access technologies

RB resource block

Rel Release

ReLU Rectified linear activation function

RIC RAN Intelligent Controller

RIS reflective intelligent surface

RRC radio resource control

RRM radio resource management

RSMA rate-splitting multiple access

SC-PTM single cell point-to-multipoint

SCS subcarrier spacing

SD spatial diversity

SFN single frequency network

SIC successive interference cancellation

SINR signal-to-interference-plus-noise ratio

SLA service level agreement

SMO Service Management and Orchestration

SMOTE Synthetic Minority Oversampling Technique

S-NOMA subgrouping based on NOMA

SO specific objective

S-OMA subgrouping based on OMA

s.t. subject to

SVC scalable video coding

TDM time-division multiplexing

THz Terahertz

TN terrestrial networks

TTI transmission time interval

UAV unmanned aerial vehicle

UE user equipment

UL upper layer

UMi urban microcell

UP user plane

URLLC Ultra-Reliable and Low-Latency Communications

V2X Vehicular-to-Everything

VR virtual reality

XR extended reality

Bibliography

[1] Ericsson. *Mobility report, forecasts*. www.ericsson.com/mobility-report. 2022.

[2] C. C. González et al. "Multicasting over 6G Non-Terrestrial Networks: a Softwarization-based Approach". In: *IEEE Vehicular Technology Magazine* (2023).

[3] R. Liu et al. "A Vision and An Evolutionary Framework for 6G: Scenarios, Capabilities and Enablers". In: *arXiv preprint arXiv:2305.13887* (2023).

[4] R. Liu et al. "Beginning of the Journey towards 6G: Vision and Framework". In: *IEEE Communications Magazine* (2023), pp. 1–2.

[5] S. Bisoyi et al. "IMT 2030 CANDIDATE TECHNOLOGIES: A white paper by IIT Hyderabad & WiSig Networks". In: *arXiv preprint arXiv:2307.14978* (2023).

[6] M. Polese et al. "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges". In: *IEEE Communications Surveys & Tutorials* (2023).

[7] 3. T. 1. 502. "5G; 5G multicast-broadcast services; User service architecture (3GPP TS 26.502 version 17.6.0 Release 17)". In: (2023).

[8] N. Chukhno et al. "Approaching 6G Use Case Requirements with Multicasting". In: *IEEE Communications Magazine* 61.5 (2023), pp. 144–150.

[9] V. K. Shrivastava, S. Baek, and Y. Baek. "5G evolution for multicast and broadcast services in 3GPP release 17". In: *IEEE Communications Standards Magazine* 6.3 (2022), pp. 70–76.

[10] 3GPP. "Study on Single-Cell Point-to-Multipoint Transmission for E-UTRA, v13.0.0". In: *3GPP TR 36.890* (July 2015).

[11] Y. Zhang et al. "MBSFN or SC-PTM: How to efficiently multicast/broadcast". In: *IEEE Transactions on Broadcasting* 67.3 (2021), pp. 582–592.

[12] 3GPP. "Technical Specification Group Services and System Aspects; System architecture for the 5G System (5GS); Stage 2 (Release 16)". In: *3GPP TS 23.501 v16.7.0* (December 2020).

[13] 3GPP. "Service Requirements for the 5G system (Release 18)". In: *3GPP TS 22.261 V19.1.0* (Dec. 2022).

[14] X. Lin. "An overview of 5G advanced evolution in 3GPP release 18". In: *IEEE Communications Standards Magazine* 6.3 (2022), pp. 77–83.

[15] E. Fontes Pupo et al. "Machine Learning-based Multicasting Radio Resource Management over 6G O-RAN Framework". In: (2023).

[16] S. Han, T. Xie, and C.-L. I. "Greener Physical Layer Technologies for 6G Mobile Communications". In: *IEEE Communications Magazine* 59.4 (2021), pp. 68–74.

[17] P. Mach and Z. Becvar. "Device-to-device relaying: Optimization, performance perspectives, and open challenges towards 6G networks". In: *IEEE Communications Surveys & Tutorials* 24.3 (2022), pp. 1336–1393.

[18] J. Wang et al. "Artificial Intelligence-Assisted Network Slicing: Network Assurance and Service Provisioning in 6G". In: *IEEE Vehicular Technology Magazine* (2023).

[19] P. S. Rufino Henrique and R. Prasad. "6G Networks for Next Generation of Digital TV Beyond 2030". In: *Wireless Personal Communications* 121 (2021), pp. 1363–1378.

[20] M. Rasti et al. "Evolution toward 6G multi-band wireless networks: A resource management perspective". In: *IEEE Wireless Communications* 29.4 (2022), pp. 118–125.

[21] Y. Zhang et al. "Mode selection algorithm for multicast service delivery". In: *IEEE Transactions on Broadcasting* 67.1 (2020), pp. 96–105.

[22] A. B. Hassouna, H. Koubaa, and L. A. Saidane. "Multi-user diversity wireless multicast: A Survey". In: *Computer Networks* 175 (2020), p. 107282.

[23] E. F. Pupo et al. "Dynamic Multicast Access Technique in SC-PTM 5G Networks: Subgrouping with OM/NOM". In: *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE. 2022, pp. 1–6.

[24] E. Iradier et al. "Nonorthogonal multiple access and subgrouping for improved resource allocation in multicast 5G NR". In: *IEEE Open Journal of the Communications Society* 3 (2022), pp. 543–556.

[25] M. N. Dani et al. "Resource allocation for layered multicast video streaming in NOMA systems". In: *IEEE Transactions on Vehicular Technology* 71.11 (2022), pp. 11379–11394.

[26] A. de la Fuente, G. Interdonato, and G. Araniti. "User Subgrouping and Power Control for Multicast Massive MIMO Over Spatially Correlated Channels". In: *IEEE Transactions on Broadcasting* 68.4 (2022), pp. 834–847.

[27] N. Chukhno et al. “The use of machine learning techniques for optimal multicasting in 5G NR systems”. In: *IEEE Transactions on Broadcasting* (2022).

[28] Y. Xue et al. “Using layered division multiplexing for mixed unicast-broadcast service delivery in 5G”. In: *2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE. 2019, pp. 1–6.

[29] A. Yazar, S. Dogan-Tusha, and H. Arslan. “6G vision: An ultra-flexible perspective”. In: *ITU Journal on Future and Evolving Technologies* 1.1 (2020), pp. 121–140.

[30] P. Jain et al. “Performance Evaluation of Cooperative OMA and NOMA Systems in 6G Deployment Scenarios”. In: *Sensors* 22.11 (2022), p. 3986.

[31] J. Ghosh et al. “On the comparison of optimal NOMA and OMA in a paradigm shift of emerging technologies”. In: *IEEE Access* 10 (2022), pp. 11616–11632.

[32] J. Ghosh et al. “Performance investigation of NOMA versus OMA techniques for mmWave massive MIMO communications”. In: *IEEE Access* 9 (2021), pp. 125300–125308.

[33] G. Araniti et al. “A solution to the multicast subgroup formation problem in LTE systems”. In: *IEEE Wireless Communications Letters* 4.2 (2015), pp. 149–152.

[34] G. Brancati et al. “Reconfigurable Intelligent Surface Deployment and Orientation in Beyond 5G Multicast Networks”. In: *2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE. 2023, pp. 1–6.

[35] S. Manap et al. “Survey of radio resource management in 5G heterogeneous networks”. In: *IEEE Access* 8 (2020), pp. 131202–131223.

[36] G. J. Sutton et al. “Enabling technologies for ultra-reliable and low latency communications: From PHY and MAC layer perspectives”. In: *IEEE Communications Surveys & Tutorials* 21.3 (2019), pp. 2488–2524.

[37] J. S. Vardakas et al. “Machine Learning-Based Cell-Free Support in the O-RAN Architecture: An Innovative Converged Optical-Wireless Solution Toward 6G Networks”. In: *IEEE Wireless Communications* 29.5 (2022), pp. 20–26.

[38] C. C. González et al. “Dynamic radio access selection and slice allocation for differentiated traffic management on future mobile networks”. In: *IEEE Transactions on Network and Service Management* 19.3 (2022), pp. 1965–1981.

[39] E. F. Pupo et al. “Thresholds of outperformance among Broadcast/Multicast access techniques in 5G networks”. In: *2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE. 2021, pp. 1–6.

[40] E. Fontes Pupo et al. “Dynamic Single/Multi-Rate Multicasting Aided NOMA for Addressing the Multiuser Diversity in 5G Networks”. In: (2023).

[41] E. F. Pupo et al. “Artificial Intelligence Aided Low Complexity RRM Algorithms for 5G-MBS”. In: *IEEE Transactions on Broadcasting* (2023).

[42] E. F. Pupo et al. “5G Link-Level Simulator for Multicast/Broadcast Services”. In: *2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE. 2023, pp. 1–6.

[43] D. Mi et al. “Demonstrating Immersive Media Delivery on 5G Broadcast and Multicast Testing Networks”. In: *IEEE Transactions on Broadcasting* 66.2 (2020), pp. 555–570.

[44] X. (Shen et al. “Toward immersive communications in 6G”. In: *Frontiers in Computer Science* 4 (2023).

[45] M. Xu et al. “A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges”. In: *IEEE Communications Surveys & Tutorials* (2022).

[46] D. Sim et al. “Low-latency haptic open glove for immersive virtual reality interaction”. In: *Sensors* 21.11 (2021), p. 3682.

[47] H. Alves et al. “Beyond 5G URLLC evolution: New service modes and practical considerations”. In: *arXiv preprint arXiv:2106.11825* 7 (2021).

[48] M. Ali et al. “Metaverse Communications, Networking, Security, and Applications: Research Issues, State-of-the-Art, and Future Directions”. In: *arXiv preprint arXiv:2212.13993* (2022).

[49] L.-a. Matti et al. *Key drivers and research challenges for 6G*. 2019.

[50] A. Ghandri, H. E. Nouri, and M. B. Jemaa. “Deep Learning for VBR Traffic Prediction-Based Proactive MBSFN Resource Allocation Approach”. In: *IEEE Transactions on Network and Service Management* (2023).

[51] G. Araniti et al. “A hybrid unicast-multicast network selection for video deliveries in dense heterogeneous network environments”. In: *IEEE Transactions on Broadcasting* 65.1 (2018), pp. 83–93.

[52] M. M. Azari et al. “Evolution of non-terrestrial networks from 5G to 6G: A survey”. In: *IEEE communications surveys & tutorials* (2022).

[53] F. Zhou et al. “DRL-based low-latency content delivery for 6G massive vehicular IoT”. In: *IEEE Internet of Things Journal* 9.16 (2021), pp. 14551–14562.

[54] L. Feng et al. “Smart mode selection using online reinforcement learning for VR broadband broadcasting in D2D assisted 5G HetNets”. In: *IEEE Transactions on Broadcasting* 66.2 (2020), pp. 600–611.

[55] L. Zhong et al. “Decentralized Optimization for Multicast Adaptive Video Streaming in Edge Cache-Assisted Networks”. In: *IEEE Transactions on Broadcasting* (2023).

[56] H. Hao et al. “Multicast-aware optimization for resource allocation with edge computing and caching”. In: *Journal of Network and Computer Applications* 193 (2021), p. 103195.

[57] D. Nguyen et al. “Scalable multicast for live 360-degree video streaming over mobile networks”. In: *IEEE Access* 10 (2022), pp. 38802–38812.

[58] H.-H. Liu and H.-Y. Wei. “5G NR Multicast and Broadcast QoS Enhancement With Flexible Service Continuity Configuration”. In: *IEEE Transactions on Broadcasting* 68.3 (2022), pp. 689–703.

[59] J. Mu et al. “Machine learning-based 5g ran slicing for broadcasting services”. In: *IEEE Transactions on Broadcasting* 68.2 (2021), pp. 295–304.

[60] P.-Y. Su et al. “Priority-Aware Resource Allocation for 5G mmWave Multicast Broadcast Services”. In: *IEEE Transactions on Broadcasting* (2022).

[61] C. C. González et al. “Hybrid Terrestrial-Airborne Connectivity for Unicast/Broadcast Services Beyond 5G”. In: *2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE. 2023, pp. 1–6.

[62] J.-M. Vella and S. Zammit. “A survey of multicasting over wireless access networks”. In: *IEEE Communications Surveys & Tutorials* 15.2 (2012), pp. 718–753.

[63] J. Montalban et al. “Multimedia multicast services in 5G networks: Subgrouping and non-orthogonal multiple access techniques”. In: *IEEE Communications Magazine* 56.3 (2018), pp. 91–95.

[64] R. O. Afolabi, A. Dadlani, and K. Kim. “Multicast scheduling and resource allocation algorithms for OFDMA-based systems: A survey”. In: *IEEE Communications Surveys & Tutorials* 15.1 (2012), pp. 240–254.

[65] W. Rhee and J. M. Cioffi. “Increase in capacity of multiuser OFDM system using dynamic subchannel allocation”. In: *VTC2000-spring. 2000 IEEE 51st vehicular technology conference proceedings (Cat. No. 00CH37026)*. Vol. 2. IEEE. 2000, pp. 1085–1089.

[66] T.-P. Low et al. “Optimized opportunistic multicast scheduling (OMS) over wireless cellular networks”. In: *IEEE Transactions on Wireless Communications* 9.2 (2010), pp. 791–801.

[67] P. Polacek, T.-Y. Yang, and C.-W. Huang. "Opportunistic multicasting for single frequency networks". In: *Wireless Communications and Mobile Computing* 16.15 (2016), pp. 2253–2262.

[68] G. Araniti et al. "Adaptive resource allocation to multicast services in LTE systems". In: *IEEE Transactions on Broadcasting* 59.4 (2013), pp. 658–664.

[69] S. Zhang et al. "Rate-Splitting Multiple Access-Based Satellite–Vehicular Communication System: A Noncooperative Game Theoretical Approach". In: *IEEE Open Journal of the Communications Society* 4 (2023), pp. 430–441.

[70] H. Schwarz, D. Marpe, and T. Wiegand. "Overview of the scalable video coding extension of the H. 264/AVC standard". In: *IEEE Transactions on circuits and systems for video technology* 17.9 (2007), pp. 1103–1120.

[71] X. Zhang et al. "Caching scalable videos in the edge of wireless cellular networks". In: *IEEE Network* (2022).

[72] Y. Mao et al. "Rate-splitting multiple access: Fundamentals, survey, and future research trends". In: *IEEE Communications Surveys & Tutorials* (2022).

[73] L. Militano et al. "Radio resource management for group-oriented services in LTE-A". In: *IEEE Transactions on Vehicular Technology* 64.8 (2014), pp. 3725–3739.

[74] G. Araniti et al. "A low-complexity resource allocation algorithm for multicast service delivery in OFDMA networks". In: *IEEE Transactions on Broadcasting* 60.2 (2014), pp. 358–369.

[75] W. Guo and B. Mouhouche. "A method to tailor broadcasting and multicasting transmission in 5G new radio". In: *2019 European Conference on Networks and Communications (EuCNC)*. IEEE. 2019, pp. 364–368.

[76] E. Iradier et al. "Adaptive resource allocation in LTE vehicular services using LDM". In: *2016 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE. 2016, pp. 1–6.

[77] L. Zhang et al. "Layered-division-multiplexing: Theory and practice". In: *IEEE Transactions on Broadcasting* 62.1 (2016), pp. 216–232.

[78] M. S. Islam et al. "Layer division multiplexing for 5G DL transmission within ultra-dense heterogeneous networks". In: *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*. IEEE. 2020, pp. 1–7.

[79] E. Iradier et al. "Advanced NOMA-based RRM schemes for broadcasting in 5G mmWave frequency bands". In: *IEEE Transactions on Broadcasting* 68.1 (2021), pp. 143–155.

[80] E. Iradier et al. "Using NOMA for enabling broadcast/unicast convergence in 5G networks". In: *IEEE Transactions on Broadcasting* 66.2 (2020), pp. 503–514.

[81] P. Wang, J. Xiao, and L. Ping. "Comparison of orthogonal and non-orthogonal approaches to future wireless cellular systems". In: *IEEE Vehicular Technology Magazine* 1.3 (2006), pp. 4–11.

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

[83] Z. Wei et al. "On the performance gain of NOMA over OMA in uplink communication systems". In: *IEEE Transactions on Communications* 68.1 (2019), pp. 536–568.

[84] L. Zhang et al. "Using non-orthogonal multiplexing in 5G-MBMS to achieve broadband-broadcast convergence with high spectral efficiency". In: *IEEE Transactions on Broadcasting* 66.2 (2020), pp. 490–502.

[85] H. Park et al. "An incremental multicast grouping scheme for mmWave networks with directional antennas". In: *IEEE communications letters* 17.3 (2013), pp. 616–619.

[86] A. Biason and M. Zorzi. "Multicast via point to multipoint transmissions in directional 5G mmWave communications". In: *IEEE Communications Magazine* 57.2 (2019), pp. 88–94.

[87] Y. Li et al. "Radio resource management considerations for 5G millimeter wave backhaul and access networks". In: *IEEE Communications Magazine* 55.6 (2017), pp. 86–92.

[88] N. Chukhno et al. "Efficient management of multicast traffic in directional mmWave networks". In: *IEEE Transactions on Broadcasting* 67.3 (2021), pp. 593–605.

[89] N. Chukhno et al. "Optimal multicasting in millimeter wave 5G NR with multi-beam directional antennas". In: *IEEE Transactions on Mobile Computing* (2021).

[90] Y. Xu et al. "A survey on resource allocation for 5G heterogeneous networks: Current research, future trends, and challenges". In: *IEEE Communications Surveys & Tutorials* 23.2 (2021), pp. 668–695.

[91] I. A. Bartsiokas et al. "ML-based radio resource management in 5G and beyond networks: A survey". In: *IEEE Access* 10 (2022), pp. 83507–83528.

[92] M. Z. Chowdhury et al. "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions". In: *IEEE Open Journal of the Communications Society* 1 (2020), pp. 957–975.

[93] B. Agarwal et al. "A Comprehensive Survey on Radio Resource Management in 5G HetNets: Current Solutions, Future Trends and Open Issues". In: *IEEE Communications Surveys & Tutorials* (2022).

[94] N. Naderializadeh et al. “Resource management in wireless networks via multi-agent deep reinforcement learning”. In: *IEEE Transactions on Wireless Communications* 20.6 (2021), pp. 3507–3523.

[95] B. Brik, K. Boutiba, and A. Ksentini. “Deep learning for B5G open radio access network: Evolution, survey, case studies, and challenges”. In: *IEEE Open Journal of the Communications Society* 3 (2022), pp. 228–250.

[96] L. Bonati et al. “Intelligence and learning in O-RAN for data-driven NextG cellular networks”. In: *IEEE Communications Magazine* 59.10 (2021), pp. 21–27.

[97] O-RAN. “O-RAN Working Group 2 AI/ML workflow description and requirements”. In: *O-RAN.WG2.AIML-v01.03* (July 2021).

[98] 3. T. 1. 214. “5G; NR; Physical layer procedures for data (3GPP TS 38.214 version 16.2.0 Release 16)”. In: (2020).

[99] A. J. Goldsmith and S.-G. Chua. “Variable-rate variable-power MQAM for fading channels”. In: *IEEE transactions on communications* 45.10 (1997), pp. 1218–1230.

[100] X. He et al. “Link layer abstraction in MIMO-OFDM system”. In: *2007 International Workshop on Cross Layer Design*. IEEE. 2007, pp. 41–44.

[101] Y. Wang, W. Liu, and L. Fang. “Adaptive modulation and coding technology in 5G system”. In: *2020 International Wireless Communications and Mobile Computing (IWCMC)*. IEEE. 2020, pp. 159–164.

[102] P. Dent, G. E. Bottomley, and T. Croft. “Jakes fading model revisited”. In: *Electronics letters* 13.29 (1993), pp. 1162–1163.

[103] J.-Y. Le Boudec. “Rate adaptation, congestion control and fairness: A tutorial”. In: *on line* (2008).

[104] 3GPP TR 38.901. “5G; Study on channel model for frequencies from 0.5 to 100 GHz (3GPP TR 38.901 version 16.1.0 Release 16)”. In: (2020).

[105] M. Pauli, U. Wachsmann, and S.-H. S. Tsai. *Quality determination for a wireless communications link*. US Patent 7,231,183. 2007.

[106] G. Piro et al. “Simulating LTE cellular systems: An open-source framework”. In: *IEEE transactions on vehicular technology* 60.2 (2010), pp. 498–513.

[107] H. Lee, H. Kim, and H. Park. “Novel Calibration of MIESM and Reduction of CQI Feedback for Improved Fast Link Adaptation”. In: *Electronics* 8.3 (2019), p. 278.

[108] A. B. Constantine. *Antenna Theory: Analysis and Design, Hoboken*. 2005.

[109] Python. *time-Time access and conversions*. https://docs.python.org/3 // library/time.htmltime.process$_t$*ime*. 2023.

[110] S. Gulwani, K. K. Mehra, and T. Chilimbi. "Speed: precise and efficient static estimation of program computational complexity". In: *ACM Sigplan Notices* 44.1 (2009), pp. 127–139.

[111] M. Shafi et al. "Microwave vs. millimeter-wave propagation channels: Key differences and impact on 5G cellular systems". In: *IEEE Communications Magazine* 56.12 (2018), pp. 14–20.

[112] L. Possenti et al. "A Study on mm-wave Propagation in and around Buildings". In: *IEEE Open Journal of Antennas and Propagation* (2023).

[113] T. S. Rappaport et al. *Radio Propagation Measurements and Channel Modeling: Best Practices for Millimeter-Wave and Sub-Terahertz Frequencies*. Cambridge University Press, 2022.

[114] J. Kennedy and R. Eberhart. "Particle swarm optimization". In: *Proceedings of ICNN'95-international conference on neural networks*. Vol. 4. IEEE. 1995, pp. 1942–1948.

[115] Y. Shi et al. "Particle swarm optimization: developments, applications and resources". In: *Proceedings of the 2001 congress on evolutionary computation (IEEE Cat. No. 01TH8546)*. Vol. 1. IEEE. 2001, pp. 81–86.

[116] X. Hong et al. "A group mobility model for ad hoc wireless networks". In: *Proceedings of the 2nd ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*. 1999, pp. 53–60.

[117] D. G. Altman. "Confidence intervals for the number needed to treat". In: *Bmj* 317.7168 (1998), pp. 1309–1312.

[118] Z. Zhang, M. Tao, and Y.-F. Liu. "Learning to Beamform in Joint Multicast and Unicast Transmission with Imperfect CSI". In: *IEEE Transactions on Communications* (2023).

[119] M. Aly. "Survey on multiclass classification methods". In: *Neural Netw* 19.1-9 (2005), p. 2.

[120] C. C. González et al. "From MFN to SFN: Performance Prediction Through Machine Learning". In: *IEEE Transactions on Broadcasting* 68.1 (2021), pp. 180–190.

[121] Y. Kim et al. "New radio (NR) and its evolution toward 5G-advanced". In: *IEEE Wireless Communications* 26.3 (2019), pp. 2–7.

[122] N. V. Chawla et al. "SMOTE: synthetic minority over-sampling technique". In: *Journal of artificial intelligence research* 16 (2002), pp. 321–357.

[123] A. Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow.* " O'Reilly Media, Inc.", 2022.

[124] S.-i. Kim et al. "Hybrid data-scaling method for fault classification of compressors". In: *Measurement* 201 (2022), p. 111619.

[125] P. Liashchynskyi and P. Liashchynskyi. "Grid search, random search, genetic algorithm: a big comparison for NAS". In: *arXiv preprint arXiv:1912.06059* (2019).

[126] B. W. Suter. "The multilayer perceptron as an approximation to a Bayes optimal discriminant function". In: *IEEE transactions on neural networks* 1.4 (1990), p. 291.

[127] P. Geurts, D. Ernst, and L. Wehenkel. "Extremely randomized trees". In: *Machine learning* 63 (2006), pp. 3–42.

[128] A. F. Agarap. "Deep learning using rectified linear units (relu)". In: *arXiv preprint arXiv:1803.08375* (2018).

[129] D. P. Kingma and J. Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[130] S. Lloyd. "Least squares quantization in PCM". In: *IEEE transactions on information theory* 28.2 (1982), pp. 129–137.

[131] *Study on Channel Model for Frequencies from 0.5 to 100 GHz (Release 14)*. Tech. rep. 3GPP TR 38.901 V14.1.1, 2017.

[132] P. Nain et al. "Properties of random direction models". In: *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.* Vol. 3. IEEE. 2005, pp. 1897–1907.

[133] K. B. Letaief et al. "The Roadmap to 6G: AI Empowered Wireless Networks". In: *IEEE Communications Magazine* 57.8 (2019), pp. 84–90.

[134] A. S. Abdalla et al. "Toward next generation open radio access networks: What O-RAN can and cannot do!" In: *IEEE Network* 36.6 (2022), pp. 206–213.

[135] L. Bonati et al. "Intelligence and Learning in O-RAN for Data-Driven NextG Cellular Networks". In: *IEEE Communications Magazine* 59.10 (2021), pp. 21–27.

[136] N. Chukhno et al. "Models, Methods, and Solutions for Multicasting in 5G/6G mmWave and sub-THz Systems". In: *IEEE Communications Surveys Tutorials* (2023), pp. 1–1.

[137] N. Kato et al. "Ten challenges in advancing machine learning technologies toward 6G". In: *IEEE Wireless Communications* 27.3 (2020), pp. 96–103.

[138] S. D'Oro et al. "dApps: Distributed applications for real-time inference and control in O-RAN". In: *IEEE Communications Magazine* 60.11 (2022), pp. 52–58.

[139] M. Katwe et al. "Dynamic user clustering and optimal power allocation in UAV-assisted full-duplex hybrid NOMA system". In: *IEEE Transactions on Wireless Communications* 21.4 (2021), pp. 2573–2590.

[140] C. C. González et al. "Deep Reinforcement Learning for Dynamic Radio Access Selection over Future Wireless Networks". In: *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE. 2022, pp. 1–6.

[141] M. K. Shehzad et al. "Artificial intelligence for 6G networks: Technology advancement and standardization". In: *IEEE Vehicular Technology Magazine* 17.3 (2022), pp. 16–25.

[142] A. Ihsan et al. "Energy-Efficient NOMA Multicasting System for Beyond 5G Cellular V2X Communications With Imperfect CSI". In: *IEEE Transactions on Intelligent Transportation Systems* 23.8 (2022), pp. 10721–10735.

[143] M. F. Pervej et al. "Mobility, communication and computation aware federated learning for internet of vehicles". In: *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2022, pp. 750–757.

[144] M. Hasselwander et al. "Towards sustainable transport in developing countries: Preliminary findings on the demand for mobility-as-a-service (MaaS) in metro Manila". In: *Transportation Research Part A: Policy and Practice* 155 (2022), pp. 501–518.

[145] E. Alyavina, A. Nikitas, and E. T. Njoya. "Mobility as a service (MaaS): A thematic map of challenges and opportunities". In: *Research in Transportation Business & Management* 43 (2022), p. 100783.