UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO

Multi-Agent Reinforcement Learning for Inter-Cell Interference Management in Hotspots Scenarios

Iago Diógenes do Rego

Orientador: Prof. Dr. Vicente Angelo de Sousa Jr

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Computação da UFRN (área de concentração: Telecomunicações) como parte dos requisitos para obtenção do título de Doutor em Engenharia.

Número de ordem PPgEEC: D382
Natal, RN, abril de 2025

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Rêgo, Iago Diógenes do.
Multi-Agent Reinforcement Learning for Inter-Cell Interference Management in Hotspots Scenarios / Iago Diógenes do Rêgo. - 2025.
112f.: il.

Tese (Doutorado) - Universidade Federal do Rio Grande do Norte, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica e de Computação, Natal, 2025.
Orientação: Dr. Vicente Angelo de Sousa Junior.
Coorientação: Dr. Dario Vieira.

1. Inter-cell Interference - Tese. 2. Fractional Frequency Reuse - Tese. 3. Reinforcement Learning - Tese. 4. Multi-agent - Tese. 5. Q-Learning - Tese. 6. MAB - Tese. I. Sousa Junior, Vicente Angelo de. II. Vieira, Dario. III. Título.

RN/UF/BCZM CDU 621.3

Elaborado por Raimundo Muniz de Oliveira - CRB-15/429

To my parents, who sacrificed so much to give me a quality education, and to my wife, who has been a true and faithful helper.

Agradecimentos

Ao Senhor, Deus soberano, que tudo sustenta e é o merecedor de toda a glória, rendo minha mais profunda gratidão. É Ele quem me capacita, e é a Ele que dedico todo o meu reconhecimento. Apesar das minhas falhas, Ele me usa como instrumento em Suas mãos.

Agradeço à minha esposa, auxiliadora idônea, com quem compartilhei todas as alegrias e dificuldades desta jornada acadêmica. Espero que você compreenda o quão importante é o seu apoio e constante encorajamento.

À minha família, que esteve presente ao longo de todo o caminho. Sou grato a Deus pela bênção de ter crescido em um lar que me incentivou a trilhar meus próprios caminhos, estando sempre disponível para me aconselhar ou corrigir quando necessário.

Ao meu orientador e amigo de longa data, Vicente, que sempre demonstrou uma admirável capacidade técnica, mas, acima de tudo, sempre se mostrou disponível e disposto a ajudar. Sua orientação moldou não apenas a tese, mas também minha trajetória profissional.

Em meio a diferentes projetos, universidades e parcerias, levo comigo uma longa lista de colegas e queridos amigos que me ajudaram a concluir essa jornada. À equipe do projeto nRIC, que contribuiu imensamente para meu amadurecimento como pesquisador e profissional; à Dario, que me deu a oportunidade de continuar trabalhando com pesquisa em novos horizontes; aos meus colegas da Efrei, que nunca recusaram uma boa conversa; e à Christiano, cuja constante disposição em ajudar e passar horas discutindo problemas técnicos, muitas vezes às custas de seu próprio trabalho, foi marcante.

Aos meus queridos amigos de longa data, sempre presentes nas conquistas e nos momentos difíceis, carrego cada um de vocês onde quer que eu vá. Aos meus amigos em Trinity, que, em tão pouco tempo, se tornaram amigos tão importantes. Estiveram comigo tanto nos dias difíceis quanto nos dias bons, e sua amizade fez toda a diferença.

Por fim, agradeço à UFRN por ter me proporcionado um excelente ambiente acadêmico desde a graduação até o doutorado. Sua infraestutura, corpo docente e administrativo foram essenciais durante todos esses anos. Sou igualmente grato ao PPgEEC, pelo apoio institucional e recursos disponibilizados, assim como aos professores e colegas que contribuíram de forma significativa para minha formação. Agradeço também ao Núcleo de Processamento de Alto Desempenho da UFRN (NPAD), cujo serviço de qualidade acelerou consideravelmente o desenvolvimento deste trabalho, e à Efrei, pela oportunidade de colaboração em pesquisa e pelo apoio financeiro.

A todos os citados, e a muitos outros que contribuíram com este trabalho, o meu muito obrigado!

Resumo

Interferência intercelular (Inter-cell Interference - ICI) é um desafio persistente em redes móveis. Embora os padrões atuais e futuros estejam evoluindo rapidamente, o aumento constante na demanda por dados, o surgimento de novos casos de uso, a coexistência de múltiplas tecnologias e o caráter dinâmico dos ambientes urbanos intensificam o impacto da interferência no desempenho do sistema. Ademais, a ICI torna-se especialmente desafiadora em cenários com alta densidade de pontos de acesso ou com zonas de alta densidade de usuários, conhecidas como *hotspots*. Para mitigar os efeitos negativos da ICI, o *Inter-Cell Interference Coordination* foi introduzido na *Release 8* do *3rd Generation Partnership Project* (3GPP), sendo o *Fractional Frequency Reuse* (FFR) uma das técnica de ICIC mais eficazes, amplamente aplicada em redes baseadas em *Orthogonal Frequency-Division Multiple Access* (OFDMA), como LTE e 5G. Embora originalmente desenvolvidas para o LTE, as técnicas de FFR também se mostram adequadas para redes 5G. Embora suas configurações estáticas frequentemente falhem em se adaptar a padrões de interferência dinâmicos, sua baixa complexidade e simplicidade de implementação tornam o FFR especialmente atrativo em cenários modernos. Essas técnicas oferecem uma base sólida para estratégias adaptativas de gerenciamento de interferência, especialmente quando integradas ao Aprendizado de Máquina (ML).

Esse trabalho tem o objetivo de investigar o uso de Aprendizado por Reforço (*Reinforcement Learning - RL*) para controlar dinamicamente múltiplos parâmetros de uma técnica de FFR em cenários caracterizados pelo surgimento inesperado de *hotspots*. A solução consiste em uma arquitetura hierárquica de múltiplos agentes, na qual dois agentes de RL otimizam em conjunto a alocação de banda e a classificação de usuários por meio do RSRQ *threshold*, sem uma relação de dependência direta no processo individual de aprendizado. A abordagem proposta é analisada quantitativamente e qualitativamente por meio do simulador de redes ns-3, utilizando um cenário composto por três estações rádio-base equipadas com antenas omnidirecionais, no qual uma fração dos usuários está confinada em *hotspots*. A cada intervalo de tempo pré-definido, um novo *hotspot* torna-se ativo, simulando um ambiente com demanda e interferência dinâmicos. Essa configuração é subdividida em dois cenários distintos, inspirados em especificações do 3GPP, representando, respectivamente, um cenário urbano denso e um cenário com grande número de dispositivos.

Os resultados obtidos demonstram que há uma forte interdependência entre a classificação dos usuários (utilizando o RSRQ *threshold*) e a alocação de banda, evidenciando que o controle combinado desses dois parâmetros tem forte impacto no desempenho da rede. A solução proposta apresentou melhor desempenho de forma

consistente, superando tanto a configuração estática quanto as soluções com agente único, nas quais apenas um dos parâmetros é ajustado dinamicamente. A solução multi-agente é capaz de explorar o espaço de ações conjuntas e convergir para a configuração mais eficaz, mesmo em cenários com alta interferência e distribuição de usuários em constante mudança, enquanto as soluções com agente único são limitadas pelo valor estático do parâmetro que não controlam. A solução proposta alcançou ganhos de vazão de até 99,4%, enquanto o controle isolado de um único parâmetro mostrou-se insuficiente em todos os casos avaliados.

Outra contribuição deste trabalho foi demonstrar a flexibilidade da estrutura proposta. Seu design hierárquico permite a integração de diferentes algoritmos de aprendizado por reforço. Duas estratégias de aprendizado foram implementadas e avaliadas: Q-Learning (QL) e Multi-Armed Bandits (MAB). O QL, ao considerar um processo de aprendizado que associa as ações à diferentes estados do sistema, alcançou desempenho superior, especialmente em cenários com maior interferência. No entanto, sua implementação exigiu um esforço adicional para a definição de estados representativos, além de envolver a exploração de um número maior de pares estado-ação. Por sua vez, os agentes MAB, por serem isentos de estados e mais simples de implementar, oferecem um bom equilíbrio entre complexidade e desempenho, especialmente em ambientes com capacidade de processamento limitada. Isso evidencia que a estrutura hierárquica pode ser adaptada para atender a diferentes requisitos do sistema, equilibrando desempenho e complexidade de implementação.

Esses resultados reforçam a contribuição central deste trabalho, que consiste na combinação de técnicas clássicas de ICIC com aprendizado por reforço para compor uma estratégia de mitigação de interferência flexível e de baixo *overhead*, capaz de se adaptar a condições de rede diversas e dinâmicas.

Keywords: ICI, Interferência intercelular, FFR, Aprendizado por Reforço, Multi-agente, Q-Learning, MAB

Abstract

Inter-cell interference (ICI) remains a critical challenge in mobile networks. Although current and future standards are rapidly evolving, the constant increase in data demand, the emergence of new use cases, the coexistence of multiple technologies, and the dynamic aspect of urban environments intensifies the impact of interference on system performance. ICI becomes especially challenging in dense deployments and in scenarios with zones of high user densities, referred to as hotspots. Fractional Frequency Reuse (FFR) is a well-established technique to mitigate ICI in OFDMA-based networks such as LTE and 5G, but traditional static configurations often fail to adapt to dynamic interference patterns. This thesis proposes a dynamic interference coordination framework based on reinforcement learning, designed to enhance the adaptability and performance of FFR techniques. The solution consists of a hierarchical multi-agent architecture, where two reinforcement learning agents operate in coordination, without direct exchange of information, to jointly control the allocation of bandwidth and user classification via the RSRQ threshold. The proposed approach was evaluated through network simulations using ns-3, across two different scenarios representing dense urban environments and massive connection conditions. The results show that the proposed framework consistently outperforms static and single-agent baselines, achieving throughput gains of up to 99.4%, particularly under high-interference conditions and for low-performing users. Furthermore, its modular design allows integration of different learning strategies. While Q-Learning agents delivered the highest performance, Multi-Armed Bandit (MAB) agents achieved comparable results with significantly lower computational complexity. By combining classical ICIC techniques with reinforcement learning, this work presents a flexible and low-overhead interference mitigation strategy that can adapt to diverse and evolving network conditions.

Keywords: Inter-cell Interference, FFR, Reinforcement Learning, Multi-agent, Q-Learning, MAB

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent decades, mobile networks have developed at an impressive rate, fundamentally changing how the world is connected and enabling new applications and services. From the early stages of mobile telephony to the data-driven networks of today, these advancements have changed the way we interact with the world and with each other.

Among these developments, 4G technology was an important milestone, marking a big shift in how mobile networks operate. Its introduction began with the standards of Long-Term Evolution (LTE) in 2008, which aimed to address the growing demand for faster data rates, lower latency, and higher network capacity. LTE was designed to be a fully packet-switched network from the start, a significant change from the earlier generations that still relied on circuit-switched systems for voice communications. This shift towards packet switching not only improved the efficiency of the network but also allowed for more versatile use cases, from mobile broadband to real-time services. Another key innovation of 4G networks was the use of Orthogonal Frequency-Division Multiple Access (OFDMA), a technology that significantly enhanced spectrum allocation flexibility. This enabled more efficient resource management, since multiple users can be served in orthogonal sub-carriers in the same frequency band.

Despite these advances, mobile networks and wireless standards need to continue its rapid development, since the demand for mobile data continues to grow at a fast pace, driven by changing consumption behaviors and the emergence of new applications. According to Ericsson, mobile network data traffic grew 21 percent between 2023 and 2024, driven by both rising smartphone subscriptions and increasing average data volume per subscription, since there is an increase in consumption of video content. Their report predicts that video traffic will account for 74 percent of all mobile data traffic. Generative Artificial Intelligence (AI) has also showed a strong presence in various device form factors, including smartphones, laptops and wearables. However, since AI is still in its early stages, it is hard to precise its impact on communication and traffic patterns.

The introduction of 5G brought further evolution, addressing the diverse and increasingly demanding requirements of new use cases. The 5G standard introduced a highly flexible architecture, capable of supporting a wide range of service types, each with distinct needs. Enhanced Mobile Broadband (eMBB) demands high data rates, massive Machine-Type Communications (mMTC) require efficient handling of large numbers of low-power, low-throughput devices, and Ultra-Reliable Low-Latency Communications (URLLC) focus on applications where high reliability and low latency are critical.

Ericsson's mobile report also shows that 5G subscriptions are growing at a fast pace, with one quarter of all mobile subscriptions projected to be 5G by the end of 2024 and being expected to overtake 4G as the dominant mobile access technology by subscription for the first time in 2027. And even now, 5G standards continue to evolve, with the 3rd Generation Partnership Project (3GPP) Release 18 marking the first specifications for 5G Advanced. These advancements are expected to further enhance network performance and facilitate the support of emerging applications such as Extended Reality (XR) and cellular vehicle-to-everything (C-V2X).

Despite the impressive capabilities of 5G, the rollout of these networks remains challenging, particularly in dense urban environments. For example, the use of high-frequency spectrum in such areas may require dense deployments of small cells due to higher sensitivity to obstructions and noise. To alleviate these challenges, Non-Standalone (NSA) 5G networks have been employed, leveraging existing 4G core networks while introducing 5G New Radio (NR) Radio Access Networks (RAN). It uses the 4G core network for the control plane while using NR as the radio access technology. In 2023, 75% of 5G deployments in the US were NSA (Rochman et al. 2023). Furthermore, 4G Networks can still meet the requirements of several use cases, such as demonstrated in (Malekzadeh 2023). It can provide up to 1 Gbps of data bandwidth in low-mobility (Alam et al. 2023) and real-world latency similar to some SA and NSA 5G systems (Rochman et al. 2023), (Kousias et al. 2022).

The coexistence of 4G and 5G networks has also given rise to techniques like Dual Connectivity (DC), which allows devices to simultaneously connect to both 4G and 5G base stations, enhancing system capacity. With DC, it is even possible to split the traffic between base stations from different technologies (Pan & Wu 2024). While DC presents opportunities for improving network performance, it also introduces additional complexity, particularly in managing interference between the different technologies.

Moreover, operators are increasingly offering 4G/5G Converged Core Network solutions, integrating the functionalities of both the 4G EPC and the 5GC (Cisco Systems, Inc. 2023). This approach enables both RANs to operate jointly, optimizing operations and scalability, while ensuring a smoother transition to future networks.

Consequently, with the ever-growing demand for higher data rates and increased capacity, the emergence of diverse use cases with varying requirements, the coexistence of numerous devices and technologies, and the fact that the frequency spectrum is both limited and costly, interference remains one of the most significant challenges in wireless communication. Both 4G and 5G are OFDMA-based, hence, even though intra-cell interference is not a problem, the reuse of the spectrum in neighboring cells can cause Inter-Cell Interference (ICI). ICI arises when neighboring base stations use overlapping frequency bands to transmit, leading to users in different cells using the same resource. As a result, devices near the cell edges are more vulnerable to interference, which can degrade network performance.

To address this challenge, 4G networks introduced Inter-Cell Interference Coordination (ICIC) in Release 8, which optimizes the use of spectrum to reduce the impact of interference, being also compatible to 5G networks, since it's also OFDMA-based. One effective ICIC technique is Fractional Frequency Reuse (FFR),

which divides the bandwidth of the cell in sub-regions, each with different frequency reuse patterns and transmitting power. The overall objective is to find a balance between reusing the frequency in neighboring cells while also protecting users close to the cell edges from ICI.

The advantages of FFR techniques extend beyond 4G networks. As modern wireless networks become increasingly complex, many classical FFR techniques stand out due to their relatively low overhead and simplicity of implementation. In 5G networks, which feature dense deployments of small cells and more intensive use of frequency bands, FFR-based solutions offer an effective mechanism for balancing spectrum efficiency with interference mitigation. By dividing the cell into regions with distinct frequency allocations, FFR ensures better performance by reducing interference in edge areas while maintaining efficient spectrum usage in the center.

Although traditional FFR techniques are typically based on static frequency patterns, they provide a solid foundation for more dynamic solutions. These techniques are particularly well-suited for integration with Machine Learning (ML), which can help predict and adapt to changing network conditions. ML algorithms could be used to optimize FFR patterns based on real-time traffic loads, interference levels, or mobility patterns, enabling the network to respond more efficiently to fluctuations in demand. For instance, ML could help adjust the frequency allocation dynamically in response to variations in user behavior, network congestion, or interference scenarios, creating a more flexible approach to interference management. This combination of FFR and ML holds promise for the development of intelligent, self-optimizing networks capable of providing better performance, even in complex and highly dynamic environments.

In this way, FFR techniques continue to be an important tool for both current and future networks, providing a robust solution for interference mitigation while offering potential for more adaptive, intelligent optimization as machine learning is integrated into the network management process.

1.1 State of the Art

Presenting the state of the art is an important step in any scientific work, as it establishes the foundation for the research and highlights the relevance of its novel contributions. The following sections provide the results of a literature review conducted in January 2025, focusing on key topics relevant to this study. By examining existing research, it aims to identify gaps, evaluate current approaches, and contextualize the contributions of this work within the broader scientific discourse.

1.1.1 Inter-cell interference

Despite all the advancements and innovative solutions being constantly proposed for current and future generations of mobile networks, inter-cell interference remains a significant challenge in cellular networks. The frequency spectrum is the most important but also the most expensive resource that is available, making its efficient utilization

critical for meeting the ever-growing demands of modern wireless communication systems.

Inter-cell interference arises mainly due to the overlapping transmission of signals from neighboring cells when sharing the same transmission resource block (a piece of time and frequency), which may degrade the reception quality for users, particularly those at the cell edges. As networks evolve to support higher data rates, lower latencies, and a massive number of connected devices, the management of interference becomes increasingly complex. The densification of networks with small cells, heterogeneous architectures, and advanced technologies like massive MIMO and millimeter-wave frequencies further complicates the issue.

Over the years, researchers have proposed various strategies to mitigate ICI, tailored to address specific causes of interference, optimize different performance metrics, or adapt to the characteristics of the technologies and scenarios deployed within the network. For instance, ICI has been studied in heterogeneous networks (HetNets) (Gulia et al. 2022), ultra-dense networks (UDNs) (Liu et al. 2022), unmanned aerial vehicle (UAV) networks (Burhanuddin et al. 2023), millimeter-wave (mmWave) cellular networks (Wei et al. 2022), and Massive-MIMO systems (Chikha et al. 2024). Some recent studies are outlined below to show the current focus of research in this field.

One of the commonly studied approaches for ICI mitigation is Coordinated Multipoint (CoMP), a cooperative communication technique where multiple base stations collaborate to serve users, reducing interference and enhancing spectral efficiency. The authors (Gulia et al. 2022) provide a comprehensive background on the topic, highlighting that some classical ICIC schemes are not spectrally efficient in HetNets, as the spectrum must be divided among cells of different tiers. Additionally, static or semi-static schemes may result in poor performance in dynamic environments with rapidly changing channel conditions. To address these limitations, they propose a novel approach combining CoMP transmission with time-domain enhanced inter-cell interference coordination (TD-eICIC) to mitigate ICI in HetNets.

The authors (Ahmad et al. 2023) investigate ICI in a heterogeneous network of a public safety network (PSN), UAVs, and an LTE-based railway network (LRN). They propose the use of CoMP to coordinate scheduling and data transmission to reduce interfering signals. The authors (Du et al. 2022) also use CoMP technology alongside a novel ICIC scheme, using the SINR of users in different regions to decide whether to use CoMP or the novel ICIC scheme.

Another strategy that has recently gained attention as a solution to mitigate ICI is cell-free massive MIMO (CF mMIMO), which eliminates cell boundaries (Kassam et al. 2023). In CF mMIMO, users are simultaneously served by multiple distributed access points (APs) through cooperative transmission, improving spectral efficiency and reducing interference (Chen et al. 2022). However, modern urban scenarios are often affected by zones of high user density (hotspots), which brings additional challenges. CoMP, which relies on coordination between base stations and substantial backhaul capacity, faces increased signaling overhead and potential delays as the number of users requiring coordinated service surges in hotspots. Cell-free systems, especially those using a user-centric approach, experience significant fronthaul limitations as data traffic

from numerous users in hotspots needs to be conveyed between the APs and the central processing unit.

Recent research has also proposed innovative solutions for established approaches like user scheduling, cell muting or power control. (Liu et al. 2022) focus on a distributed power control strategy for UDNs. Their method involves adjusting transmission power on specific sub-channels allocated to edge users. The paper by (Tani & Higuchi 2022) proposes an improved method for controlling the activation and deactivation of base stations (BSs) in HetNets, using not only the system throughput but also the activation states of neighboring BSs. That has led to a better convergence rate and improved system performance. (Chikha et al. 2024) leverage Radio Environment Maps (REMs) to enhance user scheduling decisions. This integration of environmental awareness through REMs enables a more informed and strategic approach to user scheduling compared to conventional methods that may not consider real-time interference conditions. (Wei et al. 2022) take into consideration the interference caused by directional antenna arrays in mmWave networks, proposing two ICIC schemes that mute certain Resource Blocks (RBs) at certain BSs based on the path loss incorporating the blockage effect or the directional array gain.

These studies demonstrate that even well-established techniques can be significantly enhanced through the integration of new technologies, algorithms, and intelligent decision-making processes. However, while recent research showcases a variety of interference management approaches, frequency allocation continues to be a key element in improving network performance, specially in OFDMA-based systems. Several solutions aim at improving how the available spectrum is allocated in terms of RBs. The authors (Benni & Manvi 2022) introduce an interference-weighted graph and a Markov Clustering Algorithm (MCA) to group cells based on user interference patterns. Second, a Markov prediction model is used to allocate channels to users within these clusters, optimizing channel selection and minimizing interference. (Burhanuddin et al. 2023) use deep reinforcement learning (DRL) to create a dynamic cell muting and time-frequency RB allocation scheme. It aims at maximizing the throughput of terrestrial users (TUEs) while guaranteeing the quality of service for UAVs, which often require reliable command and control links. (Su et al. 2023) present a novel user-centric resource allocation algorithm for heterogeneous UDNs. They use spectral clustering to form sub-networks depending on interference levels and then orthogonally allocate RBs within each sub-network and perform spatial multiplexing of RBs between sub-networks.

1.1.2 Scenarios with hotspots

Hotspots, or zones with high user density, appear frequently in studies addressing inter-cell interference, often in the context of heterogeneous networks. HetNets mitigate interference in such scenarios by deploying small cells to increase capacity and offload traffic. The authors (Konishi 2013) present an extensive analysis of HetNets and how pico cells can be used to manage traffic in hotspot areas. They highlight the effectiveness of deploying pico cells but also discuss its many challenges, such as the additional interference introduced by these pico cells and that not all pico cells might be actively

serving UEs. In order to mitigate these drawbacks, they evaluate the use of CRE and eICIC, which also come with its own advantages and challenges. (Yang et al. 2021) address the challenge of modeling and analyzing the non-uniform distribution of users and small base stations (SBSs) clustered around hotspots, as users are typically clustered around known hotspots like shopping centers or universities. They also investigate the use of FFR and CoMP to mitigate co-tier interference, which becomes a significant issue in dense small cell deployments. Their simulations show that FFR can significantly improve coverage and capacity in HetNets, and the Poisson Cluster Process (PCP)-based model better represents the spatial correlation between users and small cells, aiding in hotspot optimization. However, they also find that increasing the number of mmWave SBSs per cluster reduces coverage probability due to higher interference in denser deployments. An interesting approach was studied by (Kimura & Ogura 2021) to adapt to hotspots using UAV-mounted aerial base stations (ABSs). They proposed a distributed 3D deployment method where ABSs dynamically adjust their positions to expand the network's capacity in response to sudden traffic increases due to hotspots. While their work includes several simplified assumptions, such as unlimited ABS resources, simulation results demonstrated that this method significantly improves communication quality in hotspots and effectively adapts to dynamic network changes.

However, network densification is not always feasible or necessary. The deployment of additional small cells comes with significant costs, both in terms of installation and ongoing maintenance, which can make operators hesitant to adopt this approach. This economic and logistical challenge has contributed to the reduced momentum of technologies like mmWave, which rely heavily on dense network deployments to overcome their limited range and penetration capabilities. As a result, alternative strategies that do not depend on extensive densification are an interesting direction to address ICI in hotspot scenarios. For example, the previously mentioned work from (Chikha et al. 2024) proposed a resource allocation low-complexity ICIC solution based on additional information from REMs on a scenario with a hotspot area between two macro cells, that is, on the cell edges. Their solution was able to marginally improve the performance of the UEs outside the hotspots and greatly improve the performance of the UEs inside the hotspots.

The main subject of this thesis is the impact of hotspots in mobile networks, specifically concerning inter-cell interference. The focus is on low-complexity solutions to dynamically respond to increasing ICI from unexpected hotspots. The following sections will review the literature on FFR-based ICIC techniques, for their potential as a foundation for dynamic solutions. The objective is to discuss how FFR has been recently employed and also how ML can be used to adaptively manage frequency allocation and respond to the dynamic nature of hotspot scenarios.

1.1.3 Fractional Frequency Reuse and Dynamic FFR Solutions

FFR was first introduced during the standardization process for LTE-A systems. It has rapidly established itself as an efficient technique for mitigating inter-cell interference and remains a frequent topic in the literature. Recent studies have explored FFR parameters

and applied it in various contexts, creating solutions for new scenarios and use cases.

FFR has been effectively utilized in UDNs and NTNs, addressing challenges such as interference management, resource allocation, and dynamic network conditions. (Lam et al. 2023) proposed a modified FFR scheme where each BS works independently in a ultra-dense network, improving coverage probability and efficiency without fixing the number of RBs in each group of users. Similarly, (Lam & Tran 2021) model and examine the performance of FFR in UDNs, proposing a generalized model of FFR with users being classified into more than two groups. Their results showed that FFR was able to improve user performance without increasing the BS's power consumption. (Shahid et al. 2024) extended FFR applications to NTNs, specifically multi-beam satellite systems, by introducing the Partially Strict Frequency Reuse (Strict FR) scheme to mitigate inter-beam interference. It was able to enhance overall capacity when compared to the Strict FR.

Even classic FFR algorithms, such as Strict and Soft FR, continue to find relevance in modern network designs. For instance, a fairness-aware adaptation of Soft FR, as demonstrated by (Ming et al. 2022), has shown how the Soft FR can be refined to balance user fairness and spectral efficiency in OFDMA systems. (Susanto et al. 2024) also proposed a Soft FR to improve SINR in a multi-cell scenario with microcells for in-band D2D communications. Previously mentioned work by (Lam et al. 2023) introduces a novel FFR scheme where each BS randomly selects from all available RBs which ones will serve its center and edge users, without limiting which RBs are available to each group. Their analytical results show a significant improve in coverage probability. (Nafees et al. 2021) introduced a variable-height UAV network deployment strategy that employs a static FFR technique, dividing frequency bands into center and edge regions. The study also optimizes the SINR threshold that defines user classification within each cell, enhancing spectral efficiency and improving coverage and throughput. (Ying & Wang 2022) demonstrated that while Strict FR minimizes co-channel interference by allocating distinct frequency resources for cell edges, Soft FR enhances spectral efficiency by overlapping resource allocation with varying power levels. (Seo et al. 2023) investigated the optimal coverage of the full frequency reuse region in cellular networks, focusing on its relationship with base station power. Their analysis demonstrate that, unlike the findings from previous studies, scaling the power of a single BS results in a non-decreasing optimal full FR coverage within that cell. This highlights the ongoing value of studying classical FFR techniques, as their parameters may involve subtle complexities that can reveal new insights and opportunities for optimization. Additionally, (Luu et al. 2024) show that FFR can also adapt to emerging multiplexing strategies, such as NOMA. They investigate the combination of FFR, TDD and NOMA to enhance uplink data rates by allowing center and edge users to share the same sub-band via power-domain NOMA.

FFR solutions tailored for dynamic environments have gained significant attention in recent research, demonstrating how classical methods can be adapted to meet changing network conditions. Notably, the advent of machine learning has facilitated the development of algorithms that dynamically react to varying interference levels, traffic demands, and user distributions, thereby enhancing network performance.

Among different parameters, power control has been a frequent topic in interference

mitigation. (Lam et al. 2022) demonstrated the use of Deep Q-Networks (DQNs) to optimize power allocation dynamically for cell-center and cell-edge regions in a modified version of the Strict FR. Their modified version allows all BSs to share the cell-edge sub-band and users are classified in groups based on their distance to the serving BS. Their results indicate a significant improvement when compared to the Full FR approach. (Safdar et al. 2022) introduced a dynamic power allocation scheme based on game-theory for irregular multicellular networks. The proposed scheme optimizes the power level of each FFR sub-band independently while penalizing if a region is using excess power. This decentralized approach was able to greatly enhance overall cell data rate.

Another parameter well-suited for intelligent control in FFR techniques is user classification. (Onim et al. 2022) presents a dynamic FFR scheme that periodically calculates the SINR threshold using Otsu's and entropy methods, dynamically updating the threshold based on subscriber distribution, enabling dynamic partitioning of inner and outer regions. Similarly, (Kihato et al. 2022) proposed a dynamic SINR threshold selection method using Native Integral Ratio (NIR) and ISODATA techniques. This approach periodically updates the SINR threshold using real-time subscriber reports. The periodic updates ensure that the threshold remains aligned with fluctuating network demands, improving fairness and edge throughput.

Recent studies have also investigated dynamic bandwidth allocation. (Ullah et al. 2022) introduced an auction mechanism-based Sectored FFR (AMS-FFR) scheme to dynamically allocate bandwidth sub-carriers based on user bids that reflect real-time traffic demands. Similarly, (Ullah et al. 2021) employed the Hungarian mechanism to dynamically allocate bandwidth in Sectored FFR systems. Both works demonstrate the effectiveness of dynamic bandwidth allocation to reduce ICI, improving throughput, and enhancing resource utilization compared to static FFR methods. Moreover, the authors (Guo et al. 2024) presented a resource allocation scheme that dynamically clusters small cells to subsequently allocate sub-channels to users considering a system with FFR, ensuring fairness among users.

Lastly, some works have explored the dynamic and simultaneous control of multiple parameters to enhance system performance and better adapt to dynamic network conditions. (Sallakh et al. 2014) presented a multi-parameter Q-Learning solution that learns the best location of the edge-center boundary (ECB) and the appropriate power levels at the cell center and edge to minimize ICI. Their results show that static solutions suffer from dynamic behaviors in the network and controlling more than one parameter results in better performance. However, even though they acknowledge and reduce the number of actions in the Q-Learning, their action and state space is still large, which could hinder its feasibility for real-time implementations in highly dynamic environments.

The authors (Marinescu et al. 2018) also proposed an approach for dynamic frequency reuse using multiple parameters. In their approach, cell agents utilize local neural networks to estimate the best performance attainable for all three available bandwidth choices. A coordinator agent chooses the option that maximizes the global network performance and every cell will use the winning bandwidth. Sequentially, each cell chooses the threshold that maximizes its own gain. They evaluate their solution in various user distributions, showing improvement for edge users when comparing to full frequency

reuse and static configurations.

In previous works, the author of this thesis performed an extensive comparison of the most common FFR techniques in the literature (de O. Nóbrega et al. 2018), identifying the Strict Frequency Reuse as a good compromise between complexity and performance. This study was followed by a factorial analysis of the Strict FR, which demonstrated that the bandwidth allocation and SINR threshold have strong effects on system performance (do Rêgo et al. 2018). The last work preceding this doctoral thesis proposed a single-agent solution that controls the SINR threshold of the Strict FR, providing substantial SINR gain for all users and indicating that the algorithm could achieve greater flexibility and efficiency by extending the agent's action to multiple parameters (do Rego & de Sousa Jr. 2021).

After reviewing this extensive body of work, it is clear that inter-cell interference remains a persistent challenge in mobile networks, with resource allocation at its core. Among the various approaches to mitigate ICI, well-established solutions like FFR have demonstrated their adaptability to modern network demands, particularly through integration with dynamic technologies such as machine learning. While FFR continues to be widely studied, even in its classical forms, the increasing prevalence of dynamic scenarios, such as urban hotspots arising from unexpected events, underscores the need for real-time, adaptive solutions. Several studies propose intelligent FFR-based approaches, but relatively few focus on simultaneously controlling multiple parameters. Moreover, solutions that rely on dedicated training, operate within large state spaces or employ a centralized operation with heavy signaling may struggle to meet the real-time retraining demands posed by rapidly changing environments. This gap highlights the need for lightweight, efficient, and highly adaptive methods capable of addressing these modern challenges.

The key differences between the dynamic FFR solutions discussed in this section and those proposed in this thesis are presented on Table 1.1, highlighting that its main contribution consists of a framework that controls two key FFR parameters with a multi-agent approach that aims at lower complexity.

1.2 Research questions

The reviewed literature shows that modern urban scenarios have become increasingly dynamic, which drove the scientific community to explore solutions such as heterogeneous networks (HetNets) and ultra-dense networks to address inter-cell interference. However, network densification entails significant costs, and operators are often reluctant to deploy a large number of small base stations due to high installation and maintenance expenses. As a result, recent research has also explored dynamic or coordinated interference mitigation strategies built upon classical techniques like FFR.

Although FFR has been proven effective in dynamic approaches, controlling a single parameter may be insufficient to adapt to highly dynamic scenarios. For instance, the unexpected emergence of hotspots can drastically change user distribution and create a sudden surge in demand, which is difficult to address without a dynamic solution capable of real-time adaptation. Furthermore, few studies have focused specifically on hotspot

Table 1.1: State of the art and thesis contribution comparison.

Paper	Deployment	FFR Scheme	Approach	Parameters
(Ullah et al. 2021)	Irregular multi-cell	Sectored FFR	Auction mechanism	BW allocation
(Lam et al. 2023)	UDN	Modified FFR	Rule-based	BW allocation
(Ullah et al. 2022)	Irregular multi-cell	Sectored FFR	Hungarian method	BW allocation
(Safdar et al. 2022)	Irregular multi-cell	Sectored FFR	Game theory	Power allocation
(Ying & Wang 2022)	Multi-cell	Soft and Strict FR	Cognitive radio	Power allocation
(Lam et al. 2022)	Multi-cell indoor	Modified Strict FR	ML: Deep QL	Power allocation
(Nafees et al. 2021)	Two-tier UAV	Strict FR	ML: MAB	SINR threshold
(Onim et al. 2022)	Multi-cell	Strict FR	Otsu's and entropy	SINR threshold
(Kihato et al. 2022)	Multi-cell	Strict FR	NIR and ISODATA	SINR threshold
(Shahid et al. 2024)	NTN	Modified Strict FR	Static	-
(Luu et al. 2024)	Irregular multi-cell	NOMA UL FFR	Static	-
(Guo et al. 2024)	Multi-tier	Modified FFR	Optimization	BW allocation
(Marinescu et al. 2018)	Multi-cell with various distributions	Strict FR	ML: Multi-agent Neural Network	SINR threshold BW allocation
(Sallakh et al. 2014)	Multi-cell	Soft FR	ML: Q-Learning	Power allocation Edge boundary
This thesis	Multi-cell with dynamic hotspots	Strict FR	ML: Multi-agent Hierarchical QL	SINR threshold BW allocation

scenarios using FFR-based approaches, particularly those that do not rely on deploying additional small base stations.

In response to these challenges, this thesis investigates the use of Reinforcement Learning (RL) to dynamically control multiple FFR parameters in scenarios characterized by the unexpected emergence of user hotspots. Reinforcement learning is a model-free, lightweight technique, well-suited for rapidly changing environments, due to its ability to learn and adapt in real time. This work proposes a hierarchical multi-agent architecture, in which two RL agents jointly optimize bandwidth allocation and user classification via the Reference Signal Received Quality (RSRQ) threshold, aiming to mitigate ICI in a scalable, low-signaling, and low-complexity manner.

Consequently, this thesis addresses the following research questions:

- What are the individual and combined impacts of bandwidth allocation and user classification on throughput and interference in FFR-based ICIC, particularly in hotspot scenarios?
- How can reinforcement learning be applied to dynamically control multiple FFR parameters to effectively mitigate inter-cell interference in scenarios with emergent hotspots?
- To what extent does a multi-agent reinforcement learning approach improve system performance metrics, such as user throughput and SINR, compared to static configurations and single-agent baselines?
- What are the key challenges associated with developing RL-based solutions for interference coordination in mobile networks?
- How does the use of different learning strategies (e.g., Q-Learning vs. Multi-Armed Bandits) impact performance and computational complexity in the context of dynamic ICIC?

1.3 Societal and Environmental Impact

The increased efficiency of mobile communication systems, achieved through higher data transmission rates, leads to shorter transmission times and energy savings. This improvement has a direct environmental benefit, as it reduces the energy consumption of communication infrastructure and the frequency of battery recharges in mobile devices. Consequently, this contributes to Sustainable Development Goal (SDG) 13: Climate Action, by helping to mitigate the environmental footprint of digital technologies. Furthermore, the results presented in this work align with SDG 9: Industry, Innovation, and Infrastructure, by fostering the adoption of artificial intelligence as a tool for innovation within the telecommunications industry. By integrating advanced AI techniques into network systems, this research supports the development of smarter, more sustainable digital infrastructure.

1.4 Thesis outline

The remainder of this thesis is organized as follows. Chapter 2 presents the theoretical background on interference in mobile networks, with a focus on co-channel interference and Fractional Frequency Reuse. Chapter 3 describes the characterization of the research problem and the modeling of the simulation environment, along with an exploratory analysis that supports the hypothesis of joint parameter control. Chapter 4 introduces the proposed hierarchical reinforcement learning framework, detailing its architecture, learning strategies, and implementation. Chapter 5 presents the results of the simulation campaigns, evaluating the performance of the proposed solution across different scenarios and learning algorithms. Finally, Chapter 6 summarizes the main findings, discusses the contributions and limitations of the work, and outlines directions for future research.

Chapter 2

Interference in mobile networks

Interference is a persistent problem in mobile networks that has been widely studied in various scenarios, across different technologies, arising from multiple sources, and affecting a wide range of users. As wireless communication technologies evolve, deployments have become more complex, with increasing demand, more diverse user and base station distributions, and the coexistence of multiple radio access technologies (RATs). These improvements enable higher data rates but can also lead to higher interference.

The general term interference is typically used in mobile networks to describe the presence of undesired signals (at the same frequency and time) added to a signal being intentionally received. If not managed, interference can severely degrade system performance. In general, mobile networks experience two main types of interference: intra-cell interference and inter-cell interference (Hamza et al. 2013).

Figure 2.1 presents how interference in mobile networks can be categorized. The primary focus of this work is inter-cell interference, highlighted by yellow blocks in the figure. However, in the following sections, all categories and subtypes will be discussed to provide context and to position how inter-cell interference differs from other forms of interference. The discussion will then continue on the impact of inter-cell interference on mobile networks and its main challenges. Finally, this chapter will conclude with the introduction of Fractional Frequency Reuse as the main subject of investigation for proposed solutions.

2.1 Intra-cell interference

As the name suggests, intra-cell interference occurs within the same serving cell and is primarily caused by signals sharing the same or overlapping frequency resources within the cell. This may result from power leakage or the use of adjacent frequency channels. When LTE was introduced, OFDMA was chosen as the downlink multiple access technology for its physical layer. This decision significantly reduced intra-cell interference by ensuring that multiple users could be served on orthogonal subcarriers within the same frequency band, thereby improving spectral efficiency (Hamza et al. 2013). 5G's radio access technology, New Radio, also adopts OFDMA for similar reasons, but 5G and beyond networks have introduced new technologies like NOMA,

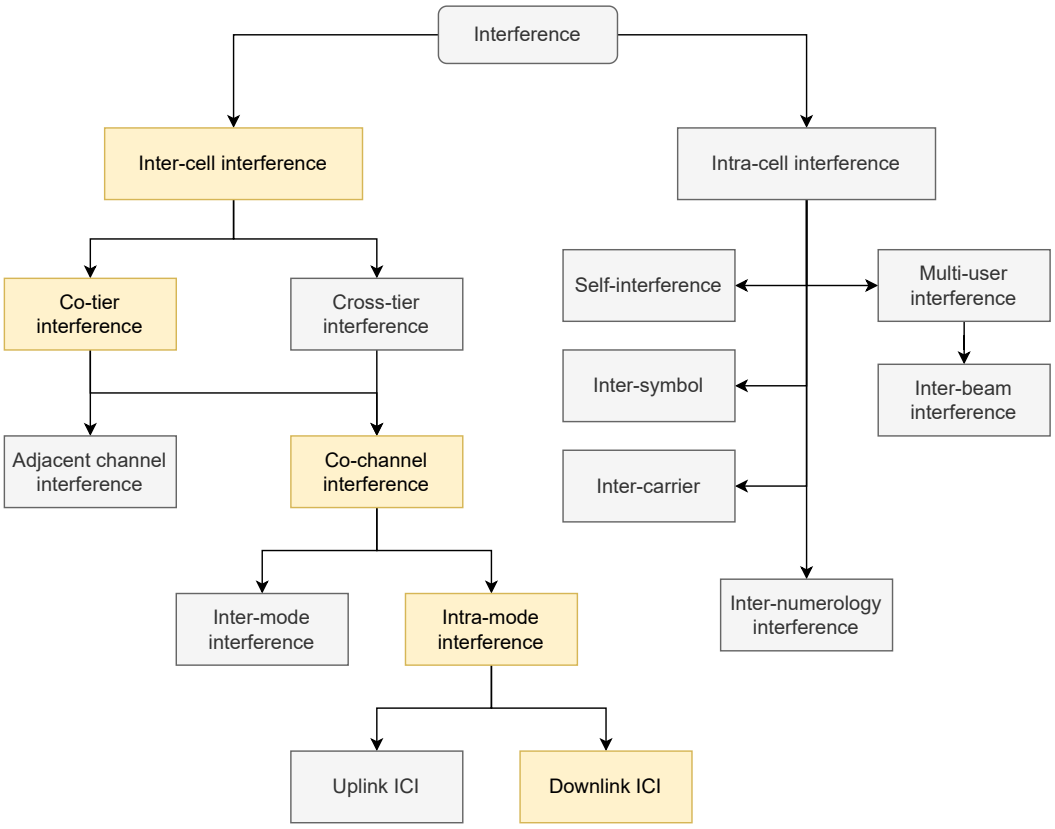


Figure 2.1: Main categories of interference highlighting the focus of this work.

which rely on non-orthogonal resource sharing and are inherently more prone to intra-cell interference (Trabelsi et al. 2024).

Self-interference is a type of intra-cell interference that occurs when a device's transmitted signal interferes with its own reception. This issue is less significant in systems that use half-duplex communication, such as LTE and 5G, which employ frequency-division duplexing and time-division duplexing to isolate uplink and downlink transmissions. However, full-duplex communication, where uplink and downlink transmissions occur simultaneously in the same frequency band, has gained more attention for future networks. Full-duplex systems have the potential to improve spectral efficiency but face the critical challenge of self-interference from the transmitted signal. Additionally, self-interference can result from hardware imperfections, such as non-linear impairments, or from signal reflections due to multipath propagation. Numerous works have focused on mitigating self-interference in full-duplex systems.

Another type of intra-cell interference comes from multiple users sharing resources such as frequency, time, or spatial dimensions within the same cell. This is referred to as multi-user interference (MUI). In OFDMA systems, MUI can occur due to frequency shifts between adjacent subcarriers. However, it is more prominent in NOMA systems,

where users share the same frequency and time resources. Multi-user interference can also take the form of inter-beam interference in Massive MIMO and mmWave networks. Beamforming is employed in these systems to improve spectral efficiency and increase data rates by directing narrow beams toward individual users, a technique known as spatial division multiplexing. However, imperfect beam separation, overlapping beams, or dense user environments can lead to interference between beams.

In OFDMA-based systems like LTE, other forms of intra-cell interference include inter-symbol and inter-carrier interference. Inter-symbol interference occurs when symbols overlap in the time domain due to delayed signal arrivals, while inter-carrier interference arises from loss of orthogonality between subcarriers in the frequency domain. These types of interference are typically not major limiting factors in LTE because the system design minimizes them. For example, LTE uses cyclic prefixes to absorb signal delays and OFDMA ensures orthogonality between subcarriers. Nevertheless, specific scenarios can cause higher interference of this nature. For instance, high-mobility environments can cause significant Doppler shifts that lead to frequency misalignment, causing inter-carrier interference. Similarly, environments with very large delay spreads can cause symbols to extend beyond the cyclic prefix, resulting in inter-symbol interference.

Finally, inter-numerology interference is a unique challenge in 5G systems. LTE is a single-numerology system, using a fixed subcarrier spacing of 15 kHz for the entire network, for both uplink and downlink, simplifying network design and reducing interference. In contrast, 5G NR introduces multiple numerologies, enabling it to support diverse use cases with varying requirements for latency, bandwidth, and mobility. While this flexibility improves network adaptability, it also introduces inter-numerology interference due to the coexistence of different subcarrier spacings, symbol durations, and power levels. Spectral leakage from overlapping subcarriers, mismatched symbols in the time domain, and uneven power levels between numerologies are common causes of this interference. Mitigating inter-numerology interference requires careful resource allocation, sufficient guard bands, and advanced scheduling techniques.

2.2 Inter-cell Interference

As opposed to intra-cell interference, inter-cell interference occurs when resources are shared between neighboring base stations. These resources span multiple domains, but the most commonly shared resource is the frequency spectrum. The reuse of frequency bands across cells is fundamental to maximizing spectral efficiency, but it can also lead to interference, specially for users along the cell edges.

In heterogeneous networks (HetNets), where different types of cells (e.g., macro and small cells) coexist, inter-cell interference is classified into co-tier interference and cross-tier interference (Agarwal et al. 2022). Co-tier interference happens between cells of the same tier, such as two macro cells, while cross-tier interference occurs between cells from different tiers, such as a macro cell and a small cell. Although the fundamental principles of inter-cell interference apply to both co-tier and cross-tier scenarios, HetNets introduce unique challenges. For instance, a dense deployment of small cells may lead

to co-tier interference among these cells, while also creating a ring of interfering users connected to the macro cell, exemplifying cross-tier interference. However, since HetNets are not the focus of this work, cross-tier interference will not be considered in subsequent sections.

Inter-cell interference can generally be caused by different cells sharing the same communication channels at the same time, known as co-channel interference (CCI), or by interference originating from neighboring channels, called Adjacent Channel Interference (ACI). Modern networks have implemented robust measures to mitigate ACI, such as careful allocation of guard bands, channel planning, and power control strategies. As a result, most ACI in practical systems is due to hardware imperfections, such as poor filtering of adjacent channels.

2.2.1 Co-channel interference

Co-channel interference is the most extensively studied type of interference in the scientific literature due to its significant impact on system performance in cell-based wireless networks. It occurs when neighboring cells use the same frequency band, which is a direct consequence of frequency reuse. While frequency reuse enhances spectral efficiency and maximizes resource utilization, it introduces interference that can degrade performance, particularly for cell-edge users. To better understand the nature of co-channel interference, it can be classified into inter-mode interference and intra-mode interference.

Inter-mode interference

Inter-mode interference happens between different transmission modes, such as uplink-to-downlink or vice versa. This type of interference is less significant in LTE due to the use of Frequency Division Duplexing (FDD) and Time Division Duplexing (TDD), which separate uplink and downlink transmissions into distinct frequency bands or time slots. Additionally, neighboring base stations in TDD systems are synchronized to avoid overlapping uplink and downlink transmissions.

However, FDD systems may experience uplink-to-downlink interference if resources are not carefully managed, such as misconfiguration of the guard period. This can lead to overlapping signals within the same frequency band. Similarly, FDD and TDD networks can interfere at cell edges when synchronization is insufficient, causing misaligned transmissions. Dynamic TDD systems, in contrast, suffer from additional inter-mode interference due to neighboring base stations dynamically changing their TDD patterns. In 5G multi-numerology networks, inter-mode interference may also occur due to mismatched subframe configurations or incompatible transmission modes, highlighting the complexities of managing diverse configurations within a shared spectrum. Although inter-mode interference is generally less critical than intra-mode interference, it remains a factor in modern network design, especially in heterogeneous deployments or advanced systems like 5G and beyond.

Intra-mode interference

Intra-mode interference, on the other hand, is more frequent and can significantly impact system performance. This type of interference occurs within the same transmission mode, either in the uplink or downlink direction. In the uplink, intra-mode interference happens when the base station experiences interference from users in neighboring cells transmitting on the same frequency. In the downlink, interference occurs when a user in one cell is affected by transmissions from neighboring base stations operating on the same frequency, degrading the user's signal quality.

An example of a specific type of intra-mode interference is inter-beam interference, which is particularly relevant in systems like Massive MIMO that employ beamforming. Inter-beam interference can occur in both uplink and downlink scenarios when neighboring base stations have overlapping beams or fail to achieve sufficient beam separation. This issue is aggravated in dense deployments, where beams directed at closely positioned users may overlap.

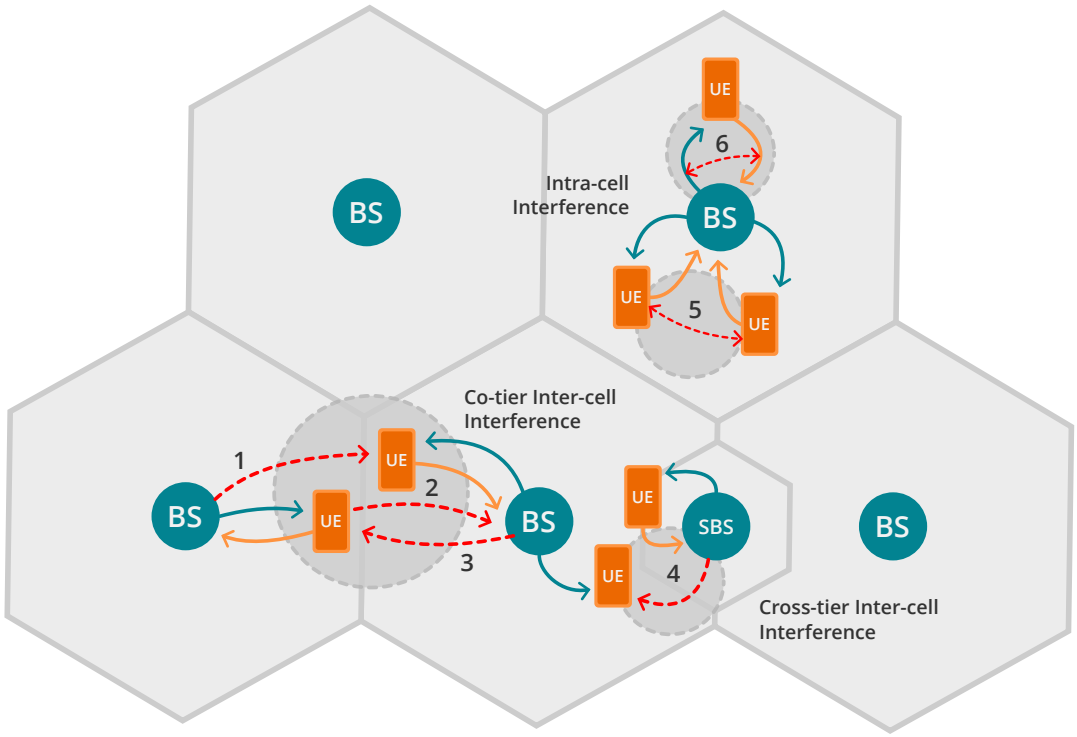
Figure 2.2 exemplifies a few of the discussed interference types, highlighting the aggressor and the victim of the interference. Each number showed in the figure is associated to an index on the table, outlining for each case who is the source of the interference (the aggressor), who suffers the interference (the victim), and what is the corresponding interference type. For example, the indexes 1, 3, and 4 refer to ICI in the downlink direction, where the aggressor is a base station and the victim is a UE.

This work will focus on downlink co-tier intra-mode co-channel interference, specifically when users experience interference from neighboring base stations operating on the same frequency due to the reuse of frequency bands. Since frequency reuse is essential for maximizing spectral efficiency in mobile networks, understanding this type of interference is crucial for achieving efficient use of the spectrum and maintaining suitable performance. However, spectrum reuse also introduces significant challenges, as seen with co-channel interference. The next section explores the reuse of the frequency spectrum in cellular systems, highlighting some of its challenges and benefits.

2.3 Frequency reuse

The frequency spectrum is a scarce and expensive resource, making its efficient utilization critical for meeting the growing demands of mobile networks. Frequency reuse, the practice of deploying the same frequency channels in multiple cells within the network, is a key component of wireless communication systems. This approach allows operators to significantly increase network capacity and coverage without purchasing additional spectrum allocations, which are often not available.

Modern networks tend to apply frequency reuse to a high degree, driven by the need to maximize system capacity. In fact, frequency reuse has been a fundamental principle in cellular networks since their conception, but its implementation has evolved over time. Early systems like GSM employed higher reuse factors, such as reuse 3 or reuse 7, where the same frequency was allocated to cells separated by a greater distance. This reduced interference but limited spectral efficiency, as a significant portion of the available



Index	Aggressor	Victim	Type
1	Macro BS	Neighboring UE	Co-tier dowlink ICI
2	Macro UE	Neighboring BS	Co-tier uplink ICI
3	Macro BS	Neighboring UE	Co-tier dowlink ICI
4	Small cell BS	Macro UE	Cross-tier dowlink ICI
5	Macro UE	UE from same cell	Multi-user interference
6	Same UE	Same UE	Self-interference

Figure 2.2: Examples of inter-cell interference and intra-cell interference.

spectrum was not utilized within each cell. These higher reuse factors were well-suited to the relatively low capacity demands and sparse deployments of earlier networks.

With the introduction of LTE, the standard adopted reuse factor 1 as the default approach. In this model, all adjacent cells have access to the entire bandwidth, enabling maximum spectral efficiency, as illustrated in Figure 2.3.

This decision was driven by the need to accommodate the growing demand for data services and the increasing density of mobile networks. Full frequency reuse allows for wider channels and greater flexibility in resource allocation, which can lead to higher capacity in many deployment scenarios. This principle is fundamentally rooted in Shannon's Capacity Theorem, which defines the theoretical maximum capacity of a communication channel in the presence of noise, as shown on Equation (2.1).

$$C = B \cdot \log_2(1 + SNR). \tag{2.1}$$

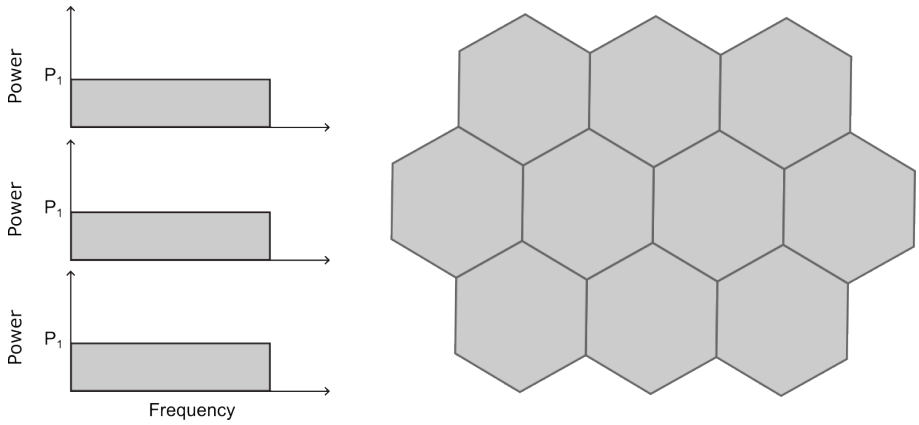


Figure 2.3: Cellular system with full frequency reuse, i.e., reuse 1.

This theorem demonstrates the direct relationship between bandwidth (B in Hz) and the achievable capacity (C in bps). The capacity grows linearly with the bandwidth, while it grows only logarithmically with the Signal-to-noise ratio (SNR). This difference may often justify higher frequency reuse, since the growth in capacity can compensate the reduction of Signal-to-interference-plus-noise ratio (SINR) (Cavalcanti, Maciel, Jr. & Silva 2018).

However, real-world networks face additional impairments, such as inter-cell interference, that complicate this idealized scenario. Interference power is highly variable and it depends on several factors, such as the interference source, user distribution, distance to the base station, frequency reuse patterns, transmission power of transmitting and interfering devices, among others.

For example, Equation (2.2) is a simplified extension of the Shannon's Capacity Theorem, as explained by (Cavalcanti, Maciel, Jr. & Silva 2018), that incorporates factors such as spectrum utilization efficiency (α), normalized user distance from the cell center (ρ), edge SINR (γ_b), and channel coding efficiency (β). This formulation emphasizes that achieving optimal capacity depends not only on bandwidth but also on efficient spectrum utilization and interference management.

$$\bar{C} = 2\alpha B \int_0^1 \log_2(1 + \frac{\gamma_b \rho^{-n}}{\beta})\rho d\rho. \tag{2.2}$$

However, reuse patterns that significantly limit the available bandwidth are no longer practical in modern networks. Conversely, aggressive reuse of the spectrum can lead to poor performance due to substantial co-channel interference, particularly for users located at cell edges. These users are farther from their serving base stations and closer to interfering signals, making them more susceptible to overlapping transmissions. Thus, this issue becomes more critical in networks with dense user deployments, especially if users are clustered in random hotspots.

Modern interference management techniques, such as beamforming, Inter-Cell Interference Coordination (ICIC), and power control strategies, are employed to mitigate

these challenges. Additional strategies like fractional frequency reuse (FFR) have been proposed to balance the trade-off by allocating different frequency resources to cell-center and cell-edge users, reducing interference while maintaining efficient spectrum utilization.

2.4 Challenges associated to co-channel interference

Co-channel interference is a complex and difficult-to-manage issue that often leads to degraded performance in wireless networks. Its impact varies due to multiple factors, including network topology, deployment strategies, and environmental conditions. In real-world scenarios, CCI can be further aggravated by design choices such as the dense deployment of base stations or scenario-dependent factors like high user density in urban areas. Additionally, the nature of wireless communication itself, including non-ideal propagation effects, user mobility, and dynamic traffic patterns, makes CCI a persistent challenge that requires sophisticated management techniques. This section explores key aspects that exacerbate CCI and increase its complexity in modern wireless networks.

2.4.1 Mobility and dynamic environments

Mobility can change how interference changes over time, especially in high-speed environments such as vehicular networks, high-speed trains, and aerial mobile networks. As users move between cells, they introduce time-varying interference patterns that challenge static interference management techniques.

A major issue in high-mobility environments is the frequent need for handovers, which leads to sudden changes in the interference landscape. When a user transitions from one base station to another, the serving cell must dynamically adjust its power control and scheduling policies to avoid degradation in SINR at the cell edge. Frequent handovers also result in increased control signaling overhead, consuming network resources that could otherwise be allocated to data transmission.

Another significant challenge is the Doppler effect, which introduces a frequency shift given by:

$$f_d = \frac{v f_c}{c}, \tag{2.3}$$

where f_d is the Doppler frequency shift, v is the velocity of the mobile user, f_c is the carrier frequency, and c is the speed of light. This effect leads to channel estimation errors, reduced orthogonality in OFDM systems, and difficulty in implementing effective interference cancellation techniques. In 5G systems, which may operate in mmWave bands, Doppler shifts are even more pronounced, degrading beamforming accuracy and increasing inter-beam interference.

Additionally, mobility impacts adaptive resource allocation algorithms, as traditional static power control and scheduling techniques become ineffective in highly dynamic environments. Interference conditions must be estimated in real time, demanding low-latency channel state information (CSI) feedback mechanisms. However, real-time

CSI reporting increases uplink signaling overhead, particularly in MIMO systems, where a large number of antennas must maintain interference coordination across multiple spatial layers.

2.4.2 Dense wireless networks

Modern cellular networks increasingly rely on dense deployments of base stations and small cells to meet the growing demand for high-capacity services. The deployment of ultra-dense and heterogeneous networks increases network capacity but significantly worsens co-channel interference due to the proximity of transmission points and coexistence of cells from different tiers. In UDNs, BSs are densely deployed to greatly reduce average inter-site distance (ISD) when compared to traditional macrocell deployments, reducing path loss. In fact, the number of BSs in UDNs might be higher than the number of connected users (Alzubaidi et al. 2022). As a result, the interfering power from neighboring BSs is often comparable to the desired signal power, particularly in scenarios with full frequency reuse.

The downlink SINR at a UE in a fully reused network can be expressed as a simple extension of the SINR calculation, as demonstrated in Equations (2.4) and (2.5).

$$\text{SINR} = \frac{P_s G_s}{I + N}, \quad \text{and} \tag{2.4}$$

$$I = \sum_{i=1}^{Z} P_i G_i. \tag{2.5}$$

where P_s and G_s are the signal power and channel gain from the serving BS, I is the sum of interference power from all neighboring BSs, and N is the thermal noise power. In a dense deployment where the number of interferers Z grows, the interference power I increases. In UDNs, both Z and P_i are bigger, increasing the overall interference power.

Additionally, the existence of small cells in both UDNs and HetNets introduce cross-tier interference, where the difference in power between macrocells and small cells cause higher interference. Small cells, often deployed with irregular distribution (e.g., inside buildings or on streetlamps), create unpredictable interference patterns. In downlink transmissions, small cell users suffer from aggressive interference from macro BSs, while in the uplink, small cell users can create interference for macro users due to proximity and non-ideal power control.

2.4.3 Coordination and management complexity

Strategies for mitigating CCI often require coordination between neighboring cells, as the primary cause of CCI is the reuse of frequency bands among adjacent base stations. Centralized coordination schemes, where a central controller collects information from multiple base stations and optimizes interference mitigation strategies, offer the potential for global interference minimization. However, they face major scalability and latency constraints, since overhead signaling is a requirement for cell coordination (Alzubaidi

et al. 2022). Real-time information exchange between BSs, such as CSI, power control parameters, and scheduling decisions increases fronthaul and backhaul traffic, making real-time coordination difficult in latency-sensitive applications.

Moreover, optimal interference coordination often requires solving non-linear optimization problems, particularly in scenarios with dynamic user distributions, appearance of hotspots, or dense deployment of small base stations. These problems may involve complex constraints such as determining the optimal transmission power for multiple small base stations while maintaining coverage, balancing the load by assigning users to base stations while mitigating interference from neighboring cells, or allocating frequency bands in a way that minimizes spectral overlap and maximizes overall throughput. These optimization problems are computationally intensive, often requiring advanced techniques such as convex optimization, game-theoretic models, or machine learning-based approaches to find near-optimal solutions in real time.

While centralized coordination offers the advantage of global interference minimization, its reliance on extensive information exchange leads to scalability issues and increased signaling delays. In contrast, distributed coordination techniques reduce backhaul load by allowing base stations to make local interference management decisions. However, these decentralized approaches often suffer from suboptimal interference mitigation, as base stations operate with incomplete or outdated information about network conditions. This trade-off between coordination complexity and interference management efficiency remains a challenge in designing scalable CCI mitigation strategies for next-generation wireless networks.

2.4.4 Spectrum scarcity

The radio spectrum is the main resource associated to wireless communications. It is a natural and reusable resource that needs efficient management to better support the coexistence of different services. It is finite, ranging from 3 kHz to 3000 GHz, and it is commonly categorized in multiple bands. Table 2.1 presents the nomenclature of these bands according to ITU (ITU-R 2015). Inside this spectrum, some bands are particularly saturated. For example, the decimetric waves (UHF), from 300 to 3000 MHz, are highly disputed for mobile communications due to its longer propagation range, better penetration and diffraction around buildings. Moreover, depending on the bands considered, the service and the altitude, which relates to the existence or not of Line of Sight (LoS), the occupancy of the spectrum may vary (Fahim Raouf et al. 2023). Hence, even though the entire radio spectrum has a wide band availability, it is in reality a scarce resource, especially in lower frequencies.

This scarcity is aggravated by the coexistence of multiple services, technologies and the emergence of use cases that demand higher data-rates, lower latency, better indoor coverage, and better spectral efficiency (Cavalcanti, Maciel, Walter & Silva 2018). Modern urban environments are characterized by a dense and diverse set of services operating within the radio spectrum, particularly in highly contested frequency bands. The growing demand for wireless connectivity across multiple domains, including cellular networks, public safety communications, Wi-Fi, IoT, broadcasting, satellite

Table 2.1: Frequency bands with their corresponding subdivisions.

Band number	Symbols	Frequency range	Corresponding metric subdivision
3	ULF	300–3 000 Hz	Hectokilometric waves
4	VLF	3–30 kHz	Myriametric waves
5	LF	30–300 kHz	Kilometric waves
6	MF	300–3 000 kHz	Hectometric waves
7	HF	3–30 MHz	Decametric waves
8	VHF	30–300 MHz	Metric waves
9	UHF	300–3 000 MHz	Decimetric waves
10	SHF	3–30 GHz	Centimetric waves
11	EHF	30–300 GHz	Millimetric waves
12	-	300–3 000 GHz	Decimillimetric waves
13	-	3–30 THz	Centimillimetric waves
14	-	30–300 THz	Micrometric waves
15	-	300–3 000 THz	Decimicrometric waves

communications, and transportation systems, has intensified spectrum congestion and interference challenges.

Cellular networks, from 2G to 5G, primarily operate in the UHF bands, such as 700 MHz, 850 MHz, 1900 MHz, and 2100 MHz. Low-band spectrum (such as sub-1 GHz) is at the core of these deployments, given its ability to provide wider coverage and require less infrastructure compared to higher frequency spectrum (CTIA 2022). With users and devices consuming more data, spectrum scarcity has driven operators to expand into higher frequencies. For example, the lower mid-band (3 GHz – 8.4 GHz) has been strategical to 5G deployments. It offers reduced coverage but greater throughput per MHz. The high-band (24-50 GHz) is also considered for current and future networks. It has more limited coverage but greater capacity. In the US, this band is currently mainly allocated to government satellites and experimental usages (CTIA 2022). Similarly, public safety operations, such as mission-critical, military, transportation, police, fire rescue and EMS, rely on dedicated bands to ensure reliable and interference-free communications during critical operations. Their scope can span over a large geographical area and some of its data might flow into broadband wireless networks such as Wi-Fi and LTE (Kumbhar et al. 2017). In the US, VHF, 700 MHz, 800 MHz, 900 MHz, and 4.9 GHz are some of the frequencies allocated to Public Safety Networks (PSNs) (Kumbhar et al. 2017).

Wi-Fi networks, widely deployed in 2.4 GHz, 5 GHz, and now 6 GHz (Wi-Fi 6E), provide wireless internet access in homes, businesses, and public spaces. However, it can significantly contribute to spectrum congestion in urban environments. The coexistence of Wi-Fi with LTE-U, NR-U (4G and 5G Unlicensed), and IoT networks in shared spectrum bands also introduces additional interference challenges (Hasan & Khan 2022). The extensive use of old and new wireless networks will lead to numerous heterogeneous networks sharing the same spectrum (Hasan & Khan 2022). IoT applications, including smart city infrastructure, industrial automation, and connected

home devices, has also increased demand for unlicensed spectrum, particularly in bands such as 868 MHz (LoRa), 915 MHz (NB-IoT), and 2.4 GHz (Zigbee, Bluetooth, Wi-Fi) (Hasan & Khan 2022).

Broadcast television continues to occupy portions of the VHF and UHF bands, though the transition to digital broadcasting and spectrum re-purposing is changing the spectrum landscape. For instance, the World Radiocommunication Conference (WRC) 12 was marked by a strong pressure from the telecommunication sector in favor of more radio spectrum, leading to the shared allocation of the band 694-790 MHz to both digital terrestrial television and mobile systems (Więcek et al. 2024). During WRC15, the allocation of additional parts of the UHF band was discussed for IMT systems, such as the 600 MHz band. Further pressure during following WRC editions led to the approval of IMT services in the 470–694 MHz range (Więcek et al. 2024). Additionally, NTNs, which includes UAVs, high altitude platforms, and satellite networks, have recently gained attention due to recent developments of aerial/space technology and reduced cost of their manufacturing and launching. They operate in various bands, such as Ka-band (26.5 to 40 GHz), Ku-band (12–18 GHz), mmWaves (24.25-52.60 GHz), C-band (5.25-5.57 GHz), among others (Sun et al. 2022), (Więcek et al. 2024). NTNs have been studied to coexist and integrate with terrestrial networks, necessitating interference mitigation techniques to minimize disruptions to both services (Rinaldi et al. 2020).

Urban transportation systems and aviation communications rely on dedicated spectrum for air traffic control (VHF: 118–137 MHz), radar (L-band: 960–1215 MHz), and V2X (Vehicle-to-Everything) communications (5.9 GHz). However, coexistence is sometimes considered, such as the deployment of 5G networks in the 3.1-3.55 GHz band, along with existing airborne radars (Fahim Raouf et al. 2023), which has raised debates and concerns on possible interference.

The constant growth of data consumption and the expansion of wireless technologies into this vast range of applications have driven industry and academia to rethink how the radio spectrum is utilized and shared among users, devices, services, and technologies. Wireless networks now support far more than mobile broadband, as illustrated with the previous discussions, even without considering coexistence of different technologies. This increasing demand has made the radio spectrum an extremely valuable and limited resource, pushing operators to adopt increasingly aggressive frequency reuse strategies, which in turn aggravates co-channel interference. As multiple services compete for limited frequency bands, effective solutions must focus on how frequency resources are allocated and how users are served within the system to minimize interference while maximizing spectral efficiency.

2.4.5 Hotspots

ICI can be significantly aggravated by the presence of hotspots, particularly in dense urban environments where their occurrence is more frequent. Hotspots can be defined as regions with a high density of active user equipments (UEs). These regions are inherently unpredictable and can shift dynamically based on user behavior. They can appear due to events such as mass gatherings, sports matches, public protests, or traffic congestion.

These ephemeral hotspots can suddenly increase the data demand, severely straining network capacity and intensifying inter-cell interference.

The severity of ICI in a hotspot can be further influenced by its distribution relative to serving cells. When a hotspot is located near a cell boundary, users within the same hotspot are often connected to different base stations. These users are more likely to suffer from ICI given multiple interfering signals from neighboring BSs have high power, as users are clustered tightly within a limited space. As BSs are competing for resources given the sudden congestion, depending on the frequency reuse pattern, it can lead to lack of available resources and higher probability of users being allocated to the same frequency.

Although these challenges posed by hotspots are well-known, their dynamic nature presents an additional layer of complexity. Static network planning assumes relatively predictable user distributions, but scenarios with hotspots can experience sudden changes in user deployment and data consumption trends. For example, a neighborhood with a historically uniform user distribution might suddenly host a few hotspots due to a car accident happening as users make their way to a public manifestation. In such situations, mobile networks that were planned and deployed statically will not be able to serve its users well.

Hotspots can also change user consumption behavior. The aforementioned events usually motivate users to consume more data, engaging in bandwidth-intensive applications such as video streaming, sharing and consuming content on social media, or delay-sensitive applications such as cloud gaming and video calls. The increased demand for bandwidth in such scenarios, especially with users consuming and sharing large volumes of content, results in reduced SINR and increased network congestion, potentially leading to increased Bit Error Rate (BER), and consequently, lower traffic capacity.

Such scenarios highlight the limitations of conventional network planning. Traditional interference management techniques, while effective to some extent, may fail to adapt to rapidly changing hotspot scenarios.

2.5 Inter-cell Interference Coordination

The standardization of LTE by 3GPP marked a significant shift in mobile network architecture, introducing advanced interference coordination techniques to enhance spectral efficiency and support the ever-growing demand for data. ICI was, from the start, a key challenge in LTE deployment, particularly in dense cellular networks where neighboring cells compete for spectral resources. To mitigate this issue, 3GPP has continuously evolved its ICIC strategies across different releases.

3GPP first introduced ICIC techniques in Release 8, providing implicit support for interference coordination (3rd Generation Partnership Project (3GPP) 2021). LTE enabled finer control over resource allocation across different domains, allowing network operators to implement interference management strategies more effectively. Its OFDMA architecture facilitated interference control by allowing fine-grained scheduling of resources in time, frequency, and power domains. However, without additional

coordination mechanisms, the interference at the edges of adjacent cells remained a significant concern.

One of the key techniques introduced under ICIC was Fractional Frequency Reuse, which aimed to partially orthogonalize users in the frequency domain, mitigating interference at the cell edges (Zhou et al. 2014). This technique allowed LTE networks to balance spectral efficiency and interference management by partitioning frequency bands based on user performance. Additionally, Release 8 introduced Home eNodeBs (HeNBs), a precursor to the modern femtocell concept, to enhance coverage and capacity in indoor environments.

As heterogeneous networks became a key deployment strategy, interference challenges became more complex. Small cells, such as pico and femtocells, were introduced within macro-cell coverage areas to improve capacity and coverage, but their deployment within shared frequency bands led to severe downlink interference at the cell edges. This challenge led to the introduction of Enhanced ICIC (eICIC) in Release 10, which extended interference coordination to the time domain.

The core mechanism of eICIC is the Almost Blank Subframe (ABS) technique. ABS are subframes during which macro cells significantly reduce or suppress data transmissions, leaving only essential control and reference signals at lower power levels. This allows small cells, such as picocells, to schedule transmissions for their users without suffering strong interference from the macro layer. By carefully configuring ABS patterns, operators can ensure that victim users at the macro cell edges can reliably connect to small cells, enhancing overall network performance.

Building on eICIC, Release 11 introduced Further Enhanced ICIC (feICIC) to address residual interference. Despite the use of ABS, some control and signaling transmissions still contributed to interference. feICIC introduced advanced interference cancellation techniques at the receiver side, refinements to the ABS mechanism, and more flexible scheduling strategies.

Thus, eICIC and feICIC are often considered an evolution of the original ICIC framework. However, despite the emergence of these advanced techniques, ICIC based on fractional frequency reuse has remained relevant, as extensively demonstrated by the related works presented on Chapter 1. Frequency allocation is still a key aspect in 5G and beyond systems. While LTE operates under fixed bandwidth allocations and predefined spectrum usage policies, modern systems are designed for flexible and dynamic spectrum utilization, creating new opportunities to revisit FFR-based ICIC as a foundation for dynamic spectrum management, allowing networks to efficiently adapt frequency reuse patterns based on real-time interference conditions.

2.5.1 Fractional Frequency Reuse

FFR is an interference coordination strategy that divides the total system bandwidth into multiple sub-bands, with different reuse factors and transmission power levels. The primary objective of FFR is to find a good compromise between spectral efficiency and co-channel interference mitigation, adjusting how the frequency is reused in neighboring base stations.

As discussed earlier in Section 2.3, modern mobile networks often operate with a frequency reuse factor of 1, meaning that all neighboring cells share the same frequency band. While this approach maximizes spectral efficiency, it also leads to severe interference at the cell edges, where users experience high signal overlap from multiple base stations. Conversely, using a reuse factor of 3, where neighboring cells are assigned entirely separate frequency bands, eliminates co-channel interference at the cost of reduced spectral efficiency. In this configuration, each cell only has access to one-third of the total available spectrum, which can significantly limit overall network capacity, as seen from the discussions in Section 2.3.

This trade-off is illustrated in Figure 2.4. It shows a cellular system with a reuse-3 model, considering omnidirectional antennas and hexagonal cells. In each cluster of 3 cells, all cells are disjoint, hence, neighboring cells do not share bandwidth. This model ensures interference-reduced operation at the cost of lower bandwidth availability per cell, since every cell has only $1/3$ of the available bandwidth.

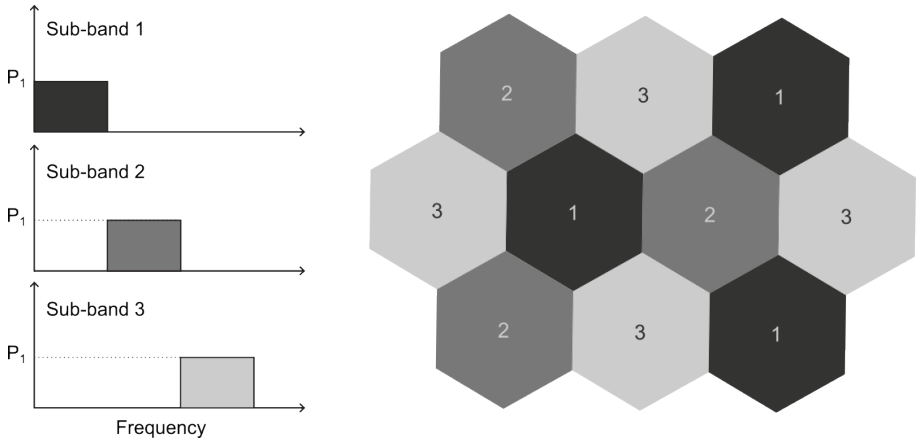


Figure 2.4: Cellular system with reuse 3.

To address this inefficiency, FFR-based ICIC techniques provide an intermediate solution between reuse-1 and reuse-3, ensuring that spectral resources are efficiently allocated while improving the SINR at the cell edges. The key motivation behind FFR is that interference is not uniformly distributed within a cell: while users near the cell center experience relatively strong signal reception, users at the cell edge suffer more interference from adjacent base stations. Hence, FFR allocates frequency resources based on user performance. Users are classified in different groups based on their interference levels, and then assigned to specific sub-bands that may be shared or not with neighboring cells. Users experiencing higher interference are allocated to sub-bands with lower reuse, reducing their exposure to interference, while users facing less interference are assigned to sub-bands with higher reuse, maximizing spectral efficiency.

Two common FFR implementations are the Strict Frequency Reuse (Strict FR) and the Soft Frequency Reuse (Soft FR), as depicted in Figures 2.5 and 2.6. In the Strict FR, each cell is divided into two distinct frequency zones: a common sub-band allocated to the cell center, which operates with a reuse factor of 1, and a private sub-band

allocated to the cell edge, which operates with a reuse factor of 3. This ensures that the center regions of neighboring cells share the same frequency resources, maximizing spectral efficiency, while the edge regions use orthogonal sub-bands across adjacent cells, thereby eliminating interference at the edges. Additionally, power control is employed to differentiate transmission power levels across these sub-bands. Typically, a higher transmission power is assigned to the cell edge users, compensating for path loss and enhancing their received signal quality.

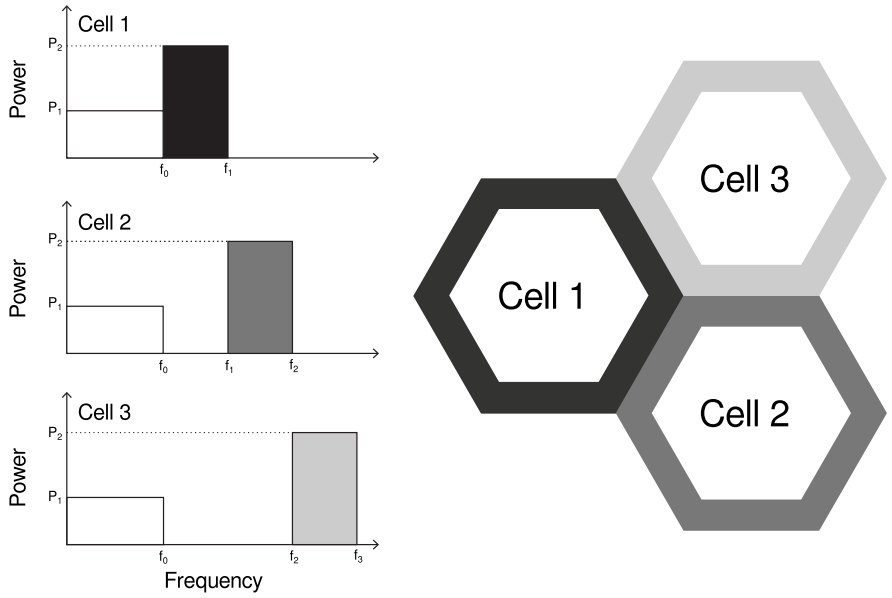


Figure 2.5: Strict Frequency Reuse.

The Soft FR also partitions the available bandwidth into two sub-bands, but with a more flexible frequency-sharing scheme. Unlike Strict FR, where the sub-band allocated to cell-edge users is completely orthogonal across neighboring cells, SFR allows partial frequency reuse at the cell edges, as illustrated in Figure 2.6. The cell-edge sub-band is shared with center users from adjacent cells. Hence, cell-edge users from one cell do not experience direct co-channel interference from cell-edge users in neighboring cells, but they can suffer interference from cell-center users in adjacent cells. Similarly, center users in one cell may experience higher interference from edge users in adjacent cells, especially in dense deployments.

Soft FR offers improved spectral efficiency compared to Strict FR, as it enables all cells to utilize the entire available bandwidth, rather than restricting certain sub-bands exclusively to specific regions. However, this improvement comes at the cost of increased interference for both cell-center and cell-edge users.

For both Strict FR and Soft FR, three key parameters determine their configuration: user classification criteria, bandwidth allocation, and power distribution across sub-bands. The user classification determines whether a user is assigned to the cell-center or cell-edge sub-band, which is typically based on an interference-related metric such as SINR. Given

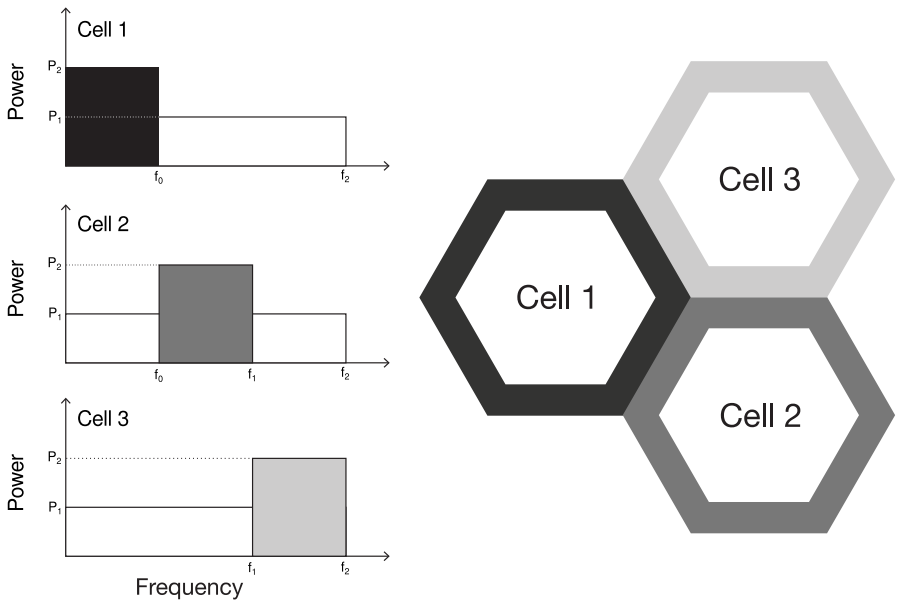


Figure 2.6: Soft Frequency Reuse.

that both Strict FR and Soft FR have two sub-bands, a single threshold is sufficient to classify users into either cell-center users (CCUs) or cell-edge users (CEUs). RSRQ is commonly employed for this classification due to its ability to reflect the overall signal quality, incorporating interference, noise levels, and resource allocation conditions, as illustrated in Equation (2.6).

$$RSRQ = \frac{N \times RSRP}{RSSI}, \tag{2.6}$$

where RSRP (Reference Signal Received Power) represents the average power received from the serving cell's reference signals, RSSI (Received Signal Strength Indicator) accounts for total received power including interference and noise, and N corresponds to the number of Physical Resource Blocks (PRBs) allocated to the user. If a user's RSRQ value is higher than the defined threshold, it is categorized as a cell-center user (CCU) and assigned to the common sub-band, which operates with a reuse factor of 1. If the RSRQ is lower than the threshold, the user is classified as a cell-edge user (CEU) and allocated to the private sub-band, which has a reuse factor of 3 for both schemes.

The second parameter is the bandwidth allocation, which determines how much spectrum is assigned to each sub-band. In LTE and 5G networks, OFDMA is the chosen multiple access technique for the downlink. OFDMA divides the available spectrum into multiple narrowband subcarriers, ensuring orthogonality between users and minimizing intra-cell interference. At the physical layer, LTE organizes radio resources into a structured time-frequency grid, as illustrated in Figure 2.7. The smallest unit of allocation is the Resource Element (RE), which consists of a single subcarrier in the frequency domain and a single OFDM symbol in the time domain. Each LTE subframe has a

duration of 1 ms and consists of two time slots, each lasting 0.5 ms. Within each time slot, multiple OFDM symbols are transmitted, typically 7 symbols per slot in the case of normal cyclic prefix (CP) and 6 symbols per slot for extended CP.

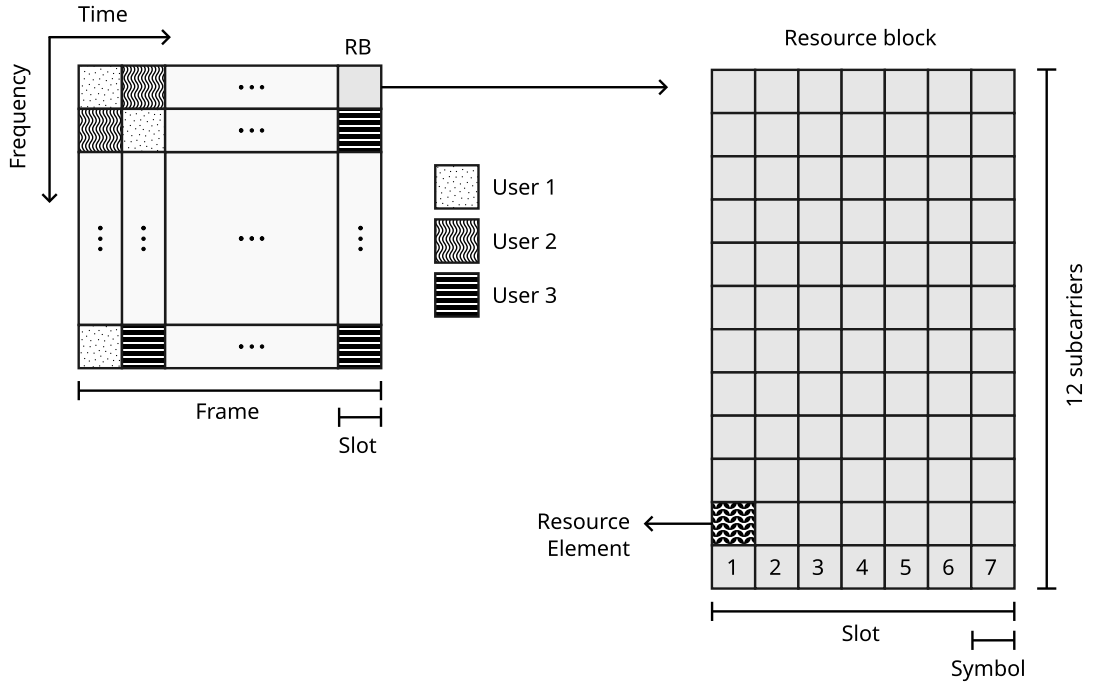


Figure 2.7: Basic OFDMA frame structure in LTE systems, adapted from (Hamza et al. 2013).

To facilitate scheduling and resource allocation, LTE groups 12 consecutive subcarriers (180 kHz) in the frequency domain over a slot duration to form a PRB, the smallest unit of resources that can be allocated to a user. These PRBs are further grouped into Resource Block Groups (RBGs). The number of PRBs per RBG depends on the total system bandwidth, as shown in Table 2.2. For instance, in a system with a 100 PRB bandwidth allocation, each RBG consists of 4 PRBs, meaning that the minimum sub-band size possible is 4 PRBs.

The third parameter is power allocation across sub-bands. In both Strict FR and Soft FR, the private sub-band (assigned to CEUs) typically operates at a higher transmission power level than the common sub-band. This ensures that users at the cell edge, who are more susceptible to interference and path loss, receive stronger signal coverage to enhance their SINR. In Soft FR configurations, increasing power for CEUs can improve their performance but may also raise interference for adjacent cells.

Table 2.2: LTE Resource Block Group (RBG) Allocation.

LTE System Bandwidth (MHz)	**Number of PRBs**	**RBG Size (PRBs per Group)**
1.4	6	1
3	15	2
5	25	2
10	50	3
15	75	4
20	100	4

Chapter 3

Problem characterization and methodology

In modern cellular networks, ICI continues to limit spectral efficiency and degrade user performance, as discussed in Section 2.2. As mobile data demand continues to rise, driven by bandwidth-intensive applications and increased number of connected devices, effective interference coordination is critical in modern systems. In particular, hotspot scenarios, where zones with high user density appear unexpectedly, aggravate co-channel interference in OFDMA systems, as shown by recent research in Section 1.1.2 and discussions in Section 2.4.5. This requires adaptive and intelligent resource allocation strategies to maintain suitable performance and maximize network capacity.

FFR has proven to be an efficient solution to mitigate interference by partitioning frequency resources between cell-center and cell-edge users, as showed in Section 1.1.3 and further discussed in Section 2.5.1. However, static FFR configurations do not adapt well to fluctuating interference patterns, potentially leading to lower SINR, lower throughput, inefficient spectrum utilization, and poor overall performance especially for cell-edge users.

Recent research has proposed adaptive interference coordination mechanisms, using ML techniques such as reinforcement learning (RL) to enhance network adaptability. While some studies have used dynamic FFR solutions, they typically control only a single parameter, as discussed in Section 1.1.3. The unexpected emergence of hotspots, localized zones of high user density, can drastically alter user distribution and create sudden surges in demand, making static or semi-adaptive FFR configurations inadequate. Despite the potential of FFR-based approaches, few studies have specifically addressed hotspot scenarios without deploying additional base stations.

Thus, the research problem addressed in this thesis is the added co-channel downlink interference caused by the unexpected appearance of hotspots in a single-tier OFDMA-based system. This chapter details the approach, tools, and techniques used to address this research problem. The methodology was designed to ensure a systematic and reproducible analysis of the investigations. The following sections describe the modeling of the system, including the network environment, traffic conditions, interference scenarios and FFR schemes. For the more complex parts of the implementation, additional details and code snippets are provided to contribute with future works aiming to reproduce or build upon the proposed scenarios and algorithms. The collection and

computation of key evaluation metrics is also detailed to ensure the validity and reliability of the results.

The chapter is concluded with an analysis of how the hotspots impact the system performance, considering multiple deployments where the hotspots are located near or far from the base stations. This study emphasizes the core problem and examines how user classification and bandwidth allocation influence the performance of the FFR technique.

3.1 Simulation tools

To evaluate the hypotheses derived from the research questions, this work used network simulation as the primary method for validating the proposed solutions in an OFDMA system. Real-world testing of mobile networks, especially in multi-user scenarios, is often impractical due to the high cost and logistical challenges of deploying multiple UEs and base stations under controlled conditions. Additionally, large-scale experimentation requires specialized infrastructure and permission to utilize the desired frequency band, making simulation a more viable alternative for early-stage research and algorithm development.

The ns-3 network simulator (ns3 2025*c*) was chosen as the main simulation tool. Ns-3 is an open-source, discrete-event network simulator widely used for research and educational purposes. Primarily implemented in C++, its modular architecture provides a range of libraries and models that allow for flexible experimentation with wireless networks.

One of the key advantages of ns-3 is its compliance with technical specifications from standardization bodies such as 3GPP and IEEE. This ensures that the simulation models used in this work closely follow real-world network behaviors, improving the validity of results. Moreover, ns-3 is well-documented and supported by an active research community. All models that are officially supported follow comprehensive documentation and citation of relevant scientific studies and/or technical specifications.

Although simpler than real-world experimentation, network-level simulations are still computationally demanding, especially when modeling large-scale cellular networks with multiple users and base stations. To address this challenge, the simulations in this work were executed using the High Performance Computing Center at UFRN (NPAD). According to NPAD, the management, administration, and maintenance of computer resources should not be part of a researcher's routine (Instituto Metrópole Digital / UFRN 2025*b*). NPAD provides high-performance hardware that is always available and well managed, allowing researchers to focus on their scientific investigations rather than the technical details of infrastructure maintenance.

NPAD offers substantial computational power, featuring a mix of Intel and AMD processors, as well as GPU-equipped nodes for specialized workloads. For example, in this work, simulations were primarily executed on a partition consisting of 56 computing nodes, each equipped with two AMD CPUs running at 2.0 GHz and 128 cores, and 512 GB of RAM. Although the LTE model in ns-3 does not support parallel execution across multiple cores, it is possible to distribute Monte Carlo simulation repetitions or independent simulation runs with varying parameter values across different cores. More

information on NPAD's hardware resources is available at (Instituto Metrópole Digital / UFRN 2025*a*).

To efficiently distribute and manage computational workloads, NPAD's cluster employs the Simple Linux Utility for Resource Management (SLURM), a widely used open-source job scheduler designed for high-performance computing (HPC) environments. SLURM allocates computational resources to user-submitted jobs, and supports job prioritization, parallel execution, and resource reservation, making it well suited for large-scale simulations.

When submitting a simulation job, users specify resource requirements such as the number of CPU cores, memory allocation, and expected runtime. SLURM then places the job in a scheduling queue, assigning it to an appropriate compute node when resources become available. This automated scheduling mechanism maximizes efficiency, preventing resource contention and enabling multiple users to share the cluster without performance degradation.

3.1.1 Creation of simulation campaigns

A Python script was developed to automate the creation of the necessary files to run a Monte Carlo campaign simulation. This section details its functionality and implementation, as a contribution for future works that may want to reuse or adapt it.

The main script responsible for the creation of the campaign files is `createCampaign.py`. It reads the configuration defined in a YAML file, `campaign.yaml`, which contains two major sections: campaign and scenario parameters. The campaign parameters define settings related to the execution of the campaign, such as number of jobs, time-limit on SLURM, and the ns-3 path. The scenario section specifies simulation parameters such as simulation time, number of users, FFR algorithm type, system bandwidth, and RL settings. The identifiers of the scenario parameters have to match the command-line arguments of the ns-3 script. If a parameter `simTime` is defined as `60`, the same parameter must exist in the ns-3 script. Listing 3.1 demonstrates a typical campaign file, as described above. In this example, the campaign named *campTest* is configured to run 50 jobs, with a 60 seconds long simulation, deploying 30 uniform users and 10 users within each hotspot.

```
campaign:
    # Output filename (and part of output dir)
    campaignName: campTest
    # Number of Jobs
    jobs: 50
    # Complete path of ns-3 simulator root directory
    ns3_path: /home/user/ns3
    # Script with the main code (inside scratch folder without .cc)
    ns3_script: hotspot-deployments
    # Run the simulation on cluster? 0 - False, 1 - True
    runOnCluster: 1
    # Time limit to simulate using the cluster
    nDays: 0
    nHours: 23
    nTasks: 1
```

```
    importRandomSeeds: false

scenario:
    # Duration in seconds
    simTime:
        - 60
    hs_interval:
        - 10
    nUniformUes:
        - 30
    npicoUes:
        - 10
    frAlgorithmType:
        - LteFrStrictAlgorithm
    rlActivated:
        - false
        - true
    pathLoss:
        - Cost231PropagationLossModel
    bandwidth:
        - 100
    rlEpsilon:
        - 0.4
    rlActionMode:
        - RsrqThreshold
        - Bandwidth
```

Listing 3.1: Example of campaign YAML configuration file.

Additionally, if a scenario parameter contains multiple values, such as `rlActivate` in Listing 3.1, the script generates multiple simulation calls. It computes the Cartesian product of all scenario parameters with more than one value and each unique combination represents a distinct simulation configuration. Once the cartesian product of the scenario parameters is created, the script formats each execution command for ns-3, using the `./ns3 -run -parameter=paramValue` format. All runs are then replicated according to the number of jobs. The generated `./ns3 -run` commands are then grouped in a bash script and the number of runs per script is an input parameter in `createCampaign.py`.

The campaign is organized as a hierarchical folder structure where the scripts and simulation results will be stored. Figure 3.1 illustrates the resulting folder structure. The script generates each job with its own random seed, and it accepts a list of random seeds as input for better reproducibility. The list of names for the result files are stored in a CSV and `RunAllScripts.sh` is the only script to be executed to submit the campaign to NPAD's cluster.

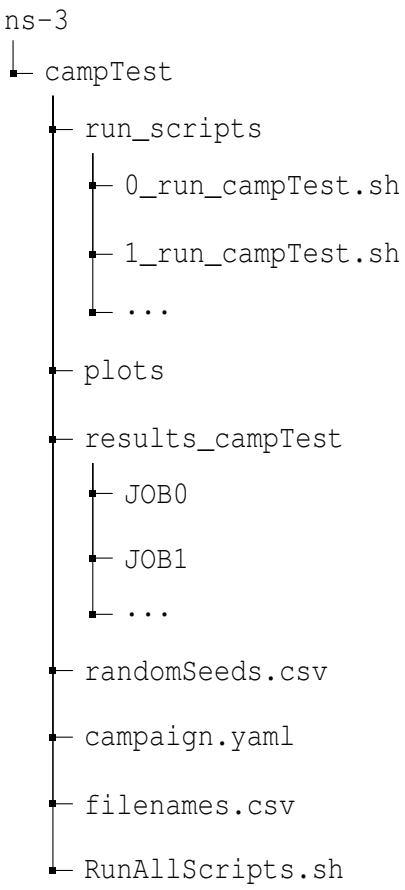


Figure 3.1: Example of campaign folder structure.

3.2 ns-3 LTE model

Figure 3.2 provides an overview of how the LTE module is implemented in ns-3. The model is divided into two main components: LTE and the Evolved Packet Core (EPC). The LTE model represents the radio access network (E-UTRAN) protocol stack and includes the RRC, PDCP, RLC, MAC, and PHY layers, all of which are implemented within the User Equipment (UE) and the eNodeB (eNB).

In summary, the Packet Data Convergence Protocol (PDCP) layer performs header compression and decompression of the IP packets, ciphers and deciphers the user plane and most of the control plane data, and handles integrity protection and verification. The Radio Link Control (RLC) layer transfers the received PDUs, with different operation modes that may involve additional steps such as error correction (Sesia et al. 2011). The Medium Access Control (MAC) layer receives the data from the RLC layer and maps the logical channels to transport channels, handling resource management through scheduling and priority. It performs multiplexing/de-multiplexing of RLC PDUs, and handles error correction through Hybrid Automatic Repeat reQuest (HARQ) (Sesia et al. 2011). Finally, the physical layer handles transmission over the air interface, which includes modulation,

coding, and interference modeling.

In the LTE model, at the radio level, the RB is the fundamental unit used for resource allocation, since packet-scheduling is done on a per-RB basis. This allows the occurrence of interference between base stations on a specific subset of RBs. According to ns-3's documentation, this could not be possible with system-level simulation, given that resource allocation happens only at call/bearer establishment (ns3 2025*a*). ns-3 is able to simulate up to tens of eNBs and hundreds of users. This would be unfeasible with link-level simulation, that models up to the symbol level, which would need the implementation of all PHY layer signal processing, limiting simulation due to computational complexity. Link-level simulators are often limited to one base station and one or a few users (ns3 2025*a*).

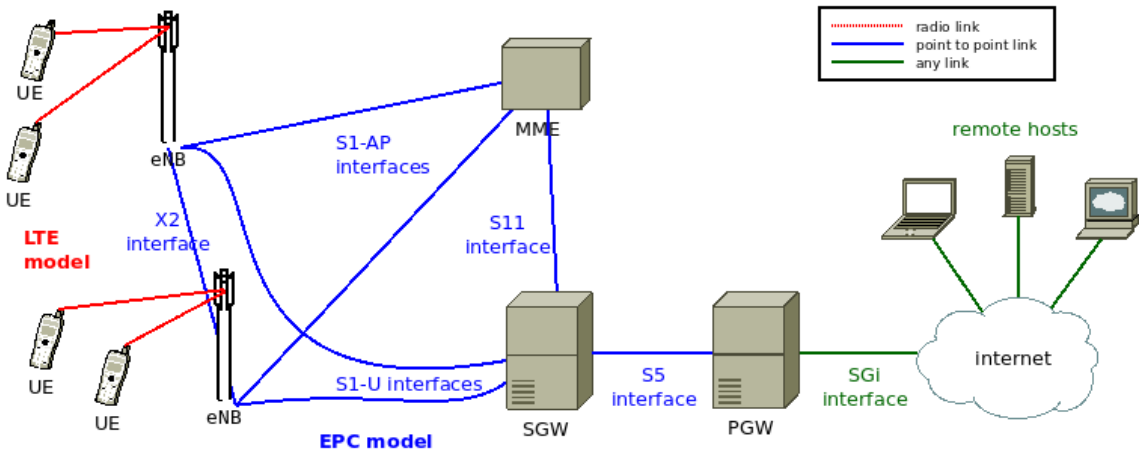


Figure 3.2: Overview of the LTE-EPC simulation model. Source: (ns3 2025*a*).

The EPC model, on the other hand, allows the simulation of end-to-end IP connectivity over the LTE model (ns3 2025*a*). It implements the Mobility Management Entity (MME), Serving Gateway (S-GW), and Packet Data Network Gateway (P-GW). ns-3 allows simulations to be performed with just the E-UTRAN, however, the EPC is needed for UE-to-internet communication using the standard IP protocol.

On LTE systems, the MME is the main control element within the EPC. It operates only in the control plane and its logical connection to the UE is the main control channel between the UE and the network (Sesia et al. 2011). It handles user authentication, subscription and mobility management. The S-GW handles user plane tunnel management and switching, also serving as the local anchor during mobility, and the P-GW is the edge router between the LTE network and external packet data networks (Sesia et al. 2011).

Figure 3.3 illustrates the data plane protocol stack implementation in ns-3 when the EPC model is used. ns-3 provides a detailed implementation of LTE and EPC protocols, including all protocol stacks (S5 protocol stack, S1-U protocol stack and the LTE radio protocol stack) specified by 3GPP (ns3 2025*a*), enabling the study of radio resource management and inter-cell interference coordination.

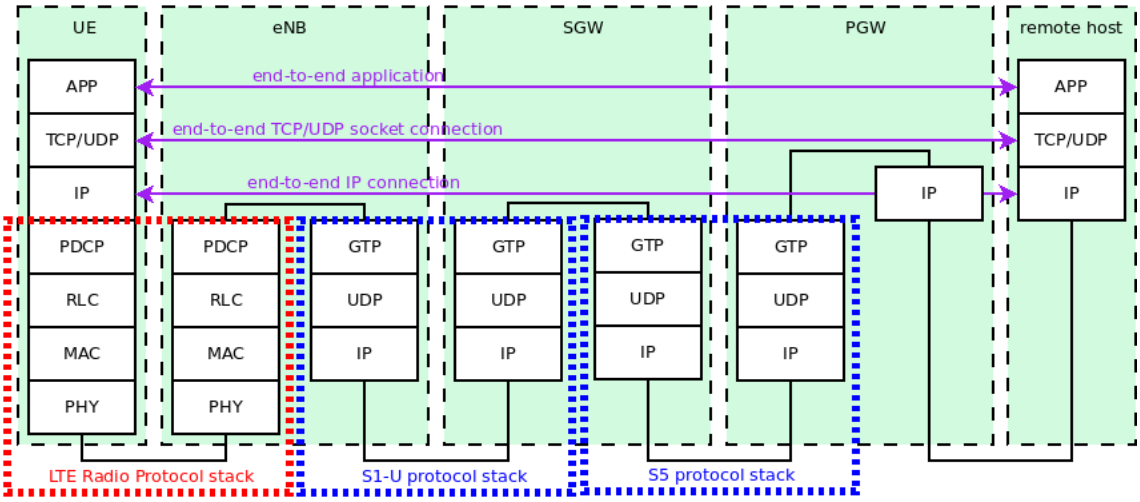


Figure 3.3: LTE-EPC data plane protocol stack. Source: (ns3 2025*a*).

3.2.1 Modeling and collection of metrics

As mentioned above, ns-3 uses 3GPP specifications to guide its implementations. Since simulation is the main tool to validate the proposed solutions in this work, it is important to verify how some relevant metrics are implemented and measured by the LTE model.

3GPP specifies that a user has to report the RSRP and the RSRQ of each eNB it perceives. RSRP measures the received power of an eNB and the RSRQ adds interference and noise to the calculation, as discussed in Section 2.5.1. RSRP is reported by PHY layer in dBm and RSRQ in dB. These values are reported every 200 ms according to the specification, but ns-3 allows configuration of this reporting interval. Section 2.5.1 also introduces how RSRQ is calculated. The equations below details its calculation and how the LTE model implements it. Given that the RB is the smallest resource unit, the calculation of RSRP and RSRQ are simplified by assuming the channel is flat within the RB, that is, all REs have the same power in an RB. The RSRP calculation can be reduced to Equation (3.1).

$$RSRP = \frac{\sum_{k=0}^{K-1} P(k)}{K}, \text{with} \tag{3.1}$$

$$P(k) = PSD_{RB}(k) \times \frac{180000}{12}, \tag{3.2}$$

where $P(k)$ is the signal power of each RE in RB k. $PSD_{RB}(k)$ is the power spectral density of the RB k, 180000 is the RB's bandwidth in Hz and 12 is the number of REs per RB in an OFDMA symbol.

The RSSI (Received Signal Strength Indicator), the total received wideband power at a given frequency (Sesia et al. 2011), is simplified in a similar manner. The RSSI includes

signal power, interference and noise, but it is only used in calculating the RSRQ value inside the UE (Sesia et al. 2011). Its general definition is presented in Equation (3.3).

$$RSSI = \sum_{k=0}^{K-1} \sum_{s=0}^{S-1} \sum_{r=0}^{R-1} \frac{I(k,s,r) + P(k,s,r) + N(k,s,r)}{S}, \tag{3.3}$$

where S is the number of OFDM symbols with reference signals in an RB and R is the number of resource elements with a reference signal in an OFDM symbol, which is fixed as 2. $I(k,s,r)$, $P(k,s,r)$, and $N(k,s,r)$ are the interference power, power of the serving cell, and noise power of the resource element r in symbol s, respectively. The RSSI can also be simplified to equation (3.4) due to all REs having equal power.

$$RSSI = RSSI = \sum_{k=0}^{K-1} 2 \times (I(k) + P(k) + N(k)). \tag{3.4}$$

To collect the necessary metrics for analyzing the results and for real-time usage in the proposed solutions, ns-3 implements a robust tracing subsystem that relies on connecting trace sources and trace sinks. Trace sources can signal any events happening in the simulation to provide access to desired underlying data. These trace sources are connected to any trace sink, which consumes the data that was generated by the trace source (ns-3 2025). This enables easy data collection by creating trace sinks for trace sources implemented by the authors of the ns-3 models. Additionally, multiple trace sinks can be connected to a same trace source, which can be triggered by time intervals or by events such as the arrival of a packet.

The solution proposed in this work collects real-time data from the system in order to assess the state of the network and calculate a reward value after the reinforcement learning agent takes an action, changing a network parameter. For throughput-related metrics, we use a trace source from the RLC layer, which notifies the reception of a Protocol Data Unit (PDU). The total amount of received bytes is summed and divided by the sampling interval to calculate throughput. The SINR is collected from a trace source provided by the PHY layer. It reports user RSRP and SINR with a configurable sampling period. The SINR in the downlink is computed per RB and it is calculated as shown in Equation (3.5). It is the division of the power of the intended signal by the sum of all interfering signals on the same RB and the noise power.

$$\gamma = \frac{P_{\text{signal}}}{P_{\text{noise}} + \sum P_{\text{interference}}}. \tag{3.5}$$

The ns-3 LTE model also has native support to exporting Key Performance Indicators (KPIs) to files, capturing data from the PHY, MAC, RLC and PDCP layers. These metrics are used for performance analysis in this work. Since the focus is on throughput and SINR evaluation, the RLC and PHY outputs are used. Both files include time stamps for each measurement, along with identification details such as cell or UE ID. The RLC KPI file has information on transmitted and received PDUs, while one of the PHY KPI files has RSRP and a linear average of the downlink SINR over all RBs, in linear units (ns-3 2025).

3.3 Implementation of FFR algorithms in ns-3

ns-3 has currently seven FFR algorithms implemented, as described in (Hamza et al. 2013). These algorithms operate at the scheduling level, under the control of the MAC layer (ns3 2025*a*). When the MAC layer initiates the scheduling process, the scheduler consults the Frequency Reuse FR entity for available RBGs. Depending on the bandwidth allocation strategy of the selected FFR algorithm, the FR returns all RBGs available in the corresponding cell. Next, the scheduler consults if the desired RBG is allowed for the UE and, if the response is true, the scheduler assigns the RBG to the UE. If not, it repeats the process for another RBG. This process is detailed in the sequence diagram shown in Figure 3.4, taken from the ns-3 LTE model documentation (ns3 2025*a*).

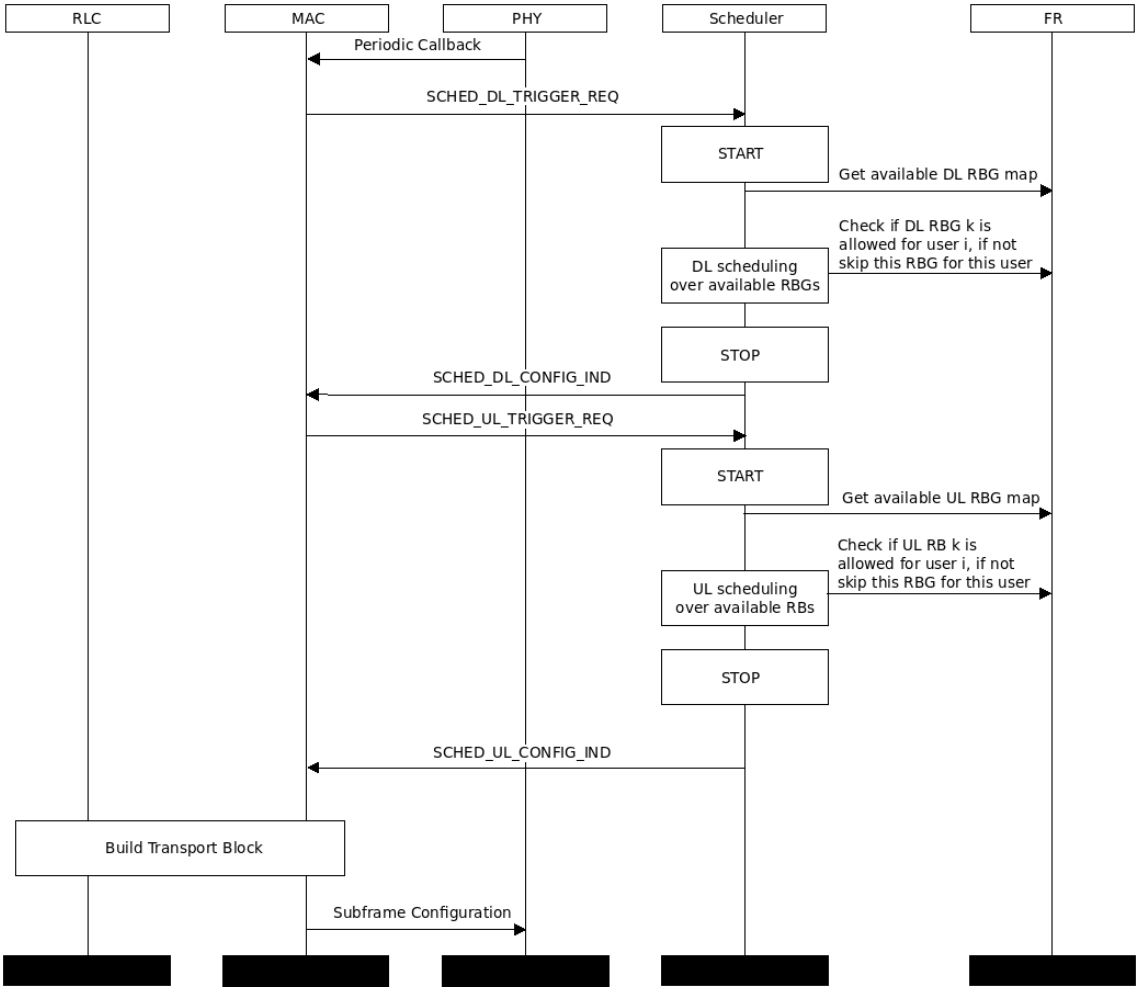


Figure 3.4: Sequence diagram of scheduling with FR algorithm. Source: (ns3 2025*a*).

ns-3 implements these algorithms in a way that users can easily configure its parameters. This configuration is managed through its attributes system, which organizes access to internal member objects within a simulation (ns-3 2025). The attribute system provides a structured approach for users to modify parameters dynamically, whether by

directly setting values in code, using command-line arguments, or via configuration paths. This flexible implementation enables fine-grained control over internal variables without modifying the core models. In this work, the attribute system is used to configure the key FFR parameters, which are described in detail in the following sections.

3.3.1 Sub-band power allocation

The power allocation for FFR allows different power levels for each sub-band. This is achieved by configuring the transmission power of the PDSCH relative to the Reference Signal. If the selected algorithm supports different power levels for different sub-bands, separate power offset variables are defined for each sub-band. For example, in the Strict FR, the attributes `CenterPowerOffset` and `EdgePowerOffset` specify the power adjustments for the center and edge sub-bands, respectively.

This work does not study power allocation of the sub-bands, hence, both attributes are set before the simulation starts using the `LteHelper`, similar to other simulation parameters, such as the scheduler type, channel model, or the choice of FFR algorithm. `LteHelper` is a class in ns-3 that simplifies the creation, configuration, and management of LTE objects. It provides an abstract layer for setting up LTE network components, such as eNodeBs, UEs, and the EPC. It also provides an easy way to configure attributes.

Listing 3.2 shows the power allocation made through the `LteHelper`. The helper has a dedicated function to set FFR attributes and the user can verify a list of all available attributes on ns3's documentation, grouped by class (ns-3 Project 2016*a*). For example, it lists the attributes of both the *LteHelper*, such as the FFR algorithm to be used, and the Strict FR, such as both attributes shown in Listing 3.2. More details on each attribute can be found within the description of each class. For instance, the attributes in Listing 3.2 are detailed in (ns-3 Project 2016*b*), which is the documentation for the class `ns3::LteFrStrictAlgorithm`. For this work, the power offset is -6 dB for the common sub-band and 3 dB for the private sub-band.

```
1 lteHelper->SetFfrAlgorithmAttribute("CenterPowerOffset",
2                 UintegerValue(LteRrcSap::PdschConfigDedicated::dB_6));
3 lteHelper->SetFfrAlgorithmAttribute ("EdgePowerOffset",
4                 UintegerValue(LteRrcSap::PdschConfigDedicated::dB3));
```

Listing 3.2: Example of attribute configuration for power allocation in Strict FR.

3.3.2 Bandwidth allocation

As discussed in Section 2.5.1, LTE systems allocate bandwidth in terms of RBGs rather than individual RBs, and its size depends on system bandwidth, as defined in Table 2.2. That constraint has to be taken into consideration when allocating RBs to sub-bands. ns-3 provides attributes to define the number of RBs in each sub-band, for DL and UL, separately. Using the Strict FR as an example, `DlCommonSubBandwidth` defines the number of RBs in the common sub-band for DL, and `DlEdgeSubBandwidth` for the edge sub-band. An additional attribute `DlEdgeSubBandOffset` defines the offset between the common and the edge sub-band in each base station, ensuring that the sub-bands are

correctly positioned within a 3-cell cluster. This parameter determines where the edge sub-band starts in the frequency spectrum. Table 3.1 shows the default configurations for the Strict FR for 25 and 100 RBs of system bandwidth, and Listing 3.3 shows how to configure the Strict FR with 28 RBs for the common sub-band and 24 for the edge sub-band. Since the edge starts with an offset of 0 on BS 1, the edge on BS 2 has to start with an offset of 24 RBs, which is the width of the edge sub-band. These attribute values can also be changed in real-time (during *runtime*) using the `Config::Set` command with the attribute's configuration path, which is the first argument of `Config::Set` in Listing 3.3.

Table 3.1: Strict FR Downlink Default Bandwidth Configuration.

Cell ID	Bandwidth	Common Bandwidth	Edge Sub-band Offset	Edge Bandwidth
1	25	6	0	6
2	25	6	6	6
3	25	6	12	6
1	100	28	0	24
2	100	28	24	24
3	100	28	48	24

```
// Base station 1
Config::Set("/NodeList/3/DeviceList/*/$ns3::LteEnbNetDevice/"
            "ComponentCarrierMap/*/LteFfrAlgorithm/"
            "$ns3::LteFrStrictAlgorithm/DlCommonSubBandwidth",
            UintegerValue (28));
Config::Set("/NodeList/3/DeviceList/*/$ns3::LteEnbNetDevice/"
            "ComponentCarrierMap/*/LteFfrAlgorithm/"
            "$ns3::LteFrStrictAlgorithm/DlEdgeSubBandOffset",
            UintegerValue (0));
Config::Set("/NodeList/3/DeviceList/*/$ns3::LteEnbNetDevice/"
            "ComponentCarrierMap/*/LteFfrAlgorithm/"
            "$ns3::LteFrStrictAlgorithm/DlEdgeSubBandwidth",
            UintegerValue (24));

// Base station 2
Config::Set("/NodeList/4/DeviceList/*/$ns3::LteEnbNetDevice/"
            "ComponentCarrierMap/*/LteFfrAlgorithm/"
            "$ns3::LteFrStrictAlgorithm/DlCommonSubBandwidth",
            UintegerValue (28));
Config::Set("/NodeList/4/DeviceList/*/$ns3::LteEnbNetDevice/"
            "ComponentCarrierMap/*/LteFfrAlgorithm/"
            "$ns3::LteFrStrictAlgorithm/DlEdgeSubBandOffset",
            UintegerValue (24));
Config::Set("/NodeList/4/DeviceList/*/$ns3::LteEnbNetDevice/"
            "ComponentCarrierMap/*/LteFfrAlgorithm/"
            "$ns3::LteFrStrictAlgorithm/DlEdgeSubBandwidth",
            UintegerValue (24));

// Base station 3
```

```
Config::Set("/NodeList/5/DeviceList/*/$ns3::LteEnbNetDevice/"
            "ComponentCarrierMap/*/LteFfrAlgorithm/"
            "$ns3::LteFrStrictAlgorithm/DlCommonSubBandwidth",
            UintegerValue (28));
Config::Set("/NodeList/5/DeviceList/*/$ns3::LteEnbNetDevice/"
            "ComponentCarrierMap/*/LteFfrAlgorithm/"
            "$ns3::LteFrStrictAlgorithm/DlEdgeSubBandOffset",
            UintegerValue (48));
Config::Set("/NodeList/5/DeviceList/*/$ns3::LteEnbNetDevice/"
            "ComponentCarrierMap/*/LteFfrAlgorithm/"
            "$ns3::LteFrStrictAlgorithm/DlEdgeSubBandwidth",
            UintegerValue (24));
```

Listing 3.3: Example of attribute configuration for sub-band allocation in Strict FR.

3.3.3 User classification

The FFR algorithms implemented in ns-3 use RSRQ as the criterion for user classification, which was further detailed in Section 2.5.1. In LTE, RSRQ is reported as an integer value, which corresponds to a specific range of RSRQ values in dB, as defined in (3GPP 2010). Table 3.2 shows the mapping of some RSRQ values as an example, while the full range extends from 0 to 34. The attribute `RsrqThreshold` defines the threshold for user classification. In the case of Strict FR, for instance, if a user's reported RSRQ is higher than the threshold, they are classified as a Cell-Center User (CCU); otherwise, they are classified as a Cell-Edge User (CEU). If the FFR scheme has more than two sub-bands, multiple threshold parameters are needed. Similar to the previous described attributes, Listing 3.4 shows how to configure the `RsrqThreshold` for the Strict FR.

Table 3.2: Mapping of RSRQ values to integer reports in LTE (3GPP TS 36.133).

RSRQ Reported Value	RSRQ Range (dB)
28	-6.0 $\leq$ RSRQ < -5.5
29	-5.5 $\leq$ RSRQ < -5.0
30	-5.0 $\leq$ RSRQ < -4.5
31	-4.5 $\leq$ RSRQ < -4.0
32	-4.0 $\leq$ RSRQ < -3.5
33	-3.5 $\leq$ RSRQ < -3.0
34	-3.0 $\leq$ RSRQ

```
Config::Set("/NodeList/*/DeviceList/*/"
            "$ns3::LteEnbNetDevice/"
            "ComponentCarrierMap/*/LteFfrAlgorithm/"
            "$ns3::LteFrStrictAlgorithm/RsrqThreshold",
            UintegerValue (30));
```

Listing 3.4: Example of attribute configuration for the RSRQ threshold in Strict FR.

3.4 Evaluating the impact of hotspots

The appearance of hotspots has become more frequent due to the higher number of users and numerous events that are drawn to big cities, such as fairs, gastronomic and sport events, riots, protests, public manifestations, or even car accidents. Many of these are not predictable and thus difficult to plan for, as discussed in Section 2.4.5. Besides potentially increasing the number of active users in the network and generating more traffic demand, hotspots also create a spatial resource management problem, forming zones that demand more radio resources and increase the likelihood of ICI.

Before exploring solutions for ICI in hotspot scenarios, this section presents a study on how hotspots can impact a mobile network depending on its location. This set of simulations investigates how bandwidth and user classification can impact in hotspot scenarios when using the Strict FR. As discussed in Section 2.5.1, these parameters are critical to the network performance in interference-limited scenarios and its better understanding facilitates the conception of intelligent RAN solutions.

Its main goals are: (i) Study how the hotspot position can impact system throughput and SINR; (ii) Investigate the interaction between user classification and bandwidth allocation, and how they impact the performance of FFR; and (iii) Provide numerical and analytical results that can guide the development of the final solution.

3.4.1 Scenario

Multiple simulations were defined based on the scenarios dense urban and urban coverage for massive connection from 3GPP (3GPP 2024). The baseline scenario has 3 base stations with omnidirectional antennas without introducing a wrap-around. Three hotspots are deployed with varying location and their radii are 5% of the inter-site distance (ISD). A group of users is distributed uniformly throughout the entire scenario, and a group of users is distributed uniformly inside each hotspot. Each hotspot has the same number of users.

Within each scenario, four distinct deployment configurations were defined to vary the placement of the hotspots. For easier reference, the scenario without hotspots is named Deployment 1. In Deployments 3 and 5, the hotspots are located closer to the base stations and in Deployments 2 and 4, the hotspots are located at the cell-edges, as illustrated in Figure 3.5. In Deployments 3 and 5, the distance between the hotspots and the closest base station is 10% of the ISD. In Deployments 2 and 4, the hotspots are located at the extreme edge of each cell. Every scenario proposed is simulated for these five deployments.

Each UE is running an application for IP-based connectivity. A full-buffer is implemented using a client application on the remote host that continuously transmits UDP packets and a server application on the UE terminals that receives these packets. The data rate offered to each user is configured as a simulation parameter by the time interval between successive packet transmissions.

Table 3.3 shows the simulation parameters that are common to all scenarios and Tables 3.4 and 3.5 highlight the differences between the scenarios. The group of scenarios A are inspired in the dense urban scenario and the group of scenarios B are inspired in the

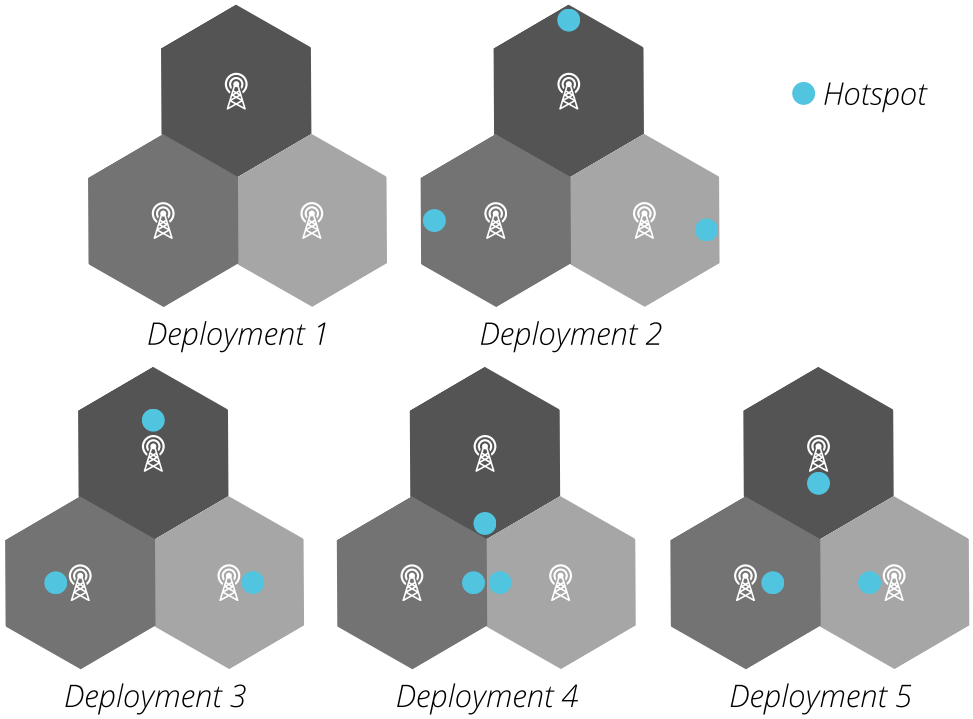


Figure 3.5: Location of the hotspots for each hotspot deployment.

urban coverage for massive connections.

Table 3.3: Simulation parameters that are common for all scenarios.

Parameter	**Value**
Scheduling algorithm	Proportional Fair
Simulation duration	5000 subframes
Channel model	Cost231
Error model	MIESM
UE mobility	no mobility
Traffic model	Non-GBR TCP-based Video

Table 3.4: Simulation parameters for group of scenarios A.

Parameter	**A1**	**A2**	**A3**
Bandwidth	20 MHz (100 RBs)		
Number of users	60: 30 uniform, 10 in each hotspot		
ISD	200 meters		
Offered data rates	500 kbps, 1 Mbps, 2 Mbps		
Sub-band allocation	28 RBs (common) 24 RBs (private)	52 RBs (common) 16 RBs (private)	64 RBs (common) 12 RBs (private)

Table 3.5: Simulation parameters for group of scenarios B.

Parameter	**B1**	**B2**	**B3**
Bandwidth	5 MHz (25 RBs)		
Number of users	120: 60 uniform, 20 in each hotspot		
Offered data rates	100 kbps, 50 kbps		
ISD	500 meters		
Sub-band allocation	6 RBs (common) 6 RBs (private)	12 RBs (common) 4 RBs (private)	18 RBs (common) 2 RBs (private)

3.4.2 Simulation results

Simulation results are discussed using throughput and SINR. Figure 3.6 shows the cell throughput for Scenario A1. Each group of bars is a different deployment. Within each deployment, each color indicates an offered data rate and the RsrqThreshold varies from 28 to 34, from left to right. Therefore, each color has 7 bars, one for each threshold value.

The results show that the difference between the threshold values increases with the offered data rate. For 500 kbps, most threshold values achieve the offered data rate for all deployments. For 1 Mbps, the threshold values provide different results but for 2 Mbps the difference becomes more significant. For instance, the threshold 34 yields 28 Mbps on Deployment 2, while the threshold 28 can only provide 15 Mbps for the same Deployment.

The threshold that provides higher throughput also changes depending on the deployment, particularly at higher offered data rates. For easier reference, Table 3.6 summarizes the achievable cell throughput for each offered data rate according to the number of active users. For 2 Mbps, the threshold 34 provides the highest throughput in Deployment 2, while in Deployment 4, this threshold yields 30% lower throughput than threshold 32. In contrast, the value 33 yields the highest throughput for Deployments 3 and 5. These results highlight the potential gain from dynamically controlling this parameter, notably when the users demand more data.

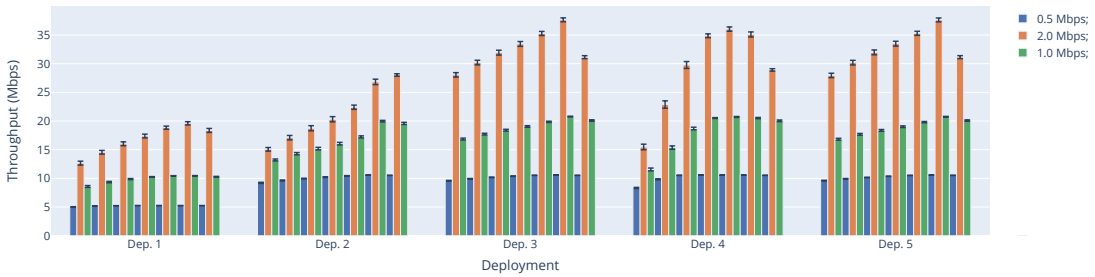


Figure 3.6: Scenario A1: cell throughput for all users.

Figure 3.7 shows the Empirical Cumulative Distribution Function (ECDF) of the SINR for all users in Deployments 3 and 4. For most thresholds, the ECDF shows a

Table 3.6: Achievable cell throughput per number of users and offered data rate.

Nr. of users	Offered data rates	Achievable cell throughput
30	500 kbps, 1 Mbps, 2 Mbps	5 Mbps, 10 Mbps, 20 Mbps
60	500 kbps, 1 Mbps, 2 Mbps	10 Mbps, 20 Mbps, 40 Mbps
	50 kbps, 100 kbps	1 Mbps, 2 Mbps
120	50 kbps, 100 kbps	2 Mbps, 4 Mbps

0.57 probability that users have less than 18 dB of SINR and a 0.4 probability that users have SINR higher than 23 dB. This gap in performance is likely due to the existence of hotspots and it happens even when hotspots are closer to the base station (Deployment 3). However, the results present an interesting situation, since the hotspots from Deployment 4 are located in the central edges, where interference is potentially higher, they have lower SINR, particularly the worst-performing users.

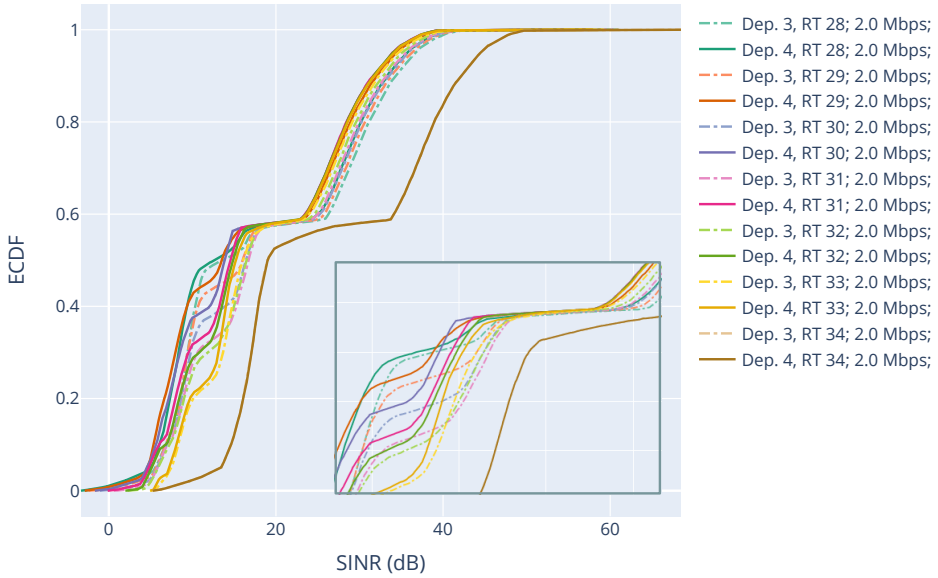


Figure 3.7: Scenario A1: ECDF of Average SINR for all users in Deployments 3 and 4.

This ECDF also shows that the threshold 34 has much higher SINR for all deployments. After inspecting the RSRQ values reported by all users, results showed that no user has RSRQ equal to 34 for any of the proposed scenarios. Consequently, when the threshold is 34, all users are classified as CEUs, leaving the common sub-band empty. If all users are allocated to the private sub-band, they do not share bandwidth with neighboring cells, which leads to higher SINR. However, that can still lead to lower cell throughput, since only $1/3$ of the bandwidth is used. This behavior is observed for Deployments 3, 4 and 5 in Figure 3.6, where the higher SINR did not lead to higher

throughput.

As an example, Figure 3.8 presents the same scenario without FFR. It shows the cell throughput for all users with all cells sharing the whole bandwidth. For Deployments 1, 2, 3 and 5, the full frequency reuse yields the same cell throughput as the corresponding best threshold value when using the Strict FR. However, for Deployment 4, the SFR is able to improve the throughput by more than 20%. This deployment is the most impacted by interference because the hotspots are closer to one another. This result indicates that an intelligent RAN solution could dynamically turn off the FFR depending on the scenario, in addition to choosing the threshold value. For example, if the scenario suddenly changed from Deployment 3 to 4, the intelligent solution would have to switch to the Strict FR and choose the threshold 32.

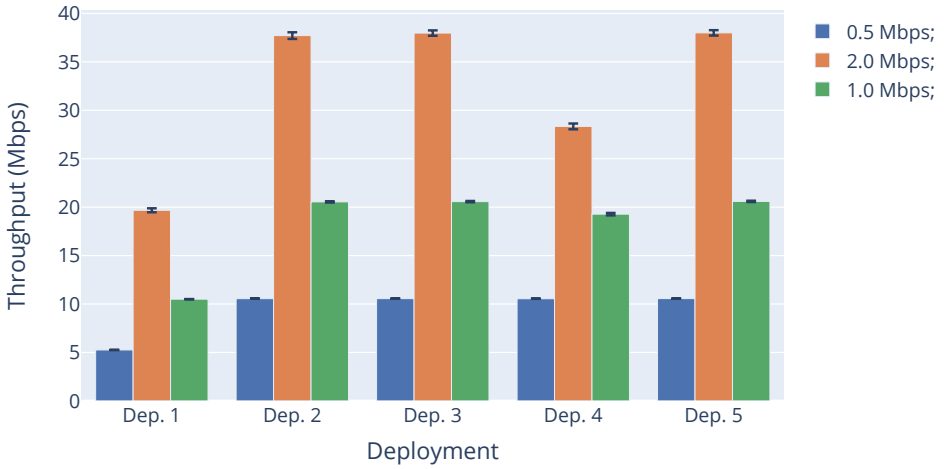


Figure 3.8: Scenario A1: cell throughput for all users without FFR.

Sequentially, Figure 3.9 shows the cell throughput for Scenario A2. This scenario allocates less RBs to the CEUs when compared to Scenario A1, as can be seen from Table 3.3. Consequently, the threshold 34 is no longer the best option in any of the deployments, providing significantly lower throughput for all cases. The lower number of RBs allocated to the private sub-band leads to poor performance if all users are classified as CEUs.

Fig. 3.10 shows throughput results for Scenario A3, which has 12 RBs allocated to the private sub-band and 64 RBs allocated to the common sub-band. Compared with the other scenarios, by giving more bandwidth to the CCUs, Deployment 2 was able to achieve even higher throughput with threshold 32, while maintaining good performance in Deployments 3 and 5, but worse results in Deployment 4 when compared to Scenario A1.

These results show not only a change in the best threshold between deployments, but also that, by changing both the bandwidth allocation (BA) and the threshold, it is possible to achieve higher throughput. For example, in Scenario A3, Deployment 2 achieves higher

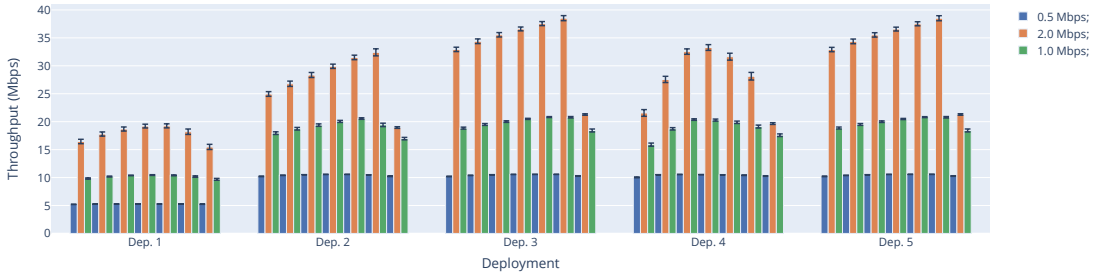


Figure 3.9: Scenario A2: cell throughput for all users

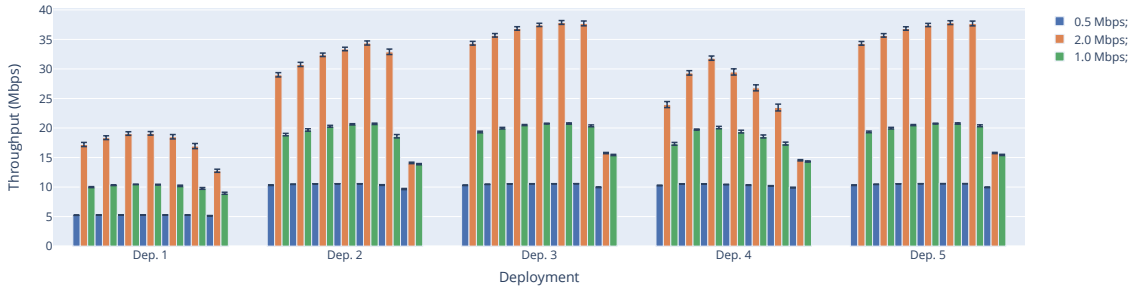


Figure 3.10: Scenario A3: cell throughput for all users.

throughput with the threshold 32, which is 34.8% higher than the higher throughput for the same Deployment in Scenario A1, which happens with threshold 34. Lastly, the results for 1 Mbps across all scenarios show that some bandwidth distributions can offer a wider range of viable threshold values. For example, if higher cell throughput is not the only metric being considered, Scenario A3 could operate on thresholds 29 to 32 for Deployment 2, while Scenario A1 only performs well with thresholds 34 and 33.

Figures 3.11, 3.12, and 3.13 present the cell throughput for Scenarios B1, B2 and B3, respectively. Since they were inspired in the Urban Coverage for Massive Connections scenario from 3GPP (3GPP 2024), they have the double amount of users, lower offered data rates, lower bandwidth and bigger ISD (see Table 3.3). The difference between these three scenarios is the bandwidth allocation, similar to the group of scenarios A.

Similar to previous scenarios, Deployments 2 and 4 have lower throughput in general because the hotspots are positioned farther from the base stations. For 50 kbps, only Scenario B2 (Figure 3.12) achieves the offered data rate for all threshold values, suggesting that this distribution offers greater flexibility in selecting thresholds. In general, allocating less bandwidth to the cell edge results in lower throughput when using higher thresholds. For instance, the best threshold for Deployment 4 is 31 in Scenario B1 and 28 in Scenario B3. This behavior is also observed in Scenarios A1, A2 and A3, but the scenarios with less bandwidth (Scenarios B) are more sensitive to changes in the bandwidth, as the number of available RBs is significantly lower.

Figure 3.11: Scenario B1: cell throughput for all users (RT is short for RsrqThreshold).

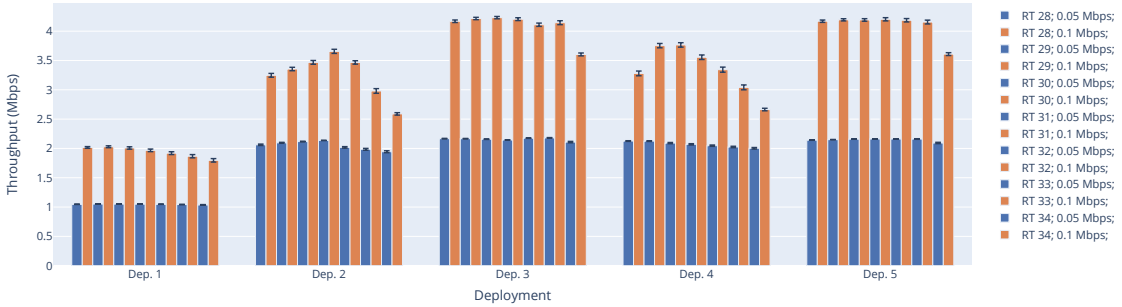


Figure 3.12: Scenario B2: cell throughput for all users (RT is short for RsrqThreshold).

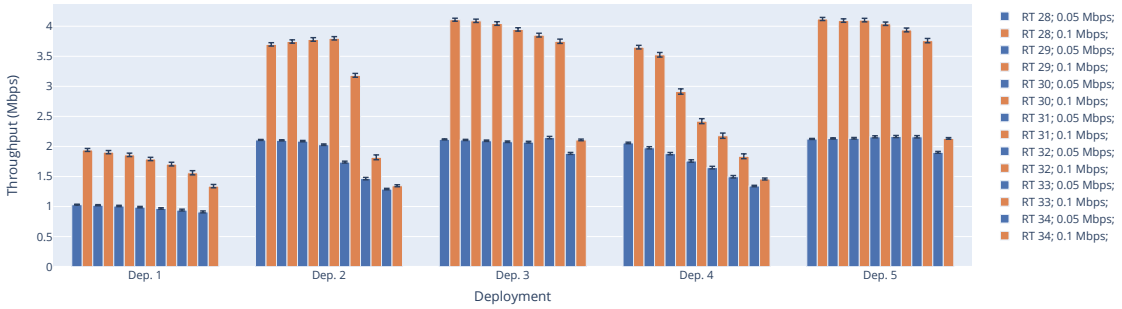


Figure 3.13: Scenario B3: cell throughput for all users (RT is short for RsrqThreshold).

Additionally, the combination of threshold and bandwidth allocation that yields the highest throughput varies with the deployment. For example, in Deployment 2, Scenario B3 (Figure 3.13) yields the highest throughput with threshold 31 and offers a broader range of well-performing threshold values compared to Scenarios B2 and B1. In Deployment 4, Scenario B1 achieves the best performance with threshold 31, whereas in Deployment 3, Scenario B1 yields the highest throughput with threshold 32. This

indicates that controlling only one of these parameters is insufficient to achieve optimal performance. Moreover, even with much lower demand and a larger ISD, reduced bandwidth can result in significant performance drops when changing the sub-band allocation. Therefore, in scenarios with limited bandwidth, fine-tuning the allocation of RBs to each sub-band is especially critical.

Lastly, we present Tables 3.7, 3.8, 3.9 and 3.10 to summarize the discussed results and highlight the differences between deployments and scenarios. The first four columns of each table indicate the highest throughput achieved in each deployment with the corresponding bandwidth allocation and threshold. Refer to these columns to identify the best combinations of bandwidth allocation and threshold for each deployment. Tables 3.7 and 3.9 compare the highest throughput in each deployment with the highest throughput achieved by other bandwidth allocations, regardless of the threshold, for scenarios A and B, respectively. Tables 3.8 and 3.10 present the gain from changing the threshold after a change in bandwidth allocation. For example, Table 3.7 indicates that if the system changed from Deployment 1 to 2, changing the bandwidth allocation and threshold can result in a 22.72% throughput gain. Meanwhile, Table 3.8 shows a potential gain of 144.25% if you re-select the best threshold after changing the bandwidth allocation in Deployment 2.

Table 3.7: Group of Scenarios A: highest throughput per deployment compared to the best throughput in other bandwidth allocations.

Deployment	Bandwidth allocation (BA)	Threshold	Throughput (Mbps)	Gain compared to best throughput in A1	Gain compared to best throughput in A2	Gain compared to best throughput in A3
1	A1 (28-24)	33	19.6	-	1.82 %	2.67 %
2	A3 (64-12)	32	34.39	22.72 %	6.14 %	-
3	A2 (52-16)	33	38.53	2.34 %	-	1.74 %
4	A1 (28-24)	32	36.04	-	8.32 %	13.23 %
5	A2 (52-16)	33	38.51	2.34 %	-	1.77 %

Table 3.8: Group of Scenarios A: highest throughput per deployment compared to using the best threshold from other bandwidth allocations.

Dep	Bandwidth allocation (BA)	Threshold	Throughput (Mbps)	Best A1 threshold	Gain compared to using best A1 threshold on BA from column 2	Best A2 threshold	Gain compared to using best A2 threshold on BA from column 2	Best A3 threshold	Gain compared to using best A3 threshold on BA from column 2
1	A1 (28-24)	33	19.6	33	-	32	4.03 %	31	12.77 %
2	A3 (64-12)	32	34.39	34	144.25 %	33	4.53 %	32	-
3	A2 (52-16)	33	38.53	33	-	33	-	32	2.61 %
4	A1 (28-24)	32	36.04	32	-	31	3.36 %	30	21.06 %
5	A2 (52-16)	33	38.51	33	-	33	-	32	2.64 %

Table 3.9: Group of Scenarios B: highest throughput per deployment compared to the best throughput in other bandwidth allocations.

Deployment	Bandwidth allocation (BA)	Threshold	Throughput (Mbps)	Gain compared to best throughput in A1	Gain compared to best throughput in A2	Gain compared to best throughput in A3
1	B1 (6-6)	31	2.03	-	0.49 %	4.64 %
2	B3 (18-2)	31	3.8	3.54 %	3.94 %	-
3	B1 (6-6)	32	4.33	-	2.36 %	5.35 %
4	B1 (6-6)	31	3.99	-	6.12 %	9.31 %
5	B1 (6-6)	32	4.25	-	1.19 %	3.15 %

Table 3.10: Group of Scenarios B: highest throughput per deployment compared to using the best threshold from other bandwidth allocations.

Dep	Bandwidth allocation (BA)	Threshold	Throughput (Mbps)	Best B1 threshold	Gain compared to using best B1 threshold on BA from column 2	Best B2 threshold	Gain compared to using best B2 threshold on BA from column 2	Best B3 threshold	Gain compared to using best B3 threshold on BA from column 2
1	B1 (6-6)	31	2.03	31	-	30	2.46 %	28	11.82 %
2	B3 (18-2)	31	3.8	33	52.63 %	31	-	31	-
3	B1 (6-6)	32	4.3	32	-	30	6.51 %	28	12.56 %
4	B1 (6-6)	31	4	31	-	30	5.25 %	28	38.75 %
5	B1 (6-6)	32	4.25	32	-	31	2.12 %	28	12.71 %

In conclusion, the presented results highlight how hotspots can impact network performance and how user classification and sub-band allocation are key parameters in FFR techniques. Interference-limited scenarios with hotspots can be highly impacted when changing the threshold for user classification or the bandwidth allocation, as seen from Deployment 4, that has the biggest variations in throughput. If the threshold value is not properly tuned, a change in hotspot deployment could result in up to 30% loss in throughput. Additionally, the SINR results confirm that the existence of hotspots creates groups of users with a big performance gap, complicating resource management, since the network has to balance the resources to satisfy such different users. Nevertheless, a dynamic solution that takes only SINR into account could end up with unused bandwidth due to misconfiguration, while tuning both the user classification and sub-band allocation can potentially lead to throughput gain. Lastly, the results demonstrated that scenarios with less bandwidth are more sensitive to changes in user classification and sub-band allocation.

Chapter 4

Proposed solution

The comprehensive review of state-of-the-art techniques in Chapter 1 and the detailed discussions on inter-cell interference in Chapter 2 highlight the persistent challenge of interference management in mobile networks. While resource allocation strategies are often employed to mitigate ICI, dynamic hotspot scenarios introduce many challenges that need real-time adaptation. Chapter 3 introduced the problem characterization and identified dynamic FFR approaches as a promising solution, especially when multi-parameter optimization is required.

These discussions were complemented by the results presented in Section 3.4, which demonstrated that hotspots can have a strong impact o n network performance, particularly if closer to the cell edges, where users are more susceptible to interference. When considering a network using FFR, the correct configuration of user classification and bandwidth allocation can greatly improve performance.

This chapter introduces a novel Machine Learning-based solution that applies RL to dynamically adjust network parameters in real time. The next sections detail some of the theoretical background, implementation strategy, and expected performance enhancements of the proposed solution, demonstrating its potential to improve network throughput and overall system efficiency in complex, real-world environments.

4.1 Machine Learning

The dynamic challenges presented by urban cellular networks, particularly the unexpected emergence of hotspots, demand solutions that can adapt in real time. Machine Learning, a subfield of artificial intelligence focused on extracting patterns from data to inform decision-making with minimal human intervention, offers promising tools for addressing these challenges.

ML techniques have been widely studied for their potential to automate and optimize network performance. Recent research has vastly explored ML-based solutions for various interference scenarios, applications and use cases (Trabelsi et al. 2024). These techniques generally fall into three categories: Supervised Learning, Unsupervised Learning, and Reinforcement Learning (RL). In Supervised Learning, a teacher provides input-output pairs, enabling the model to learn to predict outcomes based on labeled data. While effective in many scenarios, supervised methods require extensive prior knowledge

and dedicated training efforts, making them less suitable for highly dynamic environments where conditions can shift rapidly. Unsupervised Learning only has access to unlabeled data, as it has no entity that has knowledge about the system or the data. Its main goal is to find patterns or hidden structures in the data, such as grouping similar inputs or identifying anomalies.

Reinforcement Learning, by contrast, has an agent that interacts with the environment, making decisions that result in rewards or penalties. This feedback loop allows the model to continuously adjust its parameters to maximize cumulative rewards (Haykin 2009). For cellular networks facing the dynamic challenge of hotspot emergence, RL enables the system to continuously adjust critical parameters, aiming at maximizing any desired metric that reflects system performance. It is an efficient learning paradigm for interacting with unknown environments (Richard S. Sutton 2018), allowing real time decision making.

4.2 Reinforcement Learning

The author (Richard S. Sutton 2018) defines Reinforcement Learning as a computational approach to understanding and automating goal-directed learning and decision-making. It differs from other computational approaches by the existence of a learning agent that is trying to maximize a numerical reward over time. Instead of requiring system models or a teacher, as in supervised learning, it learns by directly interacting with the environment (Richard S. Sutton 2018).

Reinforcement learning problems have two main components: the agent that learns and makes decisions, and the environment, which is composed by all external entities outside the agent, that is, what the agent is interacting with. The line that separates the agent from the environment is not always intuitive. In general, anything that the agent cannot change arbitrarily is part of the environment, since it is not part of the agent. This ensures that the agent must learn to influence the environment indirectly through its actions, rather than directly controlling the external conditions.

The interaction between agent and environment occurs over a sequence of discrete time steps, denoted $t = 0, 1, 2, 3, n$. Unlike supervised learning, the agent is not given explicit instructions on which action to take. At each time step, the agent detects the current state of the environment $s_t \in S$ and chooses an action $a_t \in A$ according to its current decision-making strategy and applies that action to the environment. The environment, which might transition to a different state $s_{t+1} \in S$ given the agent's action, provides a numerical reward r_{t+1} after one time step. This interaction between agent and environment is illustrated in Figure 4.1.

The agent's decision-making strategy is formally represented as the policy π, a mapping from states to probabilities of selecting each possible action (Richard S. Sutton 2018). It maps each state $s \in S$ to a probability distribution over the available actions. Specifically, $\pi(a|s)$ is the probability of choosing an action a when the current state is s. In a deterministic policy, this distribution assigns probability 1 to a single action for each state, while in a stochastic policy, multiple actions may have non-zero probabilities, allowing for the exploration of alternative actions. Moreover, the policy is

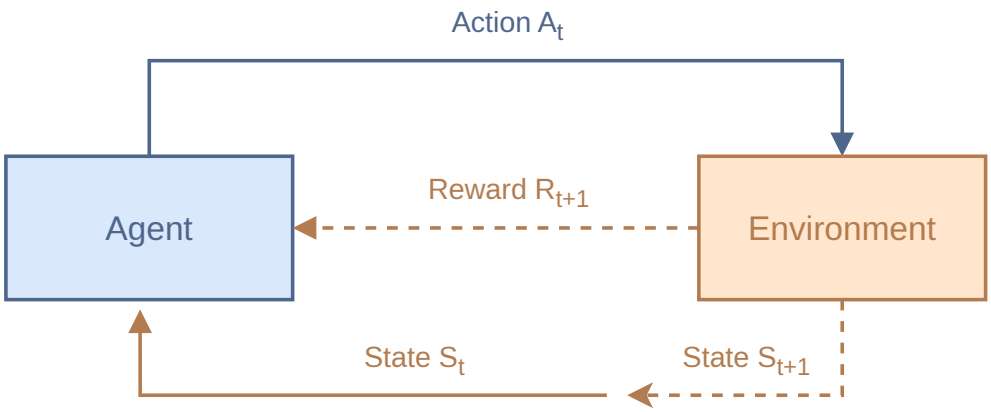


Figure 4.1: Interaction between the agent and the environment in Reinforcement Learning problems.

not fixed. Instead, it is updated at each time step through successive interactions with the environment, favoring actions that lead to higher cumulative rewards. The specific policy update strategy depends on the RL method employed.

According to (Richard S. Sutton 2018), the use of a reward signal as a formal representation of the agent's goal is a unique feature of reinforcement learning. The reward is the primary mechanism for communicating the end goal to the agent, which aims at maximizing not only the immediate reward, but the cumulative sum of future rewards. This flexible approach can be used for various domains and applications. For example, an agent learning how to drive a car in a racing game could get negative rewards whenever it bumps into another car and a positive reward for each successful overtake. However, the signal should directly represent the agent's final goal, instead of any of its sub-tasks. For instance, the former example could lead the agent to lose the race, since it was only told to avoid crashing and overtake cars.

In wireless networks, most tasks are characterized by a continuous interaction between the agent and the environment without a predefined endpoint. In such cases, the cumulative reward is better defined with the use of discount factors, since a continuous task could lead to an infinite reward. As defined in (Richard S. Sutton 2018), Equation (4.1) defines the cumulative reward G_t as the sum of the subsequent rewards, adjusted by a discount factor γ. In this formulation, future rewards contribute less than immediate rewards, and, as long as γ is lower than 1, and the reward sequence is bounded, the sum is finite. For instance, if γ is set to zero, only the immediate reward is considered.

$$G_t \doteq R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 R_{t+4} + \cdots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}. \tag{4.1}$$

Another important aspect of Reinforcement Learning is the representation of the environment's state. While the reward signal conveys to the agent its final goal, the

state signal is a collection of information about the environment that is useful for the agent's decision making. The state can contain various types of information, including current or past measurements, combinations of different metrics, or even the result of post-processing raw data into higher-level features. The state does not need to completely describe the environment, but a key assumption in most RL formulations is that the state retains all of its relevant past information. That is described as the Markov property. It states that the future state of a process depends only on its present state, not on the sequence of states that preceded it. Consequently, the next state s_{t+1} and reward r_{t+1} depend only on the current state s_t and action a_t. This assumption allows RL algorithms to predict the next state and the expected reward using only the current state and action, highlighting the importance of the state representation.

As a result, this property allows the RL problem to be modeled as a finite Markov Decision Process (MDP), if the states and actions spaces are finite. Formally, a finite MDP can be specified by the state and action spaces, the reward function, and the state transition probability function, presented in Equation (4.2), as defined in (Richard S. Sutton 2018).

$$p(s', r \mid s, a) = \Pr\{S_{t+1} = s', R_{t+1} = r \mid S_t = s, A_t = a\}, \tag{4.2}$$

where $p(s', r \mid s, a)$ defines the probability that the environment transitions to state $s' \in S$ and gets a reward r when action $a \in A$ is taken from state $s \in S$.

Based on the concepts introduced so far, it is now possible to further detail the agent's goal in a finite MDP: learning an optimal policy π_* that maximizes the expected cumulative discounted reward from any initial state. This is quantified by a value function, that estimates future performance, guiding the agent on how good each action is given each state. Some approaches estimate the state-value function, which represents the expected return from state s while following a given policy. Another approach is using an action-value function, which represents the expected cumulative discounted reward when taking action a in state s and subsequently following policy π. This allows the agent to explicitly evaluate different actions at each state, instead of assessing the desirability of states. The action-value function is presented in Equation (4.3) as the expected value of G_t, presented in Equation (4.1), given that policy π is followed.

$$q_\pi(s, a) = \mathbb{E}_\pi\big[G_t \mid S_t = s, A_t = a\big] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,\middle|\, S_t = s, A_t = a\right]. \tag{4.3}$$

The corresponding optimal value function $q_*(s, a)$, presented in Equation (4.4), yields the maximum achievable expected return under the optimal policy π_*.

$$q_*(s, a) = \max_\pi q_\pi(s, a). \tag{4.4}$$

Hence, the action-value function estimates how good a given action is when the environment is at a given state, accounting for both immediate and future rewards accumulated over time, when following a particular policy. However, computing the value function directly by summing future rewards is infeasible in most practical cases. Instead,

given the recursive nature of these value functions, the Bellman equation provides the means of approximating them recursively, by averaging all the possibilities weighted by their occurrence probabilities. The bellman equation for q_π is presented in Equation (4.5) and its optimality counterpart is presented in Equation (4.6).

$$q_\pi(s,a) = \sum_{s',r} p(s',r|s,a) \left[r + \gamma \sum_{a'} \pi(a'|s') q_\pi(s',a') \right]. \tag{4.5}$$

$$q_*(s,a) = \sum_{s',r} p(s',r \mid s,a) \left[r + \gamma \max_{a'} q_*(s',a') \right]. \tag{4.6}$$

4.2.1 Q-Learning

Q-Learning is an off-policy reinforcement learning algorithm designed to solve the RL problem by learning an optimal action-value function $Q_*(s,a)$, which estimates the maximum expected cumulative reward for each state-action pair. The Q function directly approximates q_* independently of the policy being followed (Richard S. Sutton 2018). The main requirement for correct convergence is the constant update of the Q values, while the policy determines how the agent will explore the state and action spaces to update its values. The algorithm iteratively refines its estimates of $Q(s,a)$ based on interactions with the environment, enabling the agent to learn the optimal policy π_* without requiring a model of the environment's dynamics.

Q-Learning operates by maintaining a Q-table that stores estimated action values $Q(s,a)$ for each state-action pair. $Q(s,a)$ is the value q of action a when executed from state s, which serves as predictors of long-term reward. A higher $Q(s,a)$ value indicates a more desirable action a when the system is in state s.

The update rule for the action-value function defined by Q-Learning is derived from the Bellman optimality equation. Instead of requiring the state transition probabilities $p(s',r|s,a)$ and the expected rewards, the algorithm approximates these values using samples of the agent's experience from interacting with the environment. After taking an action, the agent can update the q values using the observed reward at the following time step. The Q-function update rule is presented in Equation (4.7).

$$Q(S_t,A_t) \leftarrow Q(S_t,A_t) + \alpha[R_{t+1} + \gamma \max_a Q(S_{t+1},a) - Q(S_t,A_t)]. \tag{4.7}$$

Each time the Q-table is updated, the new q value is computed as a weighted sum of the existing q value (the agent's prior knowledge) and the new learned value. The learning rate α controls how much the learned value impacts the q value update. Higher values of α provide a faster learning, and lower values make learning more stable but slower. R_{t+1} is the reward received after taking action A_t in state S_t, and γ is the discount factor, which determines how much future rewards impact the learning, with values closer to 0 meaning future rewards have less importance and values closer to 1 giving more importance to long-term rewards. Lastly, $\max_a Q(S_{t+1},a)$ is the best q value in the new state S_{t+1}.

(Watkins 1989) demonstrates that, given sufficient exploration and a large enough number of samples, the algorithm converges to the optimal action-value function directly

derived from the Bellman equation, which is the classic solution for an MDP. Once the Q-table has been sufficiently trained, the agent chooses the action that maximizes $Q(s,a)$ for each state s. This policy is proven to approximate the optimal solution to finite MDP as the number of interactions with the environment increases (Watkins 1989).

4.3 Multi-agent hierarchical RL-based ICIC

Building on the insights gained from the analysis of related works, the challenges posed by inter-cell interference in hotspot scenarios, and the foundations of reinforcement learning and Q-Learning, this section formally presents the proposed RL-based ICIC solution. The objective is to maximize user throughput in interference-limited hotspot scenarios, using a reinforcement learning approach to adaptively mitigate interference and enhance network performance in real time.

The main goal is to design a framework that enables the dynamic control of two parameters of the Strict FR. The first parameter is the threshold that governs user classification, which determines how users are assigned to different sub-bands, and the second parameter is the number of available PRBs in each sub-band. RL agents are particularly well suited for multi-agent operation due to their simple and efficient learning process (Busoniu et al. 2006), as described in previous sections. Some environments are naturally prone to the use of multiple agents due to the existence of multiple objectives, while in other cases, multi-agent systems can be used to solve complex tasks or achieve common objectives through cooperation (Liang et al. 2025).

Cooperation among agents can be achieved through various mechanisms, often involving the exchange of information to enhance collective performance. Whenever multiple agents interact within the same environment, their dependencies must be carefully considered to achieve the global objective, which can be modeled either as a cooperative or a competitive task. In general, multi-agent algorithms are classified as fully cooperative, fully competitive, or mixed, depending on how their reward function is modeled (Liang et al. 2025). Fully competitive agents receive opposing rewards, meaning that one agent tries to maximize its own reward while minimizing the reward of competing agents. In fully cooperative settings, the agents share a common reward function, working collectively toward a common goal (Liang et al. 2025). In mixed tasks, the relationship between agents' rewards is not entirely correlated. The agents may have both cooperative and competitive interactions, depending on the scenario.

Additionally, a multi-agent problem can be characterized as a stochastic game, also referred to as a Markov game (Busoniu et al. 2006). The MDP is extended to a multi-agent setting, where multiple agents interact within the same environment. Formally, it consists of a set of states, a set of actions for each agent, a transition probability function, and a reward function that depends on the joint actions of all agents. Unlike single-agent RL, where the environment evolves solely based on an individual agent's decisions, in a stochastic game, the next state and rewards are influenced by the combined actions of all agents, leading to a dynamic and interdependent learning process. This requires coordination and a careful balance between stability and adaptation. Stability ensures that the agents converge to an optimal policy, while adaptation involves the adjustment to

the dynamic behavior of other agents, ensuring the maintenance or improvement of the performance (Busoniu et al. 2006).

In this context, given the strong interdependency between user classification and bandwidth allocation, this work uses a fully cooperative approach with indirect coordination to manage the cooperation between the two independent agents, without an explicit mechanism to enforce joint action decisions. Instead, the proposed framework defines a hierarchical structure that introduces an implicit bias in the decision process. One of the agents is considered to have a higher level of control, acting first, while the second agent acts multiple times before the first agent acts again, repeating the cycle. This creates a sequential dependency, where the second agent makes a decision within the constraints imposed by the decision of the first agent.

Hence, this hierarchical multi-agent approach has two cooperative Q-Learning agents. The first agent controls the `RsrqThreshold` and the second agent controls the parameters related to the sub-bands PRB allocation. Both agents operate in a semi-centralized mode, as their decisions affect a cluster of three cells, considering a scenario with omnidirectional antennas in hexagonal cells. This structure is presented in Figure 4.2. Agent 1 makes decisions first, which can result in users being reclassified as cell-center users (CCUs) or cell-edge users (CEUs) and being assigned to different sub-bands. Once this classification decision is made, Agent 2 acts on the bandwidth allocation, adjusting the amount of PRBs in each sub-band over the next four consecutive time steps.

Both agents share the same reward space, meaning they do not only learn from their direct interaction with the environment but also from the joint impact of their decisions. Despite this hierarchical cooperation, the agents remain independent entities, each maintaining its own Q-table and action set. Although they influence each other's learning process, their decision-making processes remain distinct and evolve separately.

One of the objectives of this work is to provide information on the challenges of developing dynamic RL-based solutions to manage interference in mobile networks. This chapter introduces the baseline configuration of the agents, which was able to improve throughput in the proposed dynamic scenario. However, certain aspects of this configuration will be varied in subsequent analyzes to investigate their impact on system performance and to evaluate the limits of the proposed solution. The next chapter will discuss the results obtained with this baseline set-up and additional simulation campaigns designed to evaluate the impact of some key parameters, such as the state representation.

The initial configuration for each agent was defined based on the background presented in the previous section, which starts with the definition of the following parameters.

- A definition of system states, $S = s_1, s_2, ..., s_n$;
- A set of possible actions, $A = a_1, a_2, ..., a_n$;
- A Q-matrix $Q(s,a)$ to store the estimated action-value function;
- The learning rate α and discount factor γ;
- A reward signal R.

The metric used to define the states of the system is usually in alignment with the objective of the solution. In mobile networks, it is common to use performance metrics

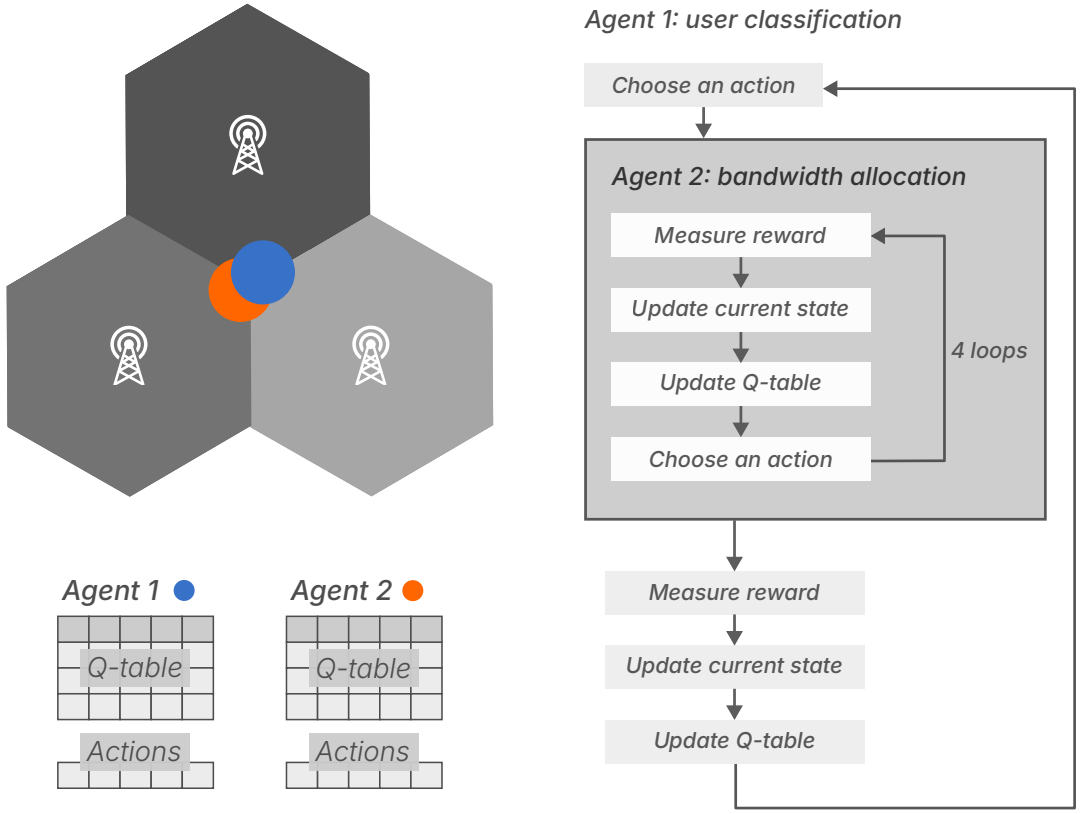


Figure 4.2: Hierarchical structure of the proposed multi-agent framework.

such as throughput, SNR, PRB occupation, etc. As described in Section 4.2, the state information should be useful for the agent's decision making, but it does not need to fully represent the system. Given the focus on improving performance, the states were defined using throughput. Since it is possible to measure the amount of data transmitted in the network, the states can be defined as a percentage of the maximum achievable throughput. For simulation purposes, the traffic is modeled as a full-buffer, hence, the states are defined as a percentage of the offered data rate, as presented in Table 4.1.

One of the challenges of using throughput for real-time solutions is whether the measured received bytes are enough to represent the current state of system. If the throughput is measured over short time windows, it may not accurately reflect the system's overall performance, as some users may not have transmitted during the last time step of the agent. Relying solely on instantaneous measurements could lead to misleading state estimations. To address this, throughput is computed over a fixed-size window, where each newly observed value replaces the oldest recorded value. The system throughput at any given time is then determined as the average of all values within the window, providing a more stable and representative measure of network performance. For example, if the window size is set to four, the throughput at each step is calculated as the mean of the four most recent throughput values, ensuring that short-term fluctuations

Table 4.1: State definition based on throughput as a percentage of the offered data rate.

State	Throughput Range (% of Offered Data Rate)
0	$tput \leq 20\%$
1	$20\% < tput \leq 30\%$
2	$30\% < tput \leq 40\%$
3	$40\% < tput \leq 50\%$
4	$50\% < tput \leq 60\%$
5	$tput > 60\%$

do not overly influence the agent's decision-making.

The action set is particular to each agent, in contrast with the states. Agent 1, which acts on the user classification, has a total of 6 possible actions and Agent 2, which acts on bandwidth allocation, has a total of either 6 or 3 possible actions, depending on the scenario, which is further explained in the following sections. The number and values of these actions were chosen based on previous works, such as (do Rêgo et al. 2018) and (do Rego & de Sousa Jr. 2021), and exploratory analyses, such as the one presented on Chapter 3. Tables 4.2 and 4.3 present the action set for both agents. For the bandwidth, the first number is the amount of RBs for the cell-center and the second number is the amount of RBs for each cell-edge.

Table 4.2: Action Set for Agent 1: RSRQ Threshold.

Action	RSRQ threshold
0	28
1	29
2	30
3	31
4	32
5	33

Table 4.3: Action Set for Agent 2: Bandwidth Allocation.

Action	Bandwidth allocation (RBs)
0	4-32
1	28-24
2	40-20
3	52-16
4	64-12
5	88-4

To balance the exploration and exploitation during the learning process, an ε-greedy policy is used. This policy determines that the agent's standard behavior is to select the action with the highest estimated q-value (exploitation), while choosing a random action

with a probability of ε (exploration). If the agent only exploits the highest q-value, it might get stuck in suboptimal policies. Instead, this random exploration may lead to higher long-term rewards. However, a high ε might add noise to the learning process, slowing it down.

Consequently, ε is set to decay exponentially over time, to encourage exploration in the initial learning phase, when the agent has limited knowledge about the environment. As time goes on, and ε decays, the agent explores less, shifting its behavior toward exploitation. Additionally, multi-agent systems can be particularly unstable if ε is high. Since the agents' exploration behaviors influence each other indirectly, frequent random actions from one agent can disrupt the learning process of the other agent, leading to unstable learning dynamics. Hence, ε_0 is kept low for both agents, set at 0.2. The described policy is formally presented in Equation (4.8). $\pi(a|s)$ is the probability of selecting action a in state s, and n is the number of possible actions. Equation (4.9) illustrates the decay of ε, where ε_t is the effective value of ε after t actions were taken.

$$\pi(a|s) = \begin{cases} 1-\varepsilon, & \text{if } a = \arg\max_{a'} Q(s,a') \\ \frac{\varepsilon}{n}, & \text{otherwise.} \end{cases} \tag{4.8}$$

$$\varepsilon_t = \frac{\varepsilon_0}{(1.06)^t}. \tag{4.9}$$

Another crucial parameter to configure is the reward signal. As detailed in Section 4.2, the reward signal is the primary mechanism to communicate the end goal to the agent. As the framework aims at maximizing throughput in a scenario with high co-channel interference, both throughput and SINR were considered as a reward signal. While SINR directly reflects interference levels, relying solely on SINR for the agent's decision-making may lead to lower throughput.

Since the Strict FR assigns disjoint sub-bands to cell-edge users, the agent can find ways to increase SINR at the cost of throughput. For example, the agent could demonstrate this behavior in the scenario explored in Section 3.4, where the RSRQ threshold of 34 significantly increased SINR but resulted in lower overall throughput. The reason for this drop was that all users were allocated to the cell edge, reducing the allocated resources to only one-third of the system bandwidth.

To better align with the agent's objective, throughput was chosen as the primary component of the reward function. However, using system throughput could lead to higher rewards when the number of active users increases, without necessarily improving user throughput or reducing interference. To address this, the reward is computed as the total received bytes averaged over the number of active users, using the same fixed-size window strategy applied to state representation.

Although a reward function based solely on user throughput could potentially lead agents to overlook users with low SINR, this concern is mitigated by the Strict FR scheme, which provides stronger protection for cell-edge users through the use of private sub-bands. As a result, the default configuration of the reward function relies exclusively on user throughput. However, in scenarios where the traffic model is not full-buffer, it may become necessary to explicitly incorporate SINR into the reward. To accommodate

such cases, the reward is formally defined as a weighted sum of throughput and SINR, with the default weight for the SINR component set to zero. In addition to evaluating the baseline configuration, the results section also presents a comparison of different reward function strategies: throughput only, SINR only, and a combination of both.

Additionally, to better guide the agent's decision-making, the reward signal is expressed as the difference between consecutive measurements. This formulation provides clearer feedback: a positive reward is assigned when an action improves performance compared to the previous measurement, while a negative reward is assigned otherwise.

Each time the agent acts, the user throughput T_t in Mbps is calculated as in Equation (4.10).

$$T_t = \frac{8 \cdot B_t}{U_t \cdot s \cdot 1000000}, \tag{4.10}$$

where B_t is the total received bytes, U_t is the number of active users, and s is the duration of the sampling interval in seconds. The average user throughput $\bar{T}_t$ is calculated using the last W time step measurements, as in Equation (4.11).

$$\bar{T}_t = \frac{1}{W} \sum_{i=t-W+1}^{t} T_i, \tag{4.11}$$

where W is the number of time steps in the window. The average SINR is calculated as in Equation (4.12).

$$SINR_t = \bar{S}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} S_i. \tag{4.12}$$

S_i is the SINR sample i and N_t is the number of SINR samples collected during time step t. The reward signal is then calculated as presented in Equation (4.13).

$$R_t = w_T \cdot (\bar{T}_t - \bar{T}_{t-1}) + w_S \cdot (\bar{S}_t - \bar{S}_{t-1}). \tag{4.13}$$

w_T and w_S are the assigned weights for the throughput and SINR, respectively.

Lastly, both agents initialize the Q-table to zero and the values for γ and α are initially configured as 0.4 and 0.3, respectively. Algorithm 1 presents the pseudo-code of the proposed multi-agent framework.

Algoritmo 1: Hierarchical Multi-Agent Q-Learning for ICIC Optimization

Initialize
for $s \in S, a \in A$ **do**
    Initialize $Q_1(s,a)$ and $Q_2(s,a)$ tables with all values set to zero;
end
Observe initial state $s \in S$;
Learning Process:
for *each time step* t **do**
    User Classification Agent (High-Level Decision);
    if $t\%3 == 0$ **then**
        Generate random number $u \in U(0,1)$;
        if $u < \varepsilon_t$ **then**
            $\varepsilon_{t+1} \leftarrow \frac{\varepsilon_t}{1.02}$;
            Select action a_1 randomly from A_1;
        else
            Select action $a_1 = \arg\max_{a'} Q_1(s,a')$;
        end
        Apply action a_1, updating user classification threshold;
    end
    Bandwidth Allocation Agent (Low-Level Decision);
    Generate random number $u \in U(0,1)$;
    if $u < \varepsilon_t$ **then**
        $\varepsilon_{t+1} \leftarrow \frac{\varepsilon_t}{1.02}$;
        Select action a_2 randomly from A_2;
    else
        Select action $a_2 = \arg\max_{a'} Q_2(s,a')$;
    end
    Apply action a_2, allocating bandwidth accordingly;
    Receive the shared reward r
    Observe new state $s' \in S$;
    Update Q-Tables:;
    $Q_1(s,a_1) \leftarrow Q_1(s,a_1) + \alpha \left[r + \gamma \max_{a'} Q_1(s',a) - Q_1(s,a_1) \right]$;
    $Q_2(s,a_2) \leftarrow Q_2(s,a_2) + \alpha \left[r + \gamma \max_{a'} Q_2(s',a) - Q_2(s,a_2) \right]$;
    Update state: $s \leftarrow s'$;
end

4.4 Implementation of RL algorithms in ns-3

To implement the proposed solution, several non-trivial modifications were made to ns-3. The most significant was the addition of new classes responsible for implementing Q-learning (QL) and Multi-Armed Bandit (MAB) agents, as conveniently proposed in Section 5.2. The initial version of these classes were developed by the authors (de Santana et al. 2019). In this work, both classes were greatly improved in terms of function and performance, in addition to being better integrated with ns-3.

The classes were implemented within the ns-3 *namespace* and fully integrated into the LTE module. As a result, when running a simulation script, the build system used by ns-3 detects and compiles the new files without requiring additional modifications. Both classes inherit from the `ns-3::Object` class, allowing them to use the smart pointers implemented in ns-3 to create multiple QL and MAB agents if needed, ensuring proper and automatic memory management.

The Object class adds some features to traditional C++ objects to provide greater functionality and features. Classes that inherit from this class can instantiate objects with special properties. The class can benefit from the attribute system, an object aggregation system, and a smart-pointer reference counting system (ns3 2025*b*). The latter implements automatic memory management using `Ptr<>` smart pointers, preventing memory leaks and ensuring objects are deallocated when no longer needed.

In practical terms, the class has to inherit from the `ns-3::Object` class and it must implement a static public member function called `GetTypeId`. The header file has to include `object.h` and the `.cc` file also has to include the macro `NS_OBJECT_ENSURE_REGISTERED`, since it defines a new `GetTypeId` method. This macro does the actual registration of the class into the system (ns3 2025*b*). As a result, all QL and MAB agents can be instantiated using ns-3 smart pointers, as illustrated in Listing 4.1.

```
Ptr<BsQLearning> bsQL = CreateObject<BsQLearning>(tputWindowSize,
                                                  rewardMode,
                                                  rlEpsilon,
                                                  rlAlpha,
                                                  rlGamma,
                                                  doubleQL,
                                                  "RsrqThreshold",
                                                  6, 6,
                                                  bandwidth);
```

Listing 4.1: Example of how to instantiate a QL agent.

The RL agents are entirely implemented within its own classes, ensuring modularity and maintainability. Each agent maintains its own Q-table, along with the current and previous values of key parameters such as state, reward, and action taken, all of which are essential for updating the Q-table. To minimize memory usage, reward metrics, including SINR and throughput, are tracked using a cumulative average rather than storing full histories. The class provides public methods for managing the agent's operation, such as retrieving the action with the highest reward, obtaining the current reward value, updating the system state, or storing received bytes from the connected trace source.

ns-3 is a discrete-event network simulator. Instead of running in a traditional loop,

it operates using discrete-event scheduling, where functions are executed at specified simulation times and the scheduling of new events can be triggered throughout the simulation. Additionally, scripts can manually schedule events using functions such as `Simulator::Schedule()`. The RL execution uses a recursive scheduling approach, wherein the function that operates the RL agent schedules itself again after a period of time defined as a scenario parameter. This ensures that the RL agent operates continuously throughout the simulation.

The code that defines the simulation scenario was built upon the `lena-frequency-reuse.cc` example, provided by ns-3. The original example only includes the LTE model, which does not support IP-based internet connectivity, as discussed in Section 3.2 and illustrated in Figure 3.2. The Evolved Packet Core (EPC) was added to provide better control over the offered data rate, creating a more realistic scenario. By including the EPC, it became possible to control installed applications on devices, allowing them to start and stop at specific time intervals to simulate the appearance of the hotspots in different periods of time.

Chapter 5

Results of hierarchical RL-based ICIC

This chapter presents the results of multiple simulation campaigns conducted to evaluate the performance of the proposed multi-agent hierarchical solution introduced in Chapter 4. The analysis begins with an assessment of the baseline configuration of the algorithm, followed by an evaluation of key parameters, including state representation and reward function. The solution is tested in two main scenarios, inspired by the dense urban and massive connections scenarios defined by 3GPP, available at (3GPP 2024). System performance is measured using SINR and throughput, while the effectiveness of the reinforcement learning agents is assessed through cumulative reward and action transition matrices. To demonstrate the flexibility of the framework, the last section of this chapter evaluates the hierarchical solution with new agents based on Multi-Armed Bandit (MAB), a simplified formulation of the RL problem that can provide similar results with lower complexity.

5.1 Hierarchical QL-ICIC solution

The first set of results in this section follows the exact configuration outlined in Chapter 4. This configuration was derived from exploratory analyses aimed at tuning the hierarchical framework to maximize throughput. The corresponding simulation scenario is described below.

5.1.1 Evaluation scenario

The implemented scenario is similar to the one used for the analysis presented in Section 3.4. It consists of three bases stations with omnidirectional antennas and users are deployed either uniformly across the entire area or within designated hotspot zones. The users' locations are drawn from a uniform distribution, but the hotspot users are confined to hotspots zones, a circular area with a configurable radius. Five hotspots are deployed in different locations and each hotspot has a radius of 8% of the ISD. Figure 5.1 shows the exact location of the base stations and hotspots with respect to the simulated area. The entire area is defined as a function of the ISD. The lower boundary is $0 - \frac{ISD}{2}$ and the higher boundary is $\frac{ISD}{2} + 100$. The figure shows the case for an ISD of 200 meters.

The location of the hotspots does not change during the simulation, but to create a

dynamic environment, the simulation starts with no hotspots and, at each fixed interval, one additional hotspot starts transmitting. The numbers assigned to each hotspot in Figure 5.1 show the sequence of their appearance. The aim is to use the proposed solution to adjust the FFR parameters in response to the appearance of the hotspots. Their location was planned to include hotspots close and far from the base stations and the edges of the cells.

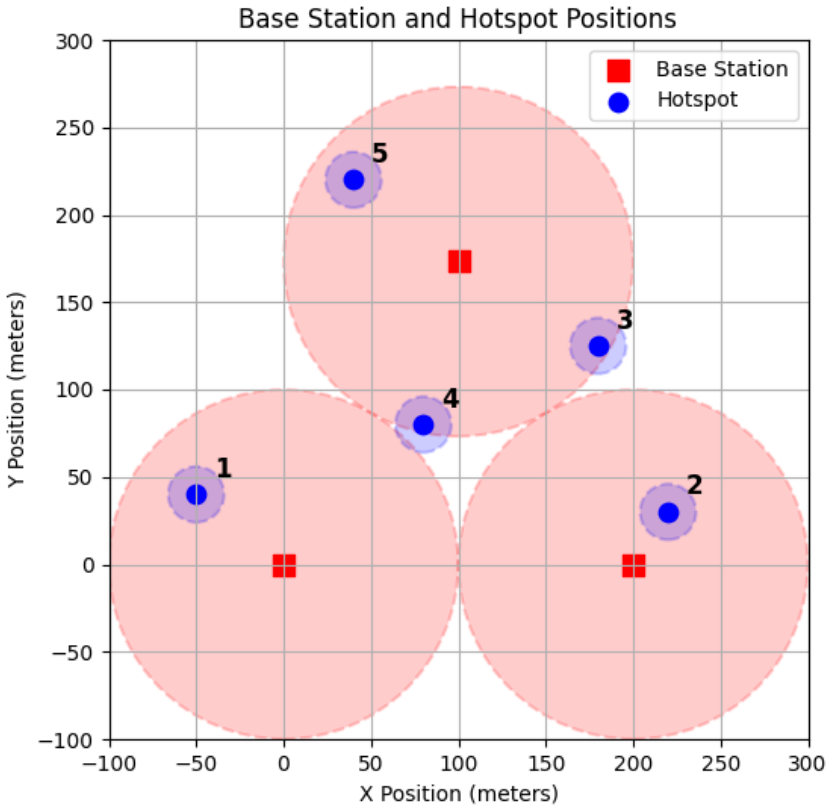


Figure 5.1: Position of the base stations and hotspots.

Two scenarios were defined based on 3GPP specifications (3GPP 2024). For clarity in the following discussions, these scenarios are referred to as Dense Urban (Scenario A) and Massive Connections (Scenario B). The Dense Urban scenario features a system bandwidth of 20 MHz, 30 uniformly distributed users, and 10 users per hotspot. It has an ISD of 200 meters and an offered data rate of 2 Mbps. The Massive Connections scenario includes a larger number of users, with 60 uniform users and 30 users per hotspot. It has a system bandwidth of 5 MHz, an ISD of 500 meters and an offered data rate of 100 kbps. The traffic in both scenarios is modeled as full-buffer, where UDP packets are continuously transmitted to users. Table 5.1 provides a summary of the key differences between the two scenarios.

Table 5.1: Comparison of Dense Urban and Massive Connections Scenarios.

Parameter	Dense Urban	Massive Connections
System Bandwidth (MHz)	20	5
System Bandwidth (RBs)	100	25
RBG size	4	1
Uniform Users	30	60
Users per Hotspot	10	30
Inter-Site Distance (m)	200	500
Offered Data Rate	2 Mbps	100 kbps

The activation of hotspots follows a sequential pattern at fixed time intervals. The total simulation duration is 180 seconds, with a new hotspot starting transmission every 30 seconds. During the first interval, no hotspots are active. Table 5.2 presents the number of users per interval for both scenarios. Other relevant parameters are summarized in Table 5.3.

Table 5.2: Number of Users per Interval.

Interval (s)	Dense Urban	Massive Connections
0 - 30	30	60
30 - 60	40	80
60 - 90	50	100
90 - 120	60	120
120 - 150	70	140
150 - 180	80	160

Table 5.3: Simulation parameters common to all scenarios.

Parameter	Value
Number of cells	3
Scheduling algorithm	Proportional Fair
Simulation duration	180 seconds
Channel model	Cost321
Error model	MIESM
UE mobility	No mobility
Traffic model	Non-GBR TCP-based Video

5.1.2 Results for dense urban scenario

This section presents the results for the dense urban scenario. The performance of the proposed multi-agent framework is compared against two single-agent approaches, where only one parameter is dynamically controlled by a single agent. In the first single-agent

case, the agent adjusts the RSRQ threshold, while in the second, the agent controls the bandwidth allocation. These three dynamic strategies are also compared to a static baseline, in which no dynamic solution is applied. The threshold is fixed on 30 and the bandwidth is set to 28 RBs to the private sub-band and 24 RBs to the common sub-band.

Figure 5.2 shows the cell throughput for all users (inside and outside hotspots). This metric is computed by summing all received bytes, grouping them by cell, calculating the throughput for each cell, and then averaging the results across all cells. At each interval, 10 new users start transmitting, due to the activation of a new hotspot. Given the offered data rate of 2 Mbps, the maximum achievable cell throughput increases by approximately 6.67 Mbps per interval. As a result, the maximum achievable cell throughput increases over intervals 1 through 6 as follows: 20, 26.67, 33.33, 40, 46.67, and 53.37 Mbps.

The results indicate that the dynamic appearance of hotspots in different locations significantly degrades cell throughput, while the gap between the actual cell throughput and this maximum achievable grows with each additional hotspot. Figure 5.2 shows that all three dynamic approaches improve performance by adjusting user allocation via the RSRQ threshold and managing bandwidth distribution through the number of RBs assigned to each sub-band.

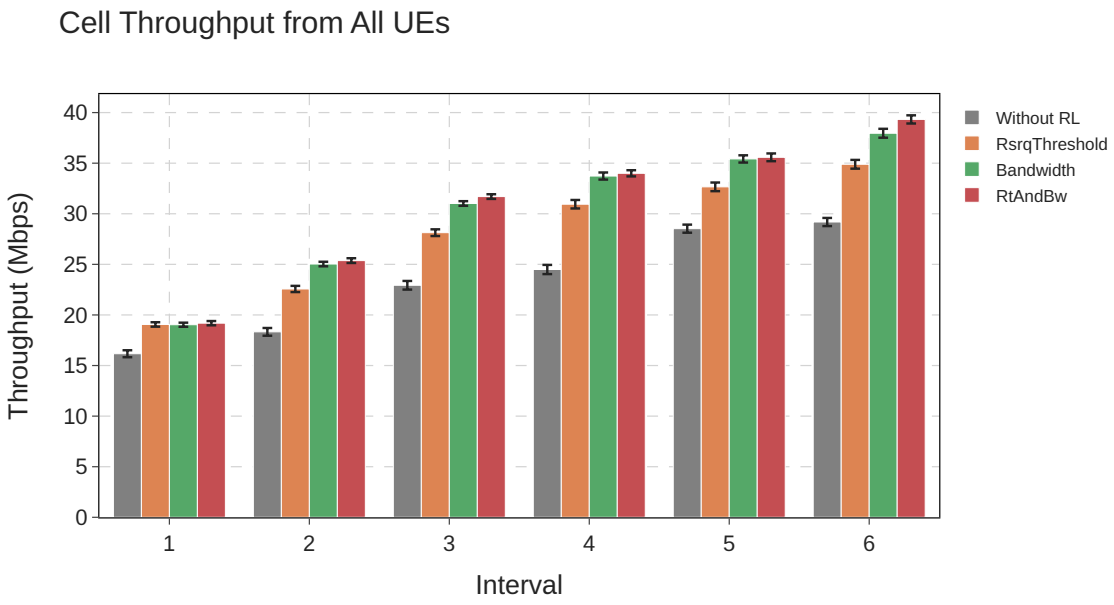


Figure 5.2: Average cell throughput for dense urban scenario.

The proposed multi-agent framework consistently achieves equal or higher overall throughput compared to both single-agent approaches. Among the single-agent configurations, controlling only the RSRQ threshold yields the smallest gain, while dynamically adjusting bandwidth has a greater impact when only one parameter is controlled. The multi-agent approach demonstrates the most significant advantage under high interference, particularly in interval 6, when all hotspots are active. During this

interval, the proposed framework resulted in a 34.7% throughput gain compared to the static configuration, 12.7% compared to the threshold agent, and 3.6% compared to the bandwidth agent.

Figure 5.3 presents the average cell throughput separately for users outside the hotspots (uniform users) and those within the hotspots. The maximum achievable cell throughput for uniform users is 20 Mbps for all intervals, and for hotspot users, over intervals 1 through 6, they are: *0.0*, *6.67*, *13.33*, *20*, *26.67*, and *33.3* Mbps. The results show that both user groups benefit from the dynamic solutions. It is important to note, however, that hotspot users are not necessarily cell-edge users. Depending on the location of the hotspots, these users may be entirely allocated to the cell-center or cell-edge sub-bands or be distributed across both.

The proposed solution provides the highest cell throughput for both uniform and hotspot users. Using interval 6 as an example, when all hotspots are active, the gains are as follows: for uniform users, the threshold, bandwidth and multi-agent approaches provide a 35.11%, 46.38%, and 53.51% gain, respectively. For hotspot users, the gains are 12.13%, 22.23%, and 25.82%, respectively. Namely, the gains are higher for uniform users, since they are highly impacted by the appearance of the hotspots. Compared to the maximum achievable throughput, uniform users experienced up to a 53% reduction in throughput on this interval, while hotspot users experienced a 40.63% throughput loss.

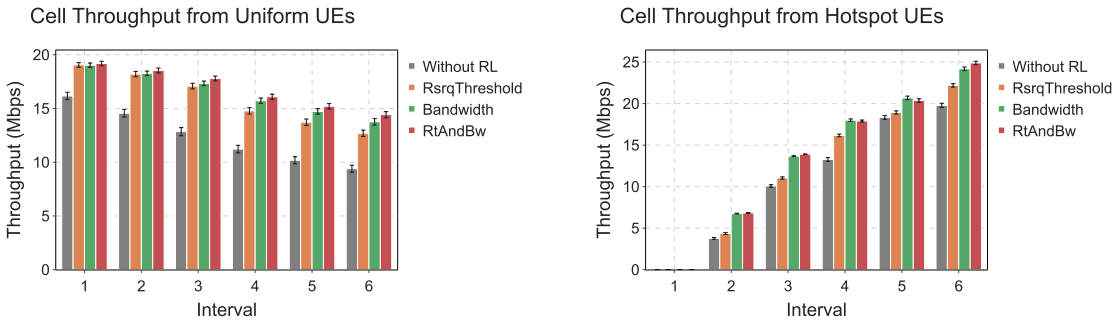


Figure 5.3: Average cell throughput for uniform and hotspot users.

Figure 5.4 presents the empirical cumulative distribution function (ECDF) of user throughput, providing insight into how each solution impacts individual user performance. The ECDF reveals a significant overall improvement in throughput for nearly all users when compared to the baseline system without dynamic control. Furthermore, the multi-agent solution outperforms the single-agent approaches, particularly for users with throughput above 1.2 Mbps and below 0.29 Mbps, indicating better performance for both high and low-throughput users.

Figure 5.5 presents the ECDF of the average user SINR, the other key metric used to evaluate system performance. The results show that the single-agent solution controlling only the RSRQ threshold yields the highest SINR overall. In contrast, both the bandwidth-only and multi-agent approaches result in lower SINR for a portion of the users. However, the multi-agent framework achieves better SINR than the bandwidth-only solution, particularly for users experiencing poor signal quality. Specifically, for users with SINR below 13.2 dB, the multi-agent approach outperforms both the static

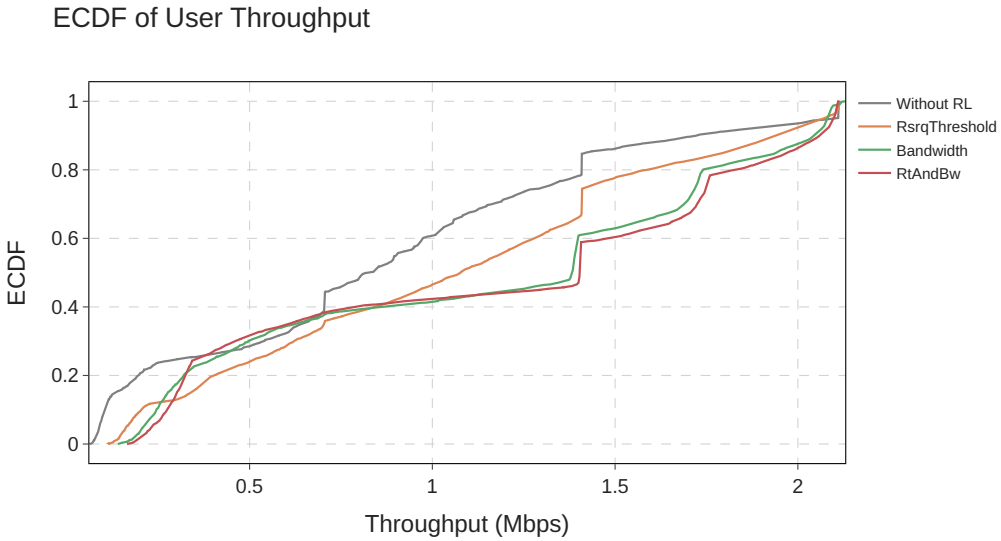


Figure 5.4: ECDF of throughput for dense urban scenario.

configuration (no RL) and the bandwidth-only agent, emphasizing its effectiveness in managing interference for lower performing users.

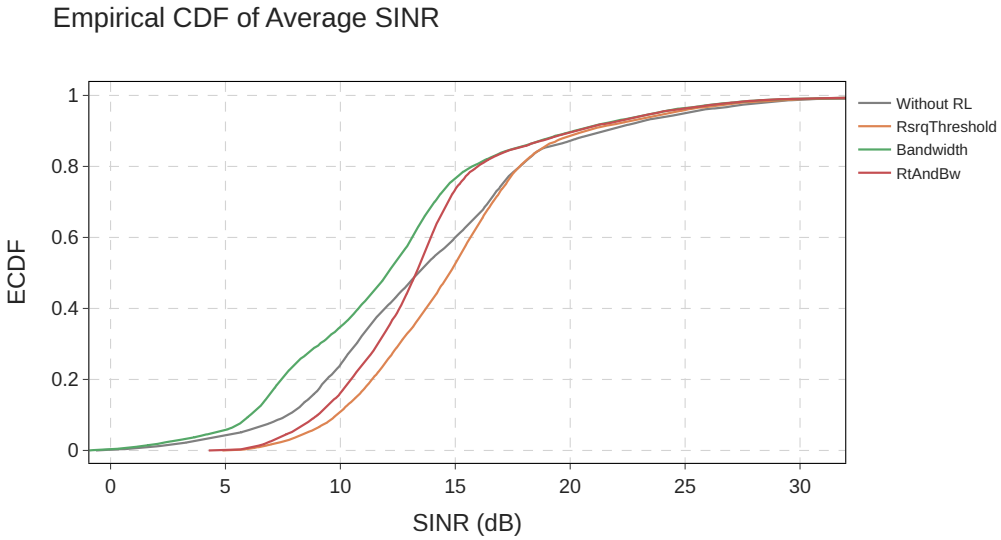


Figure 5.5: ECDF of SINR for dense urban scenario.

To provide further insight into the performance of the QL agents and the proposed framework, the following figures illustrate the frequency of each action taken per interval by each agent, followed by transition matrices. Figure 5.6 presents the bandwidth allocation actions, while Figure 5.7 shows the RSRQ threshold actions for four different

simulation jobs, chosen as representative examples. In all four cases, the proposed framework achieved the highest throughput gain across every interval. These results show evidence that the agents are effectively learning, initially exploring various actions, and eventually exploiting those that yield the highest rewards.

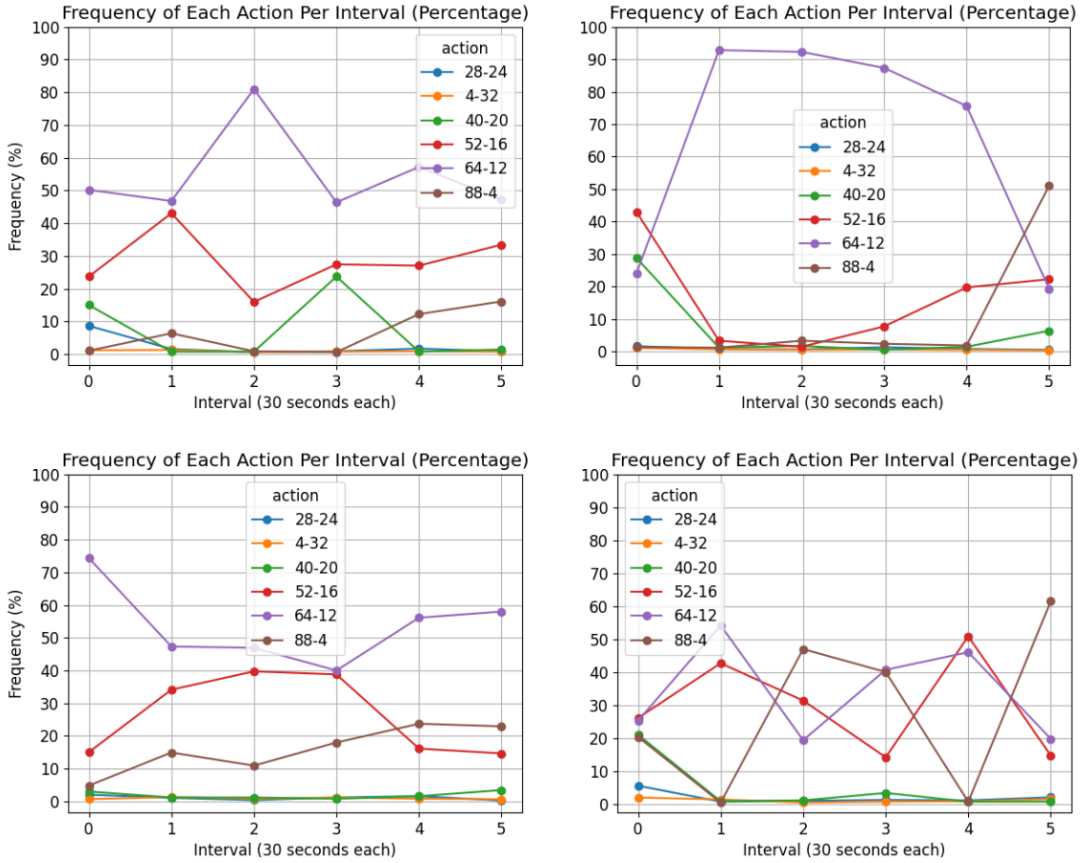


Figure 5.6: Frequency of each bandwidth action from the multi-agent framework.

In some runs, there is a clear preference for a specific action, with one action being significantly more frequent than the others. In other cases, two or more actions appear with similar frequency. This can be attributed to several factors. For instance, multiple actions may yield similar rewards in certain scenarios, leading the agent to alternate between them. Additionally, a high exploration rate, especially in the multi-agent setup, can slow convergence, as both agents independently explore without direct coordination. This can initially result in a high rate of action switching until sufficient exploration allows both agents to align and converge on an effective strategy. However, even with adequate exploration the multi-agent framework can often lead to more action switching. Despite this variability, all jobs and intervals presented in Figures 5.6 and 5.7 achieved higher throughput than the other evaluated solutions.

These figures also highlight that the optimal configurations for bandwidth allocation and user classification may vary across different jobs. Each simulation job randomly

places users in the scenario, generating distinct interference patterns. Nevertheless, because the hotspot locations remain fixed, and they are the dominant factor affecting performance, the best actions remain relatively consistent across runs. For example, in the second interval, the 64-12 bandwidth configuration is optimal for three of the jobs, while the remaining job selects 88-4, reflecting minor adaptation to the specific user distribution. Figure 5.7 also reinforces the need for different configurations across intervals, illustrating how the appearance of hotspots in varying locations demands adaptive parameter control.

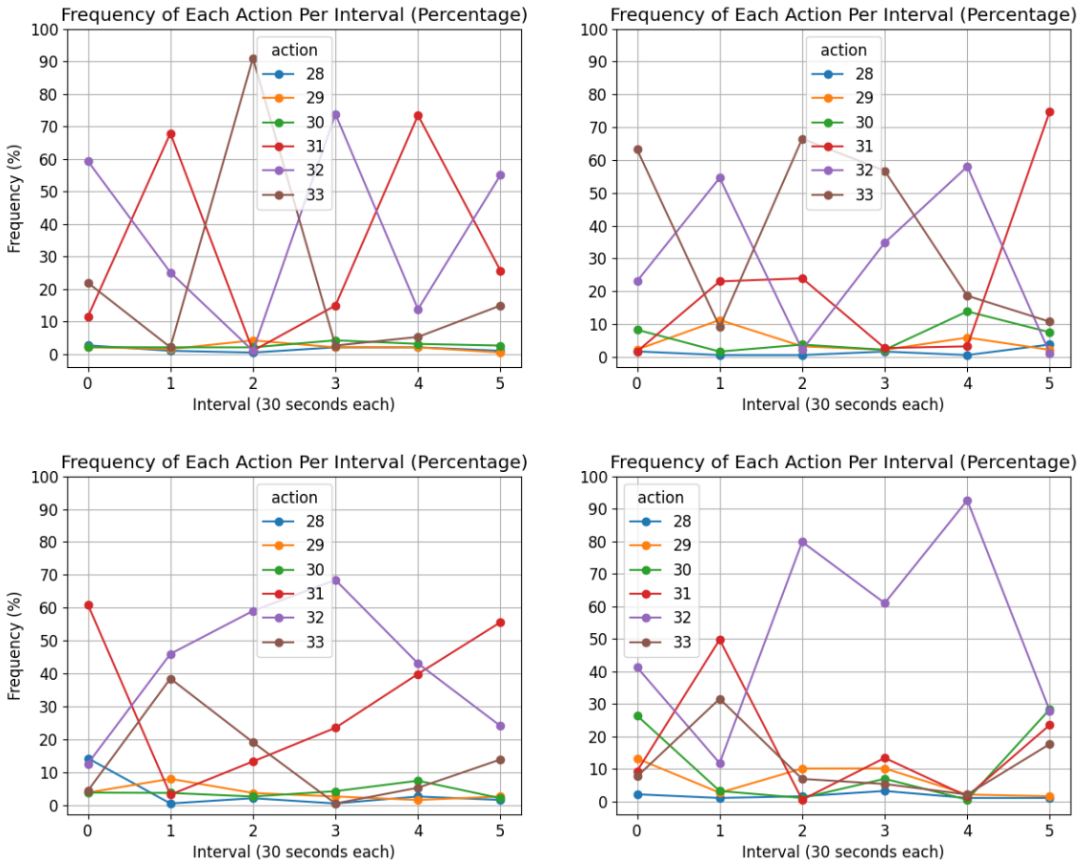


Figure 5.7: Frequency of each threshold action from the multi-agent framework.

Following the analysis of action frequencies, the transition matrices help visualize that even though the agents explore different actions, and sometimes exhibit similar selection frequencies across multiple actions, they still predominantly remain in the same action after sufficient exploration. Figures 5.8 and 5.9 show the action transition matrices for the same simulation jobs presented in the previous figures.

They show how often each agent switches from one action to another and they include the transitions from all intervals. Each element in the matrix represents the number of times the agent transitioned from a given action (row) to another action (column) during the simulation. The elements in the main diagonal indicate a self-transition case, where the agent repeated the same action in consecutive intervals. Elements outside the main

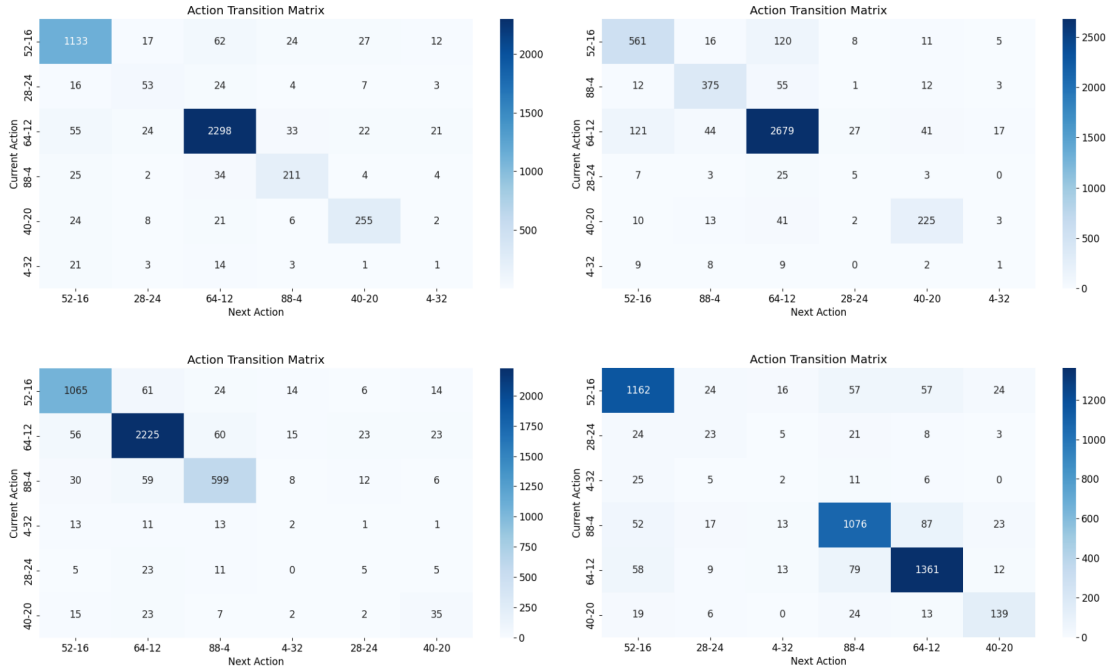


Figure 5.8: Action transition matrix for the bandwidth in the multi-agent framework.

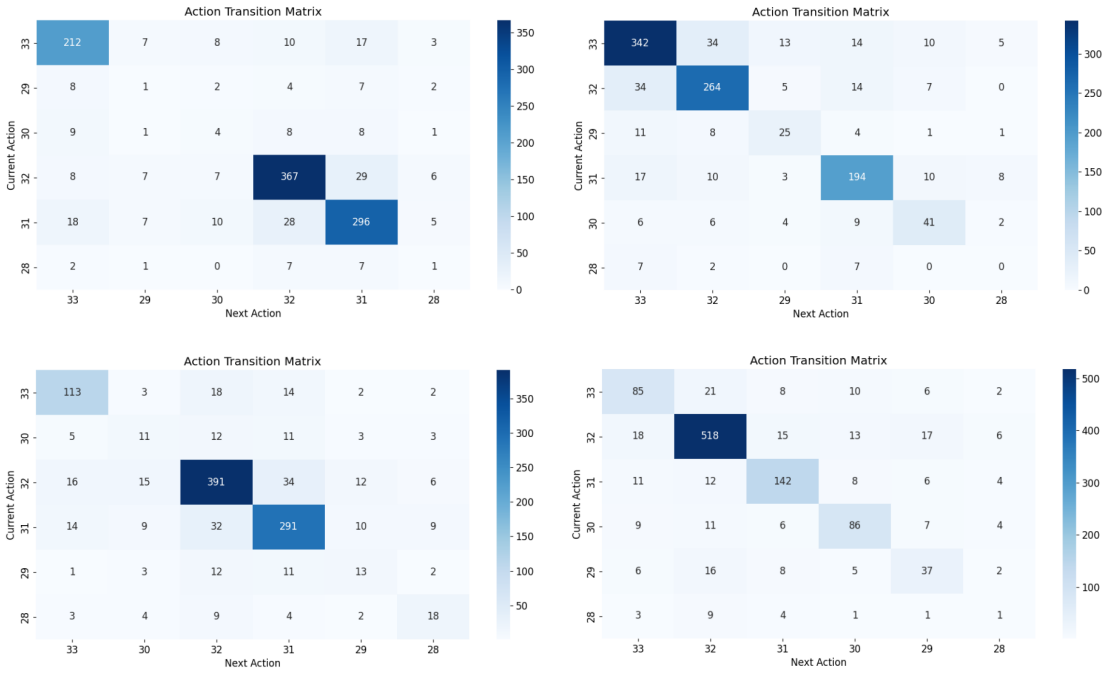


Figure 5.9: Action transition matrix for the RSRQ threshold in the multi-agent framework.

diagonal show the changes between different actions. A high number of self-transitions indicates that the agent has converged to a specific action and the opposite might suggest ongoing exploration or an unstable learning.

Figures 5.8 and 5.9 show that most transitions occur along the main diagonal, confirming that the agents achieved higher throughput by converging to higher performing actions. In the proposed scenario, it is also expected to observe more than one action with a high number of self-transitions, as the optimal action may vary across intervals. Additionally, the matrices reveal a healthy number of transitions between different actions, which is desirable to some extent. It is important that the agent explores as much of the action space as possible to learn the optimal policy.

To complete the analysis, Figure 5.10 presents the actions taken and the corresponding cumulative reward over time for one of the simulation jobs. As described in Section 4.3, the reward signal was defined as the difference between consecutive throughput values. Consequently, the cumulative reward is expected to increase as the agents learn, eventually stabilizing once the agents converge to high-performing actions. Occasional peaks or dips in the cumulative reward are natural and reflect the agents' ongoing exploration.

Figure 5.10 confirms this expected behavior, which repeats at each simulation interval. The activation of an additional hotspot at every 30 seconds increases interference and reduces overall throughput, which in turn leads to a noticeable drop in the cumulative reward. This is typically followed by a rapid recovery, driven by the agents' adaptation to the new network conditions. This adaptation is a result of changes in either the RSRQ threshold or the bandwidth allocation.

Figure 5.10: Cumulative rewards and actions taken for one simulation job of the dense urban scenario.

For example, at 150 seconds, the cumulative reward drops significantly. In response,

the bandwidth agent switches to the 40-20 and the 52-16 configurations, which continues to yield low rewards. Shortly after, it changes to 88-4, leading to a noticeable improvement in performance. Although there are a few temporary switches back to 64-12, the agent eventually stabilizes on 88-4 by the end of the interval.

All results presented thus far demonstrate the performance of the proposed solution in comparison to a fixed configuration, selected based on its good performance in the target scenario. In the case labeled *Without RL*, the system uses a fixed bandwidth configuration of 28 RBs for the common sub-band, 24 RBs for the private sub-band, and an RSRQ threshold of 30. These values were identified through exploratory analysis as well-balanced for the proposed scenario, making it a challenging baseline to surpass in terms of throughput. This also explains why the observed gains were not large, as the scenario was already operating near an optimized configuration. Moreover, it highlights that the fixed parameters relied on prior knowledge of the system, which may not be available in real deployments.

One of the key advantages of the proposed multi-agent framework lies in its ability to consistently enhance performance, even when the initial parameter settings are suboptimal. By learning and selecting the most effective combination of actions, the framework adapts to varying network conditions to maximize user throughput. In contrast, single-agent approaches can be significantly hindered by the misconfiguration of the parameter that is not under their control. Since the optimal values for both the RSRQ threshold and bandwidth allocation can vary considerably depending on traffic conditions and user distribution, single-agent solutions are more vulnerable in the absence of prior knowledge about the environment. The multi-agent approach, by jointly controlling multiple parameters, demonstrates greater robustness and consistently outperforms both single-agent strategies.

This analysis is supported by the following results, which present three additional simulation campaigns. Each highlights a different scenario where fixed configurations deviate from the well-balanced baseline. These scenarios show significant performance drops for the single-agent cases: one where the threshold agent is outperformed, one where the bandwidth agent is outperformed, and one where both are outperformed. In each case, the proposed multi-agent framework continues to adapt and provide higher throughput.

This first simulation campaign evaluates the scenario where the bandwidth is initially configured as 28 RBs for the common sub-band and 24 RBs for the private sub-band, with the RSRQ threshold set to 28. Figures 5.11 and 5.12 present the cell throughput for all users, as well as for uniform and hotspot users separately. Under this configuration, both single-agent approaches are outperformed by the proposed multi-agent framework, particularly in the later intervals where interference is higher. Notably, the static configuration without RL results in significantly lower throughput, with gains of up to 70% observed in the final interval when using the multi-agent framework. Compared to the single-agent approaches, the multi-agent framework achieved throughput gains of up to 11.8%. This performance trend holds for both uniform and hotspot users.

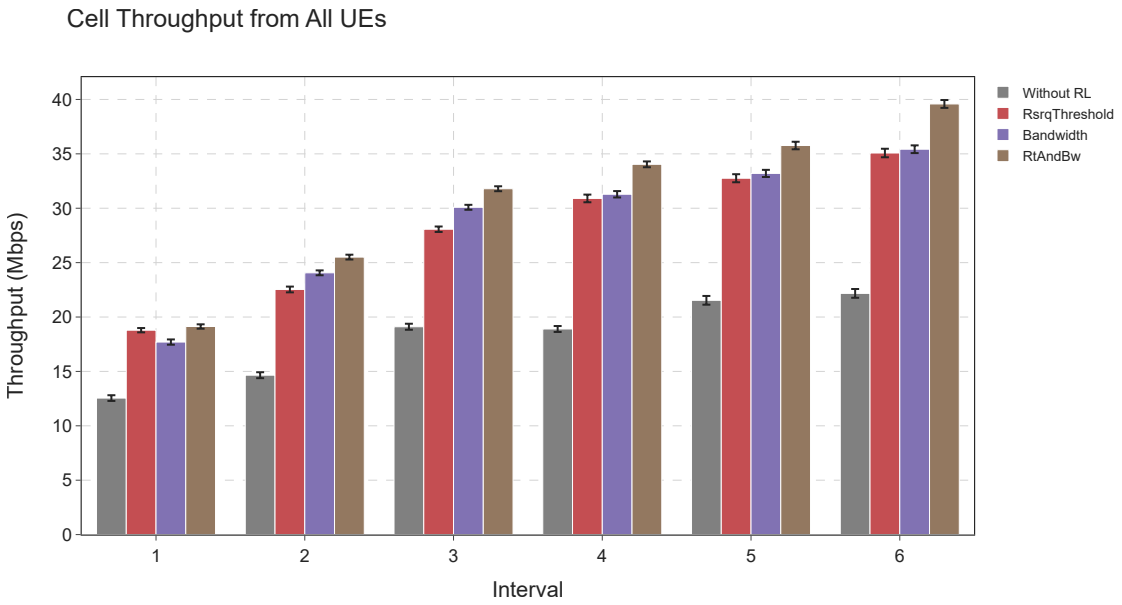


Figure 5.11: Average cell throughput for bandwidth 28-24 and threshold 28.

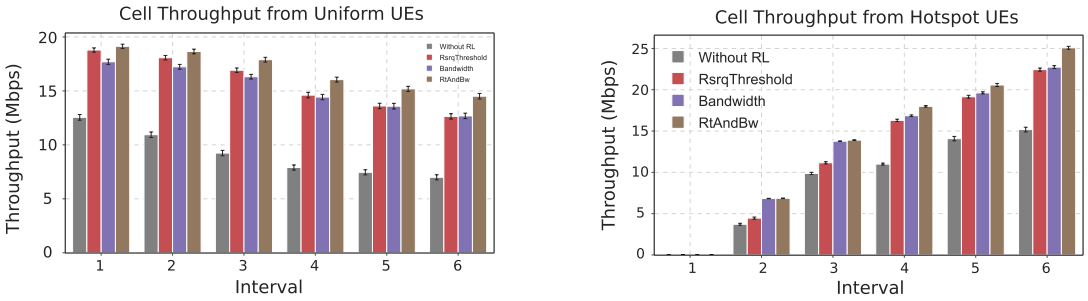


Figure 5.12: Average cell throughput of hotspot and uniform users for bandwidth 28-24 and threshold 28.

Figure 5.13 presents the ECDFs of both SINR and user throughput. In terms of SINR, the threshold agent achieves the highest values for all users, while the bandwidth agent results in the lowest SINR, with the multi-agent approach falling in between. When it comes to user throughput, all dynamic approaches outperform the static configuration without RL. Among them, the multi-agent solution provides the best overall performance for the majority of users, including those with the lowest throughput.

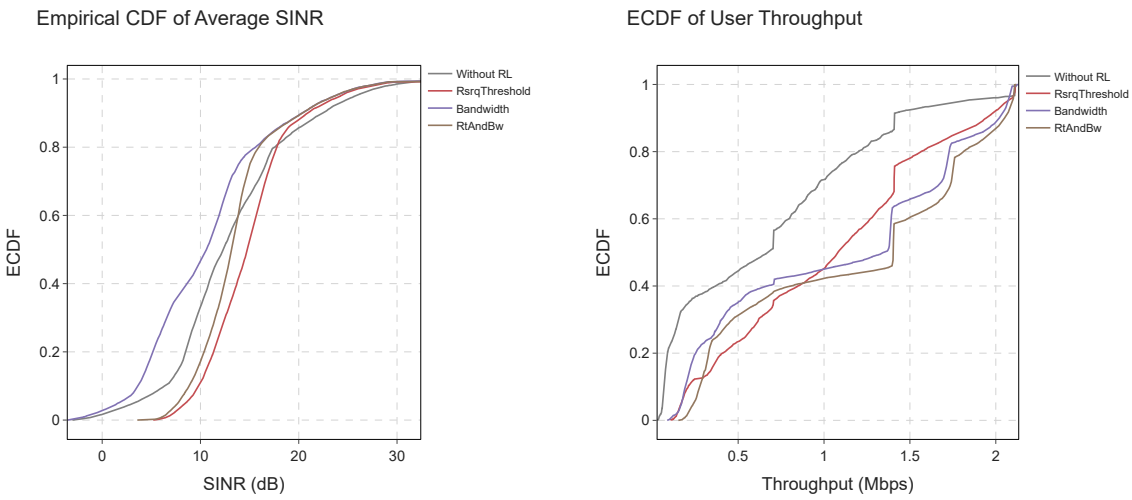


Figure 5.13: ECDF of SINR for bandwidth 28-24 and threshold 28.

The second simulation campaign considers a scenario in which the bandwidth agent is particularly outperformed. This simulation initially configures the bandwidth as 88 RBs for the common sub-band and 4 RBs for the private sub-band, while the RSRQ threshold is set to 28. As shown in Figure 5.14, the bandwidth agent fails to improve throughput relative to the static configuration and is also outperformed by the threshold agent. The proposed framework achieves the best performance, with a maximum throughput gain of 11.2% compared to the bandwidth agent, with a similar behavior on uniform and hotspot users, illustrated in Figure 5.15.

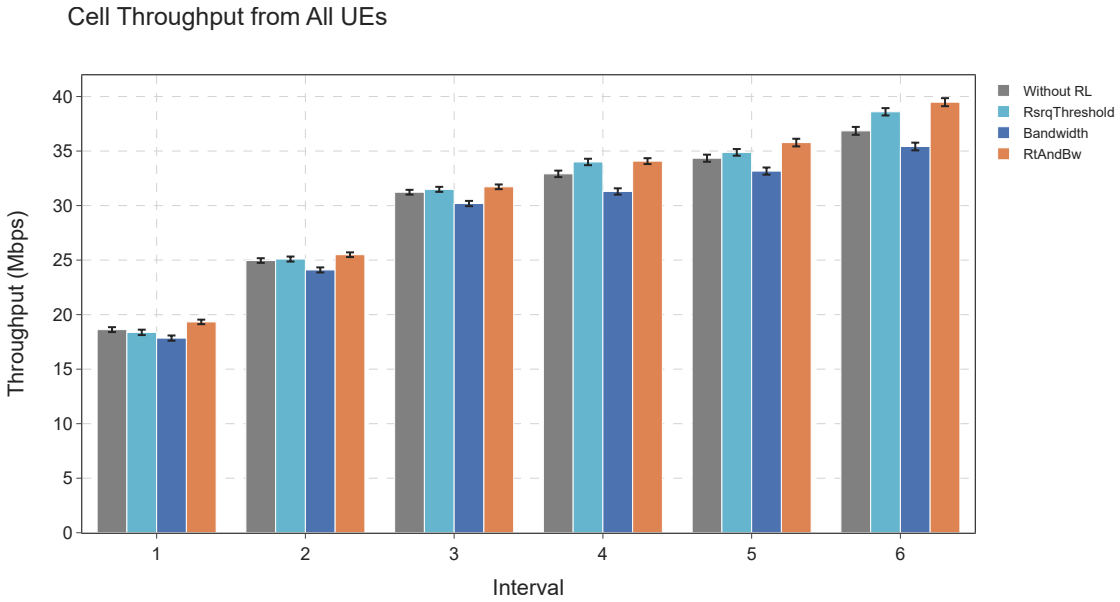


Figure 5.14: Average cell throughput for bandwidth 88-4 and threshold 28.

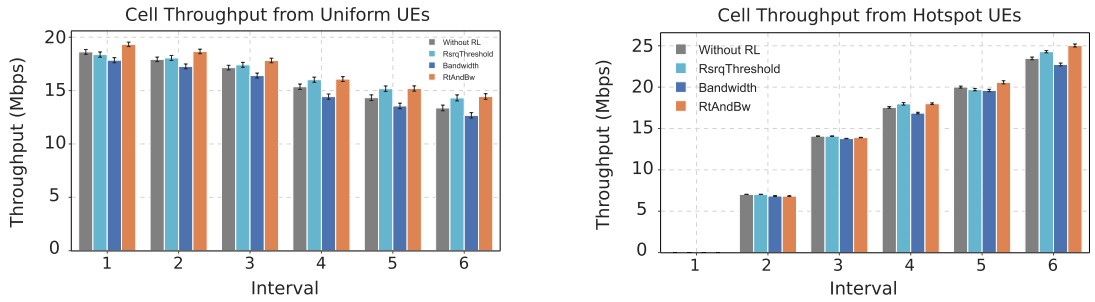


Figure 5.15: Average cell throughput of hotspot and uniform users for bandwidth 88-4 and threshold 28.

The ECDF of SINR, presented in Figure 5.16, shows that the multi-agent framework achieves higher SINR values compared to all other approaches, particularly for low-performing users. The threshold agent performs better for users with SINR below 10 dB, while the bandwidth agent provides higher SINR for users above this threshold. The ECDF of user throughput further confirms the effectiveness of the multi-agent solution: it delivers superior performance for users with lower throughput and consistently outperforms the bandwidth agent across the entire user base.

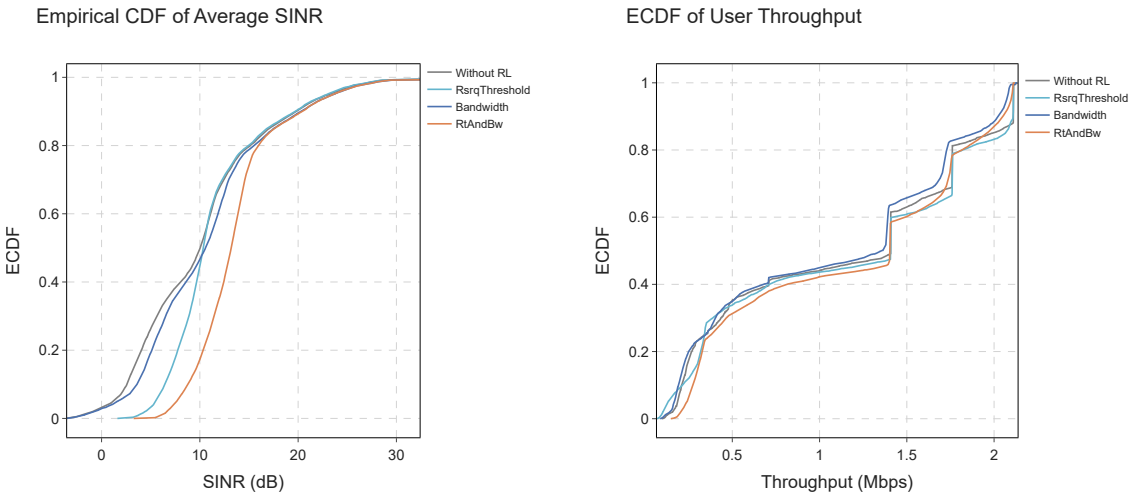


Figure 5.16: ECDF of SINR for bandwidth 88-4 and threshold 28.

The last simulation examines a scenario in which the threshold agent is greatly outperformed by the proposed solution. The initial configuration allocates 4 RBs to the common sub-band and 32 RBs to the private sub-band, with the RSRQ threshold set to 30. For this configuration, the bandwidth agent has similar performance compared to the proposed solution, but the threshold agent, while still better then the static configuration, is greatly outperformed by the other approaches. Figure 5.17 shows that the multi-agent solution achieves up to 99.4% throughput gain compared to the threshold agent, and up to 3.5% gain compared to the bandwidth agent. As in previous scenarios, this behavior

is also observed for uniform and hotspot users, confirmed by the throughput presented in Figure 5.18.

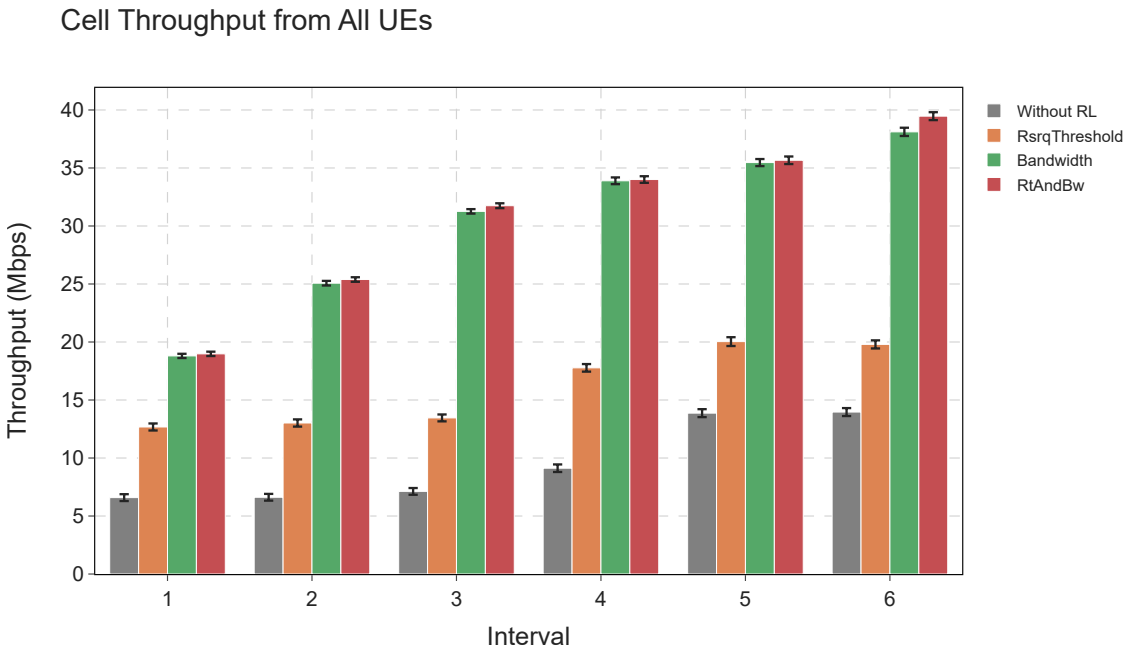


Figure 5.17: Average cell throughput for bandwidth 4-32 and threshold 30.

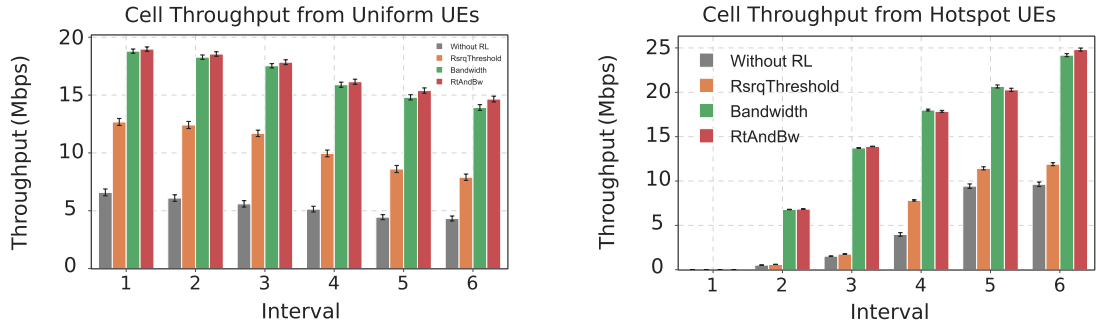


Figure 5.18: Average cell throughput of hotspot and uniform users for bandwidth 4-32 and threshold 30.

Lastly, Figure 5.19 presents the ECDFs of both SINR and user throughput for this configuration. In terms of SINR, the threshold agent provides the highest values for most users, outperforming both the bandwidth agent and the multi-agent framework. The bandwidth agent consistently yields the lowest SINR values, while the multi-agent solution once again falls between the two. Notably, the multi-agent framework delivers significantly better SINR for the lowest-performing users when compared to the bandwidth agent, although it does not reach the SINR levels achieved by the threshold

agent or static configuration for users with already high signal quality. Regarding user throughput, the ECDF reveals that both the bandwidth agent and the multi-agent framework significantly outperform the threshold agent for all users, and the latter provides the best overall performance, providing higher throughput than the bandwidth agent for most users.

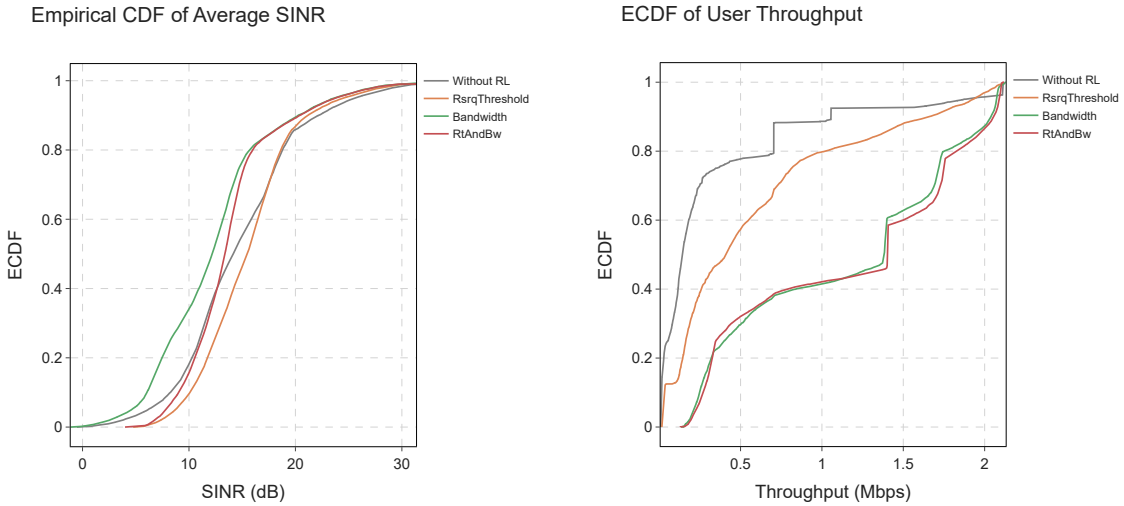


Figure 5.19: ECDF of SINR for bandwidth 4-32 and threshold 30.

All simulation campaigns discussed in this section, from the well-configured baseline scenario to the additional cases with suboptimal parameter settings, demonstrate the robustness and adaptability of the proposed hierarchical multi-agent framework. The solution is capable of not only providing moderate throughput gains in scenarios that were already carefully tuned using prior system knowledge, but also of significantly improving performance in less favorable conditions. In cases where the initial configuration was far from optimal, the multi-agent framework successfully enhanced system performance and consistently outperformed both single-agent approaches.

These results highlight a key strength of the hierarchical multi-agent approach: its ability to jointly adapt multiple control parameters to mitigate interference in a highly dynamic network, where a new hotspot becomes active every 30 seconds. Under the Strict FR scheme, there is a strong interdependence between the User Classification (using the RSRQ threshold) and bandwidth allocation, and their combined configuration has a strong impact on system performance. In this context, while single-agent solutions are limited by the static value of the parameter they do not control, the multi-agent framework can dynamically explore the joint action space and converge toward the most effective configuration, even under high interference and changing user distribution.

Moreover, the detailed analyses of cumulative rewards, action frequencies, and transition matrices confirm that the agents are able to learn and stabilize over time, balancing exploration and exploitation. The framework also proves to be effective for both uniform and hotspot users. By analyzing the ECDF of SINR and throughput, the proposed solution was also shown to consistently enhance the performance of users with

lowest throughput, which is an important aspect in interference mitigation, since these users are more vulnerable to ICI.

5.1.3 Results for massive connections scenario

The proposed multi-agent hierarchical framework was also evaluated in the massive connections scenario, whose parameters are summarized in Table 5.1. Compared to the dense urban scenario, this configuration includes a higher number of users, reduced bandwidth, a lower ISD, and a lower offered data rate. Figure 5.20 presents the average cell throughput for all users and Figure 5.21 shows the throughput separately for uniform and hotspot users.

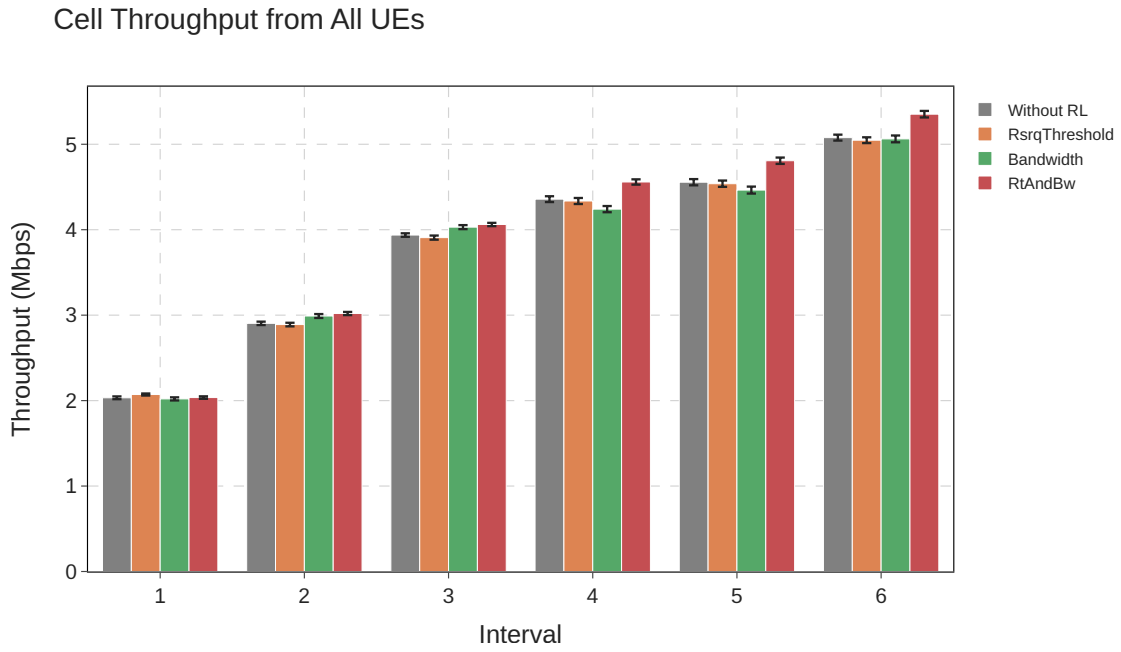


Figure 5.20: Average cell throughput for massive connections scenario.

Similar to the previous scenario, the proposed framework achieves higher throughput in intervals with increased interference. The bandwidth agent is slightly outperformed by the threshold agent in the later intervals, while the opposite is observed in intervals 2 and 3. For uniform users, the threshold agent achieves the best performance in the early intervals, but is eventually surpassed by the multi-agent framework in later intervals. For hotspot users, the threshold agent is initially outperformed by the bandwidth agent, and the proposed solution delivers the best performance under high interference conditions.

In intervals with fewer active hotspots (1, 2 and 3), the system is able to reach the offered data rate even without dynamic control. It is important to recognize that reinforcement learning-based solutions may occasionally result in lower throughput compared to a well-optimized static configuration. This is due to the agents' exploration

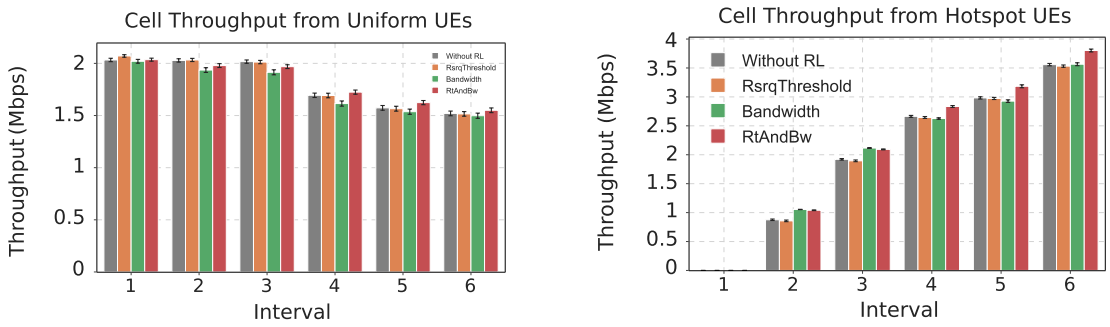


Figure 5.21: Average cell throughput of hotspot and uniform users for massive connections scenario.

behavior, which can lead to frequent action switching and, in some cases, temporary performance degradation. The impact of this behavior can vary depending on which parameters are being controlled. For example, if an agent changes the bandwidth allocation from 88-4 to 4-32 as part of its random exploration strategy, it can cause a substantial change in throughput, since the number of RBs allocated to each sub-band is drastically different.

These results reinforce the relevance of the proposed framework, as it demonstrates the ability to maintain high performance even in cases where the static configuration was already close to optimal, while still achieving gains in more challenging scenarios. This also highlights the importance of correctly choosing the exploration rate of the algorithm. Nevertheless, the overall improvements in this scenario were not particularly large, which may indicate that the system is approaching its capacity limits. This hypothesis is left for further investigation in future works.

Nevertheless, the proposed solution still enhanced cell throughput by up to 7% in the best case, while also enhancing the performance of the lowest-performing users, as shown by the ECDF of user throughput in Figure 5.22. Additionally, the multi-agent framework demonstrated the most balanced performance across the entire user base. According to the ECDF, the threshold agent performs worse for users with throughput above 550 kbps, while the bandwidth agent underperforms for users below 400 kbps. The multi-agent solution provides higher throughput than both single-agent approaches on their respective weak zones. When the bandwidth agent performs poorly, the multi-agent framework aligns with the performance of the threshold agent, and vice versa, ensuring consistently better or at least comparable performance across all user groups.

Figure 5.22 also presents the ECDF of user SINR. Both the bandwidth and threshold agents resulted in higher SINR values compared to the multi-agent approach, but not higher than the baseline without dynamic control. As discussed in previous sections, improvements in SINR do not always translate into throughput gains, particularly when using the Strict FR scheme. This is because Strict FR prioritizes interference mitigation by allocating users to disjoint sub-bands, which reduces interference but also limits the available bandwidth, potentially constraining overall throughput.

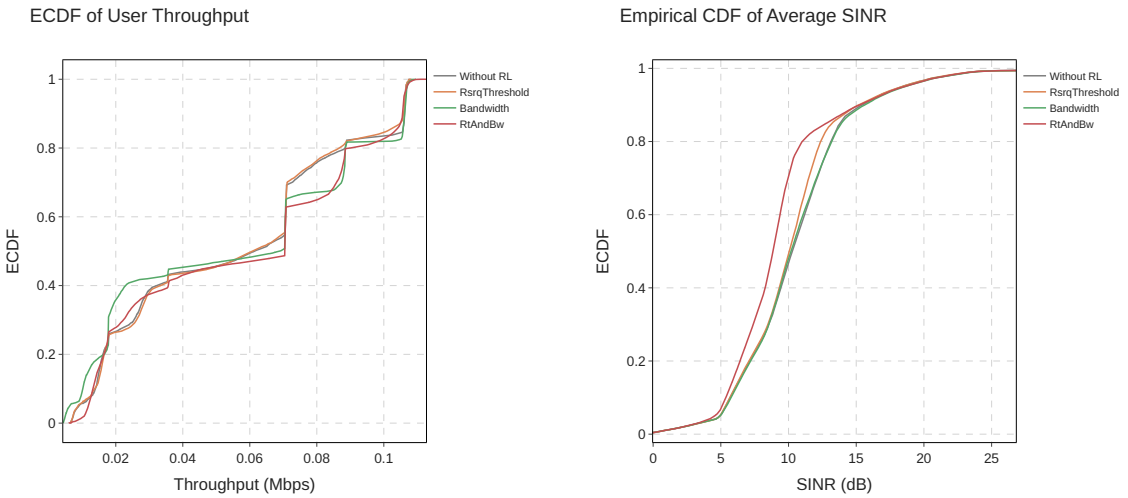


Figure 5.22: ECDF of throughput and SINR for massive connections scenario.

Lastly, the action frequency plot and transition matrices are presented in Figures 5.23 and 5.24. Since Section 5.1.2 already provides a detailed discussion on the agents' behavior and learning process, two simulation jobs are presented to offer additional insight specific to the massive connections scenario.

As observed in the dense urban scenario, the combined information from these two figures provides strong evidence that the improved performance of the proposed framework stems from the agents' ability to learn which combinations of actions maximize future rewards. Figure 5.23 shows that, in most cases, the agent prioritizes one action over the others within each interval. Furthermore, the transition between intervals often leads the agent to favor a different action, reflecting its adaptation to changing traffic conditions triggered by the activation of new hotspots.

Different simulation jobs may also require different configurations of the parameters, as users are deployed in different locations across runs, even though hotspots users remain confined within the predefined hotspots areas. For example, in interval 2, the most frequently selected bandwidth allocation in the first job (shown on the left) is 12-4, while the second job (on the right) is 6-6. These differences in bandwidth selection influenced the choice of RSRQ threshold values, which vary accordingly between the two jobs.

Additionally, these results indicate that the massive connections scenario caused the agents to switch between actions more frequently, despite using the same agent configuration as in the dense urban scenario. The exploratory analysis presented in Section 3.4 showed that the massive connections scenario maintained good performance for a wider range of threshold values, although drastic changes on its value still led to significant performance degradation, especially in deployments with higher interference. In contrast, bandwidth allocation had a bigger impact on cell throughput, likely due to the more limited bandwidth available.

Combined with the smaller ISD and lowered offered data rate, these factors create a larger set of bandwidth and threshold combinations capable of delivering good performance. Therefore, the agents are expected to switch actions more frequently, as

multiple parameter configurations may provide similar cell throughput. Nevertheless, the agents still jointly improved the overall throughput, which is also the case for both individual jobs represented in Figures 5.23 and 5.24.

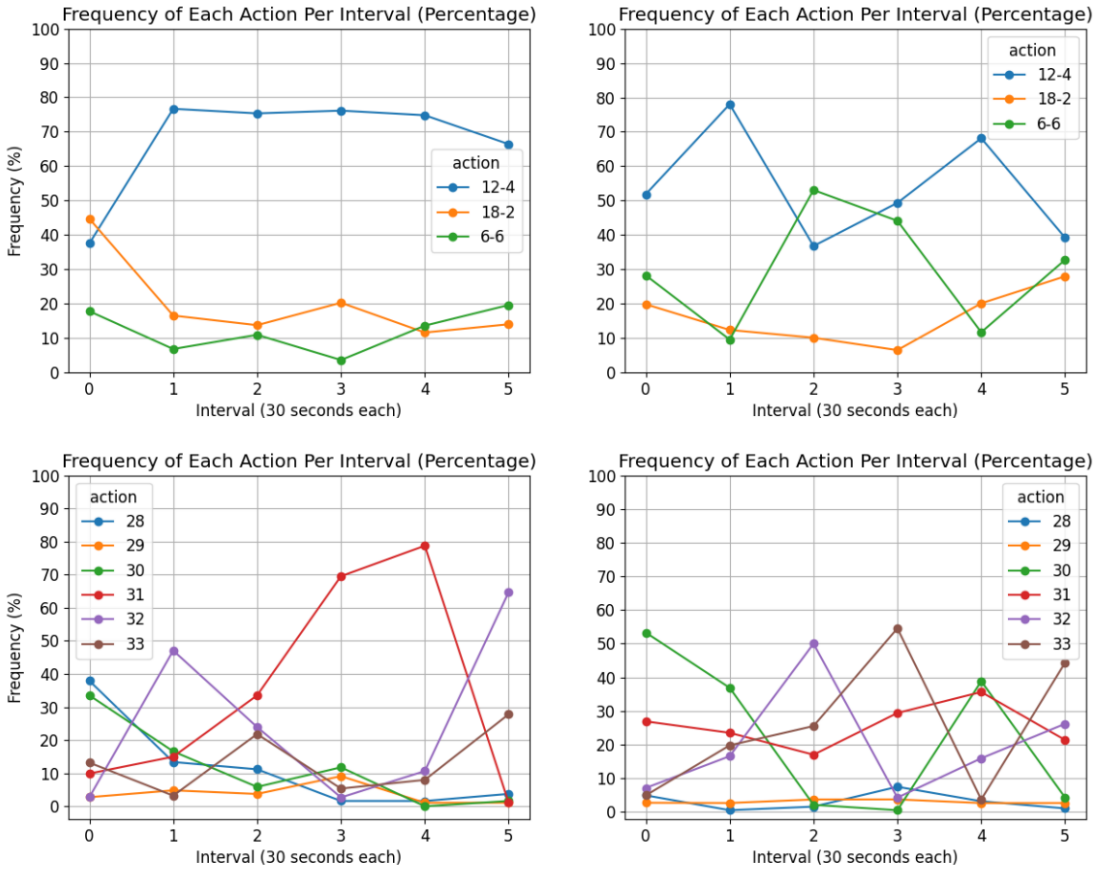


Figure 5.23: Frequency of each action from the multi-agent framework for the massive connections scenario.

Lastly, the action transition matrices shown in Figure 5.24 confirm that, although the agents switched between actions more frequently than in the previous scenario, they still converged toward high-performing actions, as indicated by the high number of self-transitions. Consequently, the matrices also reveal a higher exploration of values other than the most frequent action. When comparing the action frequency per interval with the transition matrices, it becomes clear that the most frequently selected actions in each interval correspond to those with the highest number of self-transitions, particularly for actions that remain dominant across multiple intervals.

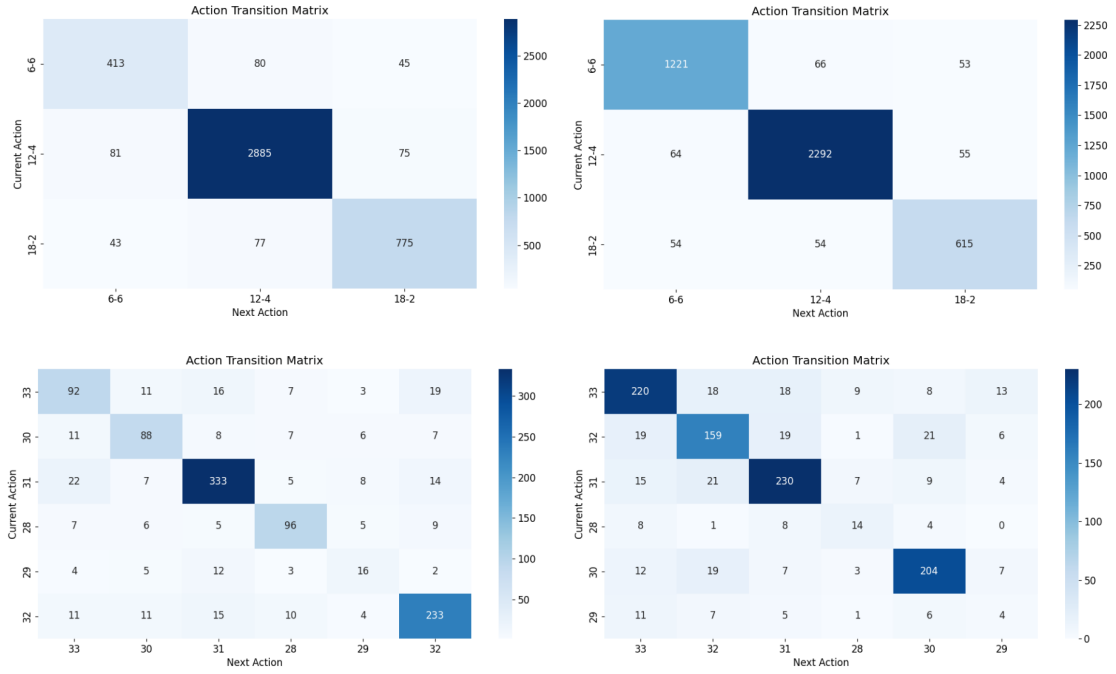


Figure 5.24: Action transition matrix for the bandwidth in the multi-agent framework for the massive connections scenario.

5.1.4 Evaluating the reward signal

The importance of the reward signal was discussed in Section 4.2 and its design was discussed in Section 4.3. The following simulation results show the impact of choosing throughput, SINR, or both as the reward signal. Figure 5.25 presents the cell throughput for all users per interval. The hierarchical multi-agent framework enhances cell throughput when the reward is based on user throughput alone or a combination of SINR and user throughput. As highlighted in Section 4.3, relying solely on SINR as the reward signal can degrade throughput, particularly in FFR scenarios with high interference. Since RL agents optimize for the objective dictated by the reward function, using only SINR encourages them to maximize SINR without directly considering throughput. When using Strict FR, neighboring CEUs are allocated to disjoint sub-bands. Hence, agents can improve SINR by minimizing bandwidth reuse. While this strategy increases SINR, it reduces overall throughput due to the restricted use of available spectrum.

This phenomenon is evident in Figure 5.25, where using SINR as the sole reward metric results in even lower throughput than not applying the dynamic solution. Additionally, using only throughput as the reward leads to the best result, since SINR is not taken into account. This behavior is confirmed by the actions taken by the RL agents. Figure 5.26 shows the frequency of each action per interval for 4 different simulation jobs when using only the SINR. The agent that controls the bandwidth allocation converged

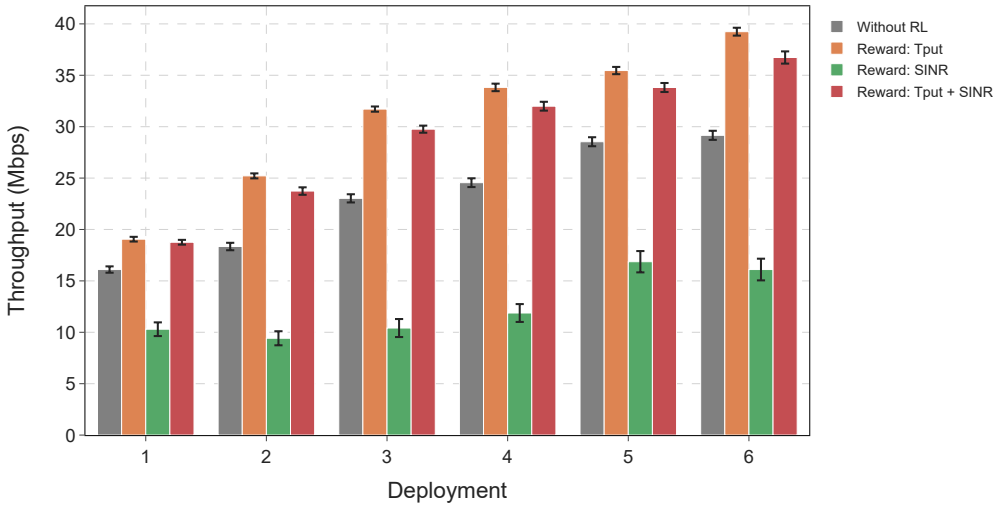


Figure 5.25: Average cell throughput for different reward signals.

to the bandwidth with the highest number of RBs to the private sub-band, increasing the available disjoint bandwidth, potentially increasing the SINR.

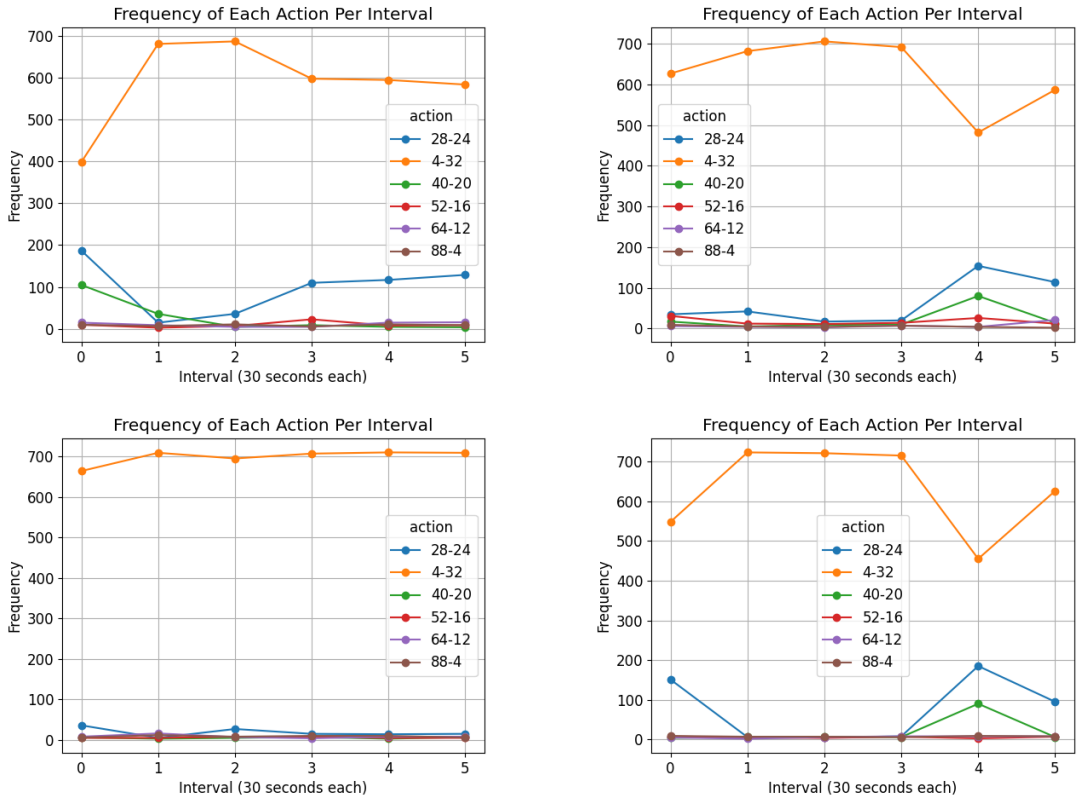


Figure 5.26: Frequency of each action taken in four different simulation jobs when using only SINR as the reward.

Figure 5.27 shows the average SINR per interval and Figure 5.28 shows the ECDF of the SINR. As expected, using SINR as the reward yields the highest average SINR for all intervals, while using only throughput yields the lowest SINR, and using both falls in between. However, the ECDF provides a better representation of the system. While using only the SINR is better for the users with highest SINR, the ECDF shows that using both metrics is better for the users with lowest SINR. A similar trend is observed for throughput, as illustrated by the ECDF of the throughput in Figure 5.28, although the difference is not substantial. Figure 5.28 also shows that using only the SINR significantly decreases throughput for all users, even when comparing with the no RL option.

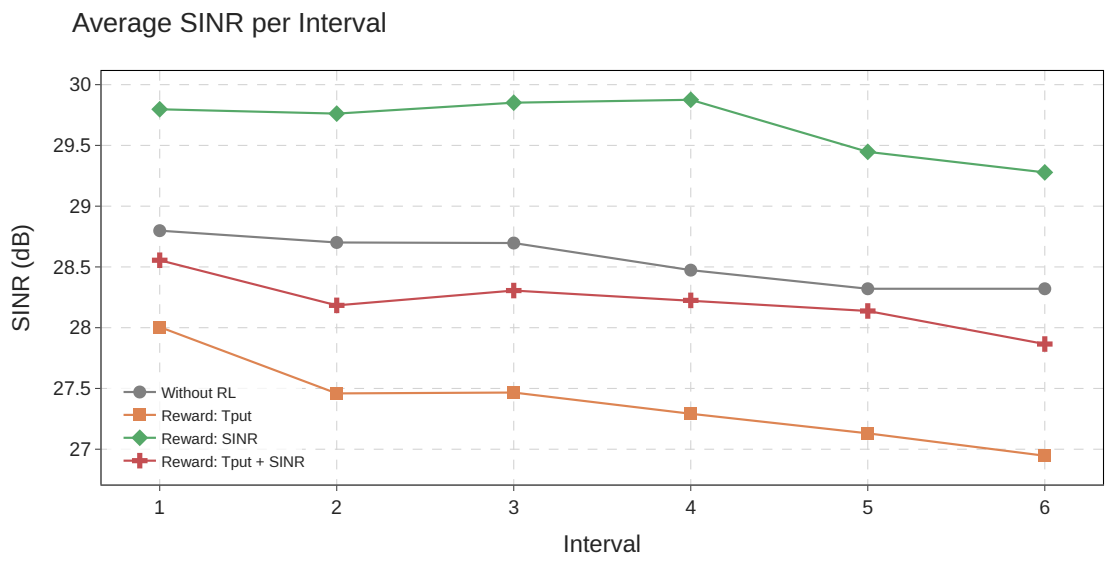


Figure 5.27: Average SINR per interval for different reward signals.

According to these results, both approaches that include throughput on the reward are viable, but for this specific scenario, there is no significant gain in including the SINR in the reward. It is important to highlight that the combined metric, which follows the formulation presented in Equation (4.13), assigns a much higher weight to the throughput. Moreover, in a scenario where the traffic is not modeled as a full-buffer, throughput might not be feasible, as the reward would be directly affected by periods where users are not transmitting, misleading the agent's actions.

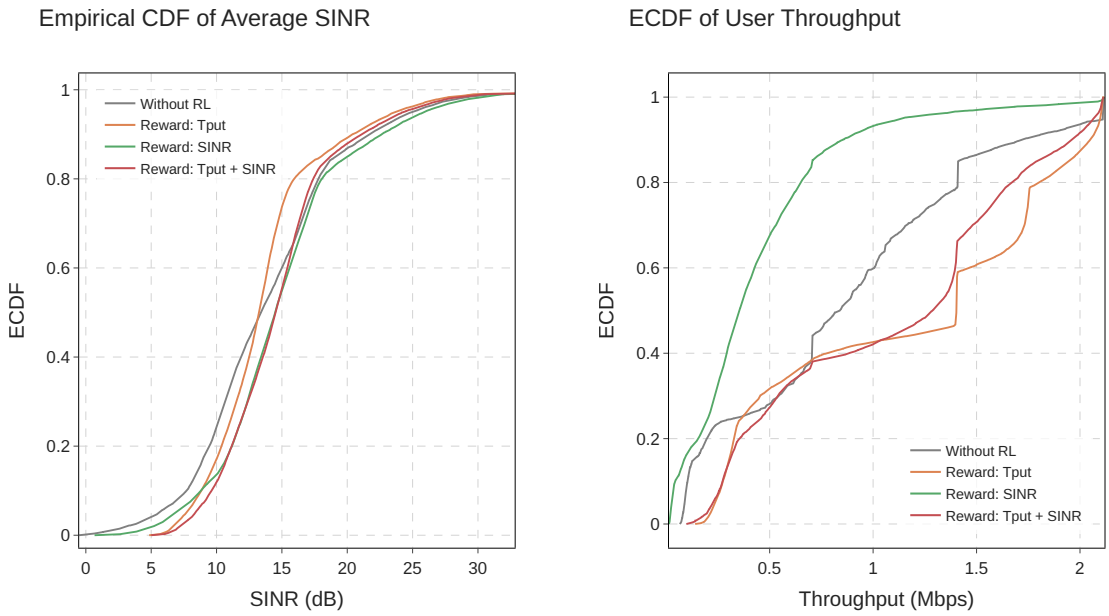


Figure 5.28: ECDF of throughput and SINR for different reward signals.

5.2 Hierarchical MAB-ICIC solution

The proposed solution in this work consists of a hierarchical framework that enables the indirect coordination of two RL agents. Throughout the previous sections, the main set of results was obtained using QL agents to solve the RL problem. However, the framework is flexible and allows for the integration of different types of RL agents, as long as they operate through direct interaction with the system, use a reward signal, and aim to estimate the rewards. Depending on the specific constraints or goals of a given scenario, alternative learning strategies may provide advantages, such as faster convergence, lower complexity, or better adaptability to environments with highly complex state representation.

To explore this possibility, the final part of this study introduces the same hierarchical framework, now implemented with Multi-Armed Bandit (MAB) agents, primarily with the goal of reducing computational complexity.

The Multi-Armed Bandit problem considers a simplified reinforcement learning formulation in which agents do not learn the best action for each state. In this *nonassociative* setting, much of the complexity of the full reinforcement learning is avoided (Richard S. Sutton 2018). The name Multi-Armed Bandit was inspired in slot machines, where each machine represents an arm that returns a reward when pulled. In the MAB problem, the machine has n levers. In other words, the agent can choose from n available actions, independently of the system's state. Through repeated selections, the agent tries to maximize the cumulative reward over time.

Instead of mapping values to state-action pairs as in Q-Learning, MAB agents maintain estimates of the expected reward for each action, denoted as $Q(a)$. Since there is no state dependency, these estimates are solely functions of the actions themselves.

A key difference from Q-Learning is the method used to estimate these action values. Considering that each action value has an expected or mean reward, instead of learning an optimal policy, the agent estimates the reward using the *sample-average* method, which is a simple average of the sample of relevant rewards (Richard S. Sutton 2018). The estimated expected reward for each action, its q value, is the average of the received rewards. Equation (5.1) presents the formal definition of $Q(a)$.

$$Q_t(a) = \frac{R_1 + R_2 + \cdots + R_{N_t(a)}}{N_t(a)}, \tag{5.1}$$

where $Q_t(a)$ is the estimated q value for action a in time step t, $R_{N_t(a)}$ is the Nth reward in time step t, and $N_t(a)$ is the number of times action a has been selected at time step t. As $N_t(a) \to \infty$, $Q_t(a)$ converges to the true q value of action a, according to the law of large numbers (Richard S. Sutton 2018). Similar to the Q-Learning, to avoid storing the entire history of received rewards, the average of the kth reward, Q_{k+1}, can be computed incrementally, as described in Equation (5.2).

$$\begin{aligned} Q_{k+1} &= \frac{1}{k}\sum_{i=1}^{k} R_i \\ &= \frac{1}{k}\left(R_k + \sum_{i=1}^{k-1} R_i\right) \\ &= \frac{1}{k}\left(R_k + (k-1)Q_k + Q_k - Q_k\right) \\ &= \frac{1}{k}\left(R_k + kQ_k - Q_k\right) \\ &= Q_k + \frac{1}{k}\left[R_k - Q_k\right]. \end{aligned} \tag{5.2}$$

The term $\frac{1}{k}$ can be considered as a step into the new estimate, similar to the learning rate in the Q-Learning computation. In this case, the learning rate decays with each time step, as more rewards are collected. The final formulation showed in Equation (5.2) was used to implement the learning mechanism of the MAB agents in this work. As a result, the agent only needs to keep track of the current reward, the current q value, and the number of times each action has been selected, greatly simplifying memory and computational requirements.

Besides these differences, the MAB agents were configured to be as similar as possible to the QL agents. They have the same set of actions, receive the same reward signal, and operate on the same time window as the QL agents. They also choose the actions based on an $\varepsilon - greedy$ policy and ε decays over time. One agent controls the bandwidth allocation and another agent controls the RSRQ threshold, as described in Chapter 4, operating under the same hierarchical logic. Additionally, these agents were also implemented as an additional class in the LTE model, available from ns-3's attribute system.

The evaluation scenario follows the setup described in Section 5.1.1, and the results presented in this section correspond to the dense urban configuration. Figure 5.29

shows the average cell throughput for the proposed multi-agent hierarchical framework, comparing the performance of Q-Learning agents and MAB agents. Both are also evaluated against the static baseline configuration, which uses fixed values of 28 RBs for the common sub-band, 24 RBs for the private sub-band, and an RSRQ threshold of 30. This baseline corresponds to the first scenario introduced in this chapter and represents the most balanced initial configuration.

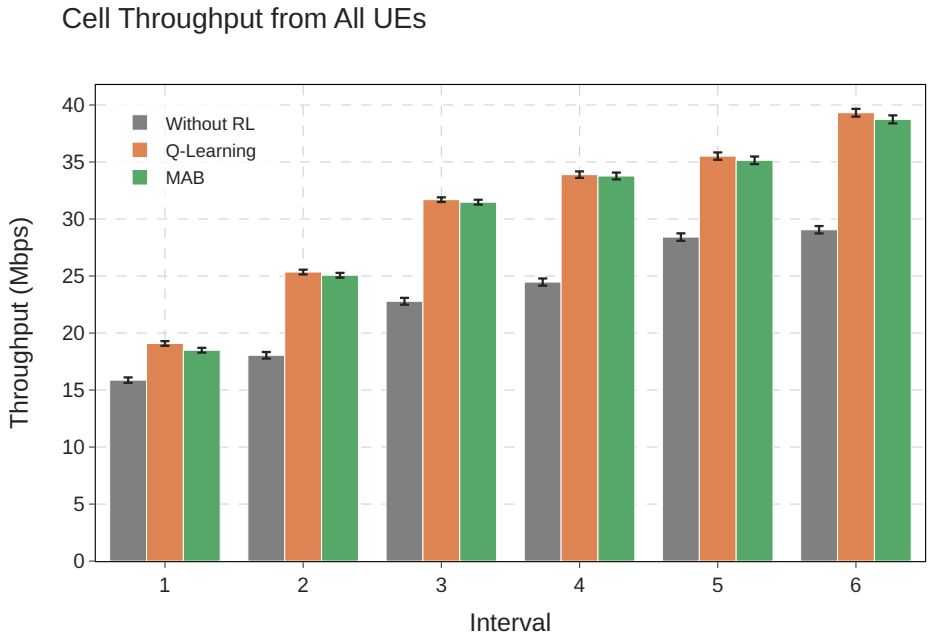


Figure 5.29: Average cell throughput comparison between Q-Learning and MAB.

The use of MAB agents within the hierarchical framework results in throughput gains over the static baseline across all intervals and user groups. Furthermore, the MAB-based solution achieves performance levels comparable to those of the QL-based implementation, although it is slightly outperformed in all intervals. The largest performance gap occurs in the final interval, which also corresponds to the period of highest interference and overall network load. This trend is consistent across both uniform and hotspot users, as illustrated in Figure 5.30.

Figure 5.31 presents the ECDFs of user throughput and SINR for both solutions. In terms of SINR, the MAB-based approach provides the lowest SINR for all users, particularly for those with lower SINR values. With regards to throughput, the Q-Learning approach shows the most significant advantage for the lowest-performing users. Nevertheless, when considering the overall average cell throughput and the ECDF distribution, the MAB solution demonstrates performance that is largely comparable to that of Q-Learning, with only minor differences across most users.

Additionally, the MAB approach requires no knowledge of the system state, which reduces implementation complexity. Unlike Q-Learning, which relies on well-defined

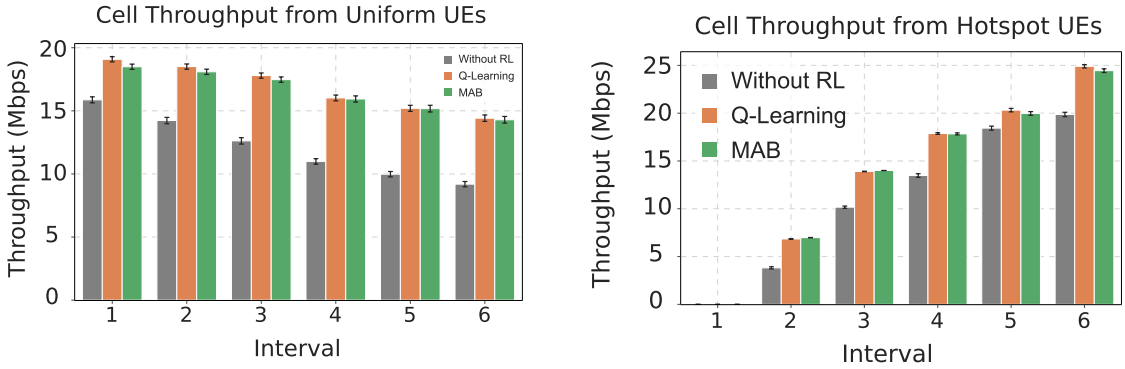


Figure 5.30: Average cell throughput comparison between Q-Learning and MAB for hotspot and uniform users.

state representations, MAB avoids the need for exploratory analysis to identify how to model the system's state space effectively. As a result, MAB offers a more lightweight alternative, especially in scenarios where defining informative states is difficult or costly.

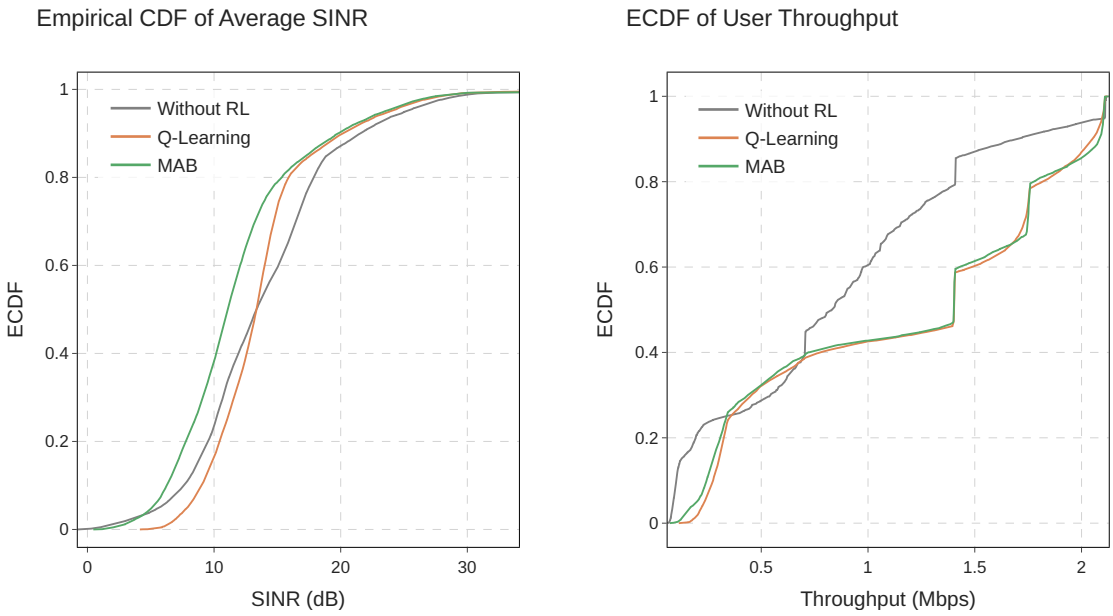


Figure 5.31: ECDF of throughput and SINR for Q-Learning and MAB.

Lastly, Figures 5.32 and 5.33 present the action frequency plots and transition matrices for two simulation jobs. Similar to the behavior observed with QL agents, the MAB learning process leads to certain actions being prioritized within each interval. In some cases, one action has significantly higher selection frequency, indicating a clearly higher expected reward. In other cases, multiple actions appear with similar frequency, suggesting that either their estimated rewards are comparable or the agent took longer to converge within a given interval.

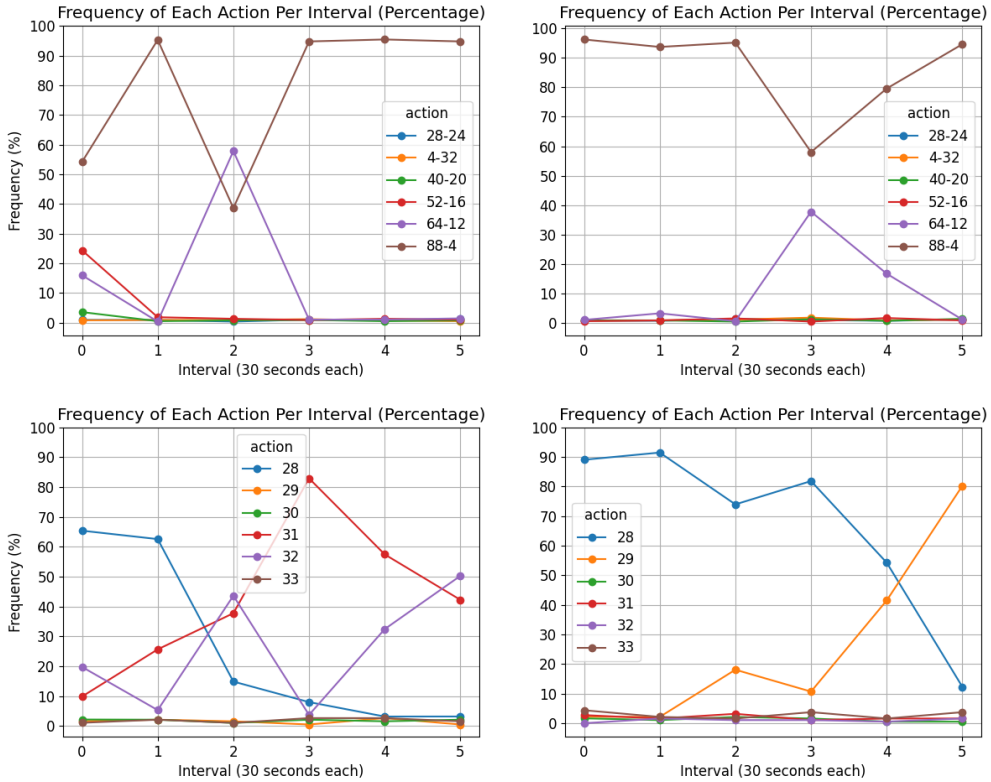


Figure 5.32: Frequency of bandwidth and RSRQ threshold actions for MAB agents.

The transition matrices confirm that the agents tend to converge toward a subset of more desirable actions, as there is a high number of self-transitions. It is worth recalling that these matrices represent transitions across all intervals combined, highlighting general learning trends rather than interval-specific behavior.

The figures also reveal that the MAB agents exhibit a clearer distinction between action frequencies, with certain actions being selected far more often than others. This behavior may be influenced by the smaller exploration space of the MAB agent, which consists solely of the set of available actions. In contrast, the Q-Learning agent must explore a much larger state-action space, which increases the complexity of the learning process. As a result, within the same time window, the MAB agent has more opportunities to repeatedly explore and refine its estimates for each action, potentially leading to faster convergence toward preferred actions.

However, although the average cell throughput results demonstrate that the MAB agents can achieve performance comparable to Q-Learning, its simpler reward estimation mechanism, based on immediate rewards, can limit its ability to consistently find the optimal solution in more complex situations. Unlike Q-Learning, MAB does not consider future rewards, which may lead to suboptimal decisions in environments where the long-term impact of an action is significant.

For example, the simulation job shown on the left in Figures 5.32 and 5.33 produced

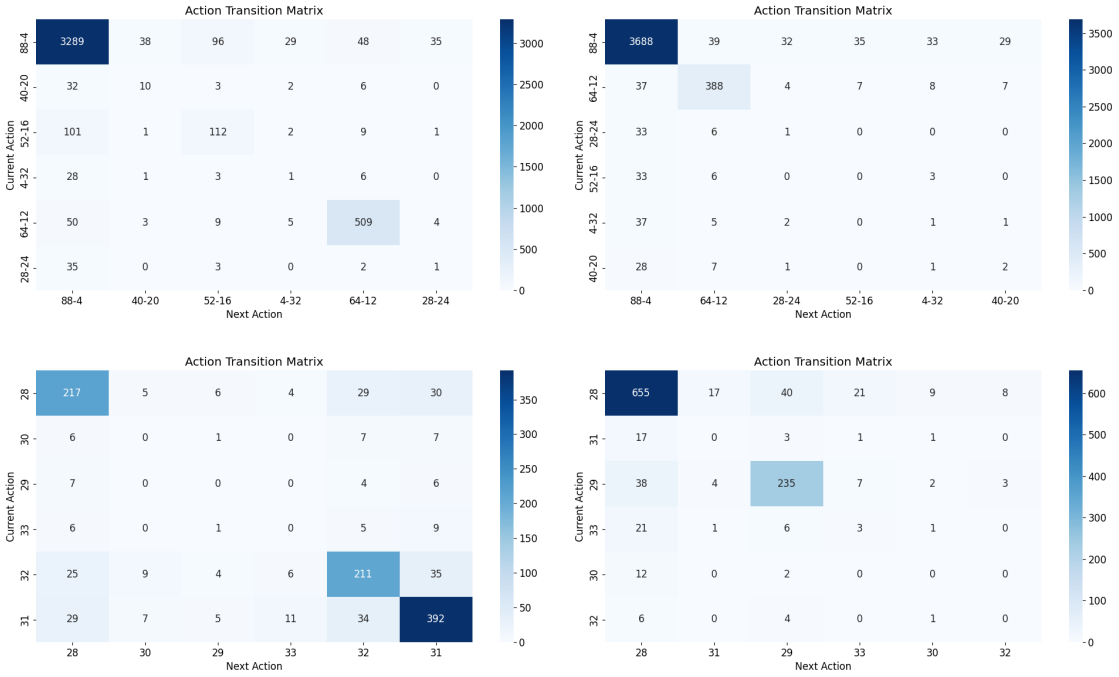


Figure 5.33: Action transition matrices for the bandwidth and RSRQ threshold for MAB agents.

results consistent with the overall average shown in Figure 5.29, where the MAB agents achieved throughput levels similar to those of the QL agents. In contrast, the job on the right resulted in noticeably lower throughput for the MAB solution compared to QL. Despite this, it still provided a considerable improvement over the static configuration without RL, confirming the MAB framework's ability to offer gains even under its simplified learning structure.

This performance degradation observed in some simulation jobs can be attributed to inherent characteristics of the MAB approach, namely, its reliance on immediate rewards through the sample-average method and its stateless nature. Because the MAB agent is nonassociative, it does not learn how the best action may vary depending on the system state, which can be a limitation in more complex or highly dynamic scenarios. However, these same characteristics also contribute to its main strengths. By avoiding the need for state representation and future reward estimation, MAB significantly reduces computational complexity and can adapt more quickly in environments where rapid changes occur or where defining meaningful states is challenging.

Despite underperforming compared to Q-Learning, the MAB-based framework still achieved considerable gains over the static configuration and demonstrated throughput levels close to those of Q-Learning in most scenarios, as shown in Figure 5.29.

These results not only position MAB as a viable alternative, but show the flexibility of the proposed multi-agent framework. Depending on the end goal, the agents can be modeled using different algorithms that best suit the requirements of the

deployment scenario. In environments where low computational complexity, faster convergence, or minimal tuning are priorities, MAB agents would be a good choice. In more complex or state-dependent environments where long-term reward estimation is important, Q-Learning or other state-aware algorithms may provide better results. The same environment might even benefit from using both the MAB and QL for different tasks that have different requirements.

Chapter 6

Conclusions and future works

This work presented a dynamic solution to mitigate inter-cell interference using Fractional Frequency Reuse in dense mobile networks, particularly in scenarios where interference is intensified by the sudden appearance of zones with high user density (hotspots). After reviewing the current direction of the state of the art, Chapter 2 introduced the interference problem in mobile networks, detailing how different types of interference can impact system performance and focusing specifically on co-channel interference (CCI), the most extensively studied form in the literature. This chapter also outlined key challenges associated with CCI and introduced FFR as an effective Inter-cell Interference Coordination (ICIC) technique.

Chapter 3 described how the research problem was characterized and modeled through network simulations. It also discussed the thesis hypothesis through exploratory analyses that assessed the impact of user classification and sub-band allocation across two baseline scenarios with different hotspot deployments. These results provided strong evidence that a multi-parameter control strategy could improve throughput in hotspot environments. Based on these findings, the proposed solution was developed and introduced in Chapter 4.

The proposed solution consisted of a hierarchical multi-agent framework based on Reinforcement Learning. The framework uses indirect coordination between two RL agents, each responsible for controlling one of the key parameters identified in the problem analysis: user classification via RSRQ threshold and bandwidth allocation via RB allocation. This coordination strategy was designed to address the dynamic interference caused by the progressive activation of hotspots during the simulation.

The results showed that the proposed framework consistently outperformed not only a statically configured system but also single-agent versions of the solution, in which only one parameter was controlled dynamically. Compared to the single-agent approaches, the multi-agent solution provided throughput gains ranging from 3% to 99.4%. Controlling only one parameter proved suboptimal in all evaluated cases, as the studied FFR parameters are strongly interdependent and their joint configuration significantly impacts system performance. The multi-agent framework was able to outperform both well-configured and poorly configured scenarios, consistently converging to high-performing configurations, especially for the lowest-performing users and during intervals with higher interference.

Another contribution of this work was demonstrating the the flexibility of the

proposed framework. Its hierarchical design allows for the integration of different RL algorithms. Two learning strategies were implemented and evaluated: Q-Learning (QL) and Multi-Armed Bandits (MAB). While QL achieved the highest overall performance, the MAB-based implementation also provided comparable throughput gains with lower computational complexity. This shows that the hierarchical framework can be adapted to meet different system requirements, balancing performance and implementation complexity.

Given the flexibility of the proposed solution, there are several directions for future work that can enhance its applicability to diverse scenarios and use cases. One option involves extending the current setup to more complex environments, such as those with a larger number of cells, the inclusion of wrap-around techniques, or the introduction of user mobility. These modifications would allow for a more robust evaluation of the framework. Additionally, further research can investigate the integration of different types of RL agents within the multi-agent system. While this thesis evaluated QL and MAB agents independently, future work could explore their combined use, such as MAB controlling the RSRQ threshold and QL controlling the bandwidth allocation. Moreover, more advanced RL methods, such as Deep Q-Learning (DQL) or Actor-Critic models, could be employed to improve learning performance. Finally, this work did not formally address the computational complexity or convergence time of the proposed approach. A detailed analysis of these aspects would be a valuable contribution in future studies.

6.1 Discussing the research questions

To conclude, this section revisits the research questions outlined in Chapter 1 and provides brief answers based on the analyses and findings presented throughout the thesis.

1. What are the individual and combined impacts of bandwidth allocation and user classification on throughput and interference in FFR-based ICIC, particularly in hotspot scenarios?
   Exploratory results presented in Chapter 3 demonstrated that hotspot deployments can significantly degrade system performance. Both the RSRQ threshold and the bandwidth allocation were shown to have strong individual impacts. However, the analysis also revealed that adjusting only one parameter is often insufficient. In some cases, improper threshold settings led to throughput losses of up to 30% following a change in hotspot location, while jointly optimizing both parameters resulted in gains of up to 22.27%. This indicates that the combined configuration of these parameters is key in maintaining performance in environments with sudden changes on user distribution.

2. How can reinforcement learning be applied to dynamically control multiple FFR parameters to effectively mitigate inter-cell interference in scenarios with emergent hotspots?

Based on the insights from recent research presented in Chapter 1 and the challenges described in Chapter 2, reinforcement learning was used to build a hierarchical framework composed of two independent agents that indirectly collaborate by sharing the reward space to control two FFR parameters. Through direct interaction with the environment, the agents learn the best joint configuration for the RSRQ threshold and bandwidth allocation in response to the sudden appearance of new hotspots. This dynamic approach enables real-time mitigation of ICI, overcoming the limitations of fixed or rule-based FFR schemes.

3. To what extent does a multi-agent reinforcement learning approach improve system performance metrics, such as user throughput and SINR, compared to static configurations and single-agent baselines?
   The results presented in Chapter 5 showed that the proposed multi-agent approach consistently outperformed both static and single-agent baselines. Compared to the single-agent solutions, the multi-agent framework achieved throughput gains ranging from 3% to 99.4%, depending on the interference level and scenario. When compared to the static configuration, the gains were even more significant, reaching up to 182%. For both cases, the gains were higher during intervals with more hotspots and for the lowest-performing users. While the framework consistently improved throughput, it occasionally resulted in lower SINR for some users.

4. What are the key challenges associated with developing RL-based solutions for interference coordination in mobile networks?
   As discussed in Chapters 4 and 5, developing RL-based ICIC can involve many challenges. For approaches such as QL, defining a reward signal and a state that is useful for the agent's learning goal can be difficult depending on the available performance metrics and the temporal window in which the agent operates. Other aspects that require careful investigation are choosing the strategy for exploring actions or state-action pairs, and tuning the exploration decay to balance the learning speed and stability. Additionally, enabling effective collaboration between independent agents proved to be the most challenging aspect. Ensuring that agents can jointly learn coordinated behavior through shared feedback alone requires careful architectural and reward design choices.

5. How does the use of different learning strategies (e.g., Q-Learning vs. Multi-Armed Bandits) impact performance and computational complexity in the context of dynamic ICIC?
   In Chapter 5, both Q-Learning and MAB agents were implemented and compared under the same framework. Q-Learning, with its state-aware learning process, achieved higher performance, particularly when interference was higher. However, it required additional implementation effort to define

meaningful states, and it involves exploring a larger number of state-action pairs. In this work, only six states were considered for Q-Learning, but the size of the Q-table grows proportionally with the number of states and available actions, increasing memory requirements and complexity. MAB agents, being stateless and simpler to implement, delivered competitive performance, offering a favorable trade-off between complexity and performance, especially in environments with limited processing capacity.

6.2 Scientific publications

During the development of this doctoral thesis, scientific works were published by the author as a direct outcome of the research presented herein. Additional publications resulted from collaborations with colleagues from the GppCom research group at UFRN, activities carried out as part of the nRIC research project in partnership with Lenovo, work as an invited researcher at Efrei Paris, and contributions made while co-supervising master's students. The list below compiles all related publications produced throughout the course of this PhD:

1. **Solution for Interference in Hotspot Scenarios Applying Q-Learning on FFR-Based ICIC Techniques** (2021 Sensors). This work was a compilation of past exploratory analysis performed during master's dissertation with additional results from the single-agent Q-Learning-based ICIC.
2. **Frequency Scanning xApps: O-RAN RIC and GNU Radio with RTL-SDR use case** (2022 SBrT). This work was a product of the research project in partnership with Lenovo, presented at the conference SBrT by the author.
3. **Prototyping near-real time RIC O-RAN xApps for Flexible ML-based Spectrum Sensing** (2022 NFV-SDN). This work was a product of the research project in partnership with Lenovo.
4. **Machine Learning Applied to Anomaly Detection on 5G O-RAN Architecture** (2023 PCS). This work was a product of the research project in partnership with Lenovo.
5. **Open-source emulation-based test environment to settle O-RAN-compliant trials** (2023 ITU). This work was a product of the research project in partnership with Lenovo.
6. **A Multiarmed Bandit Approach for LTE-U/Wi-Fi Coexistence in a Multicell Scenario** (2023 Sensors). This work was a product of a research collaboration with researchers from the research group GppCom.
7. **Gerenciamento de Largura de Banda em Sistemas 3GPP: uma Análise de Desempenho** (2024 ENCOM). This work was a product of co-orienting a master's student.
8. **A Study on Bandwidth Management and User Classification in Hotspot Scenarios** (2024 VTC-Fall). This work was a product of an exploratory analysis used to guid the development of the final solution porposed in this thesis. These results are included in Chapter 3.

Lastly, the results presented in this work are being organized to target journal publications.

Bibliography

3GPP (2010), TS 36.133: Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); requirements for support of radio resource management (Release 8), Technical report, 3GPP.

3GPP (2024), TR38.913: Technical Specification Group Radio Access Network; Study on Scenarios and Requirements for Next Generation Access Technologies; (Release 18), Relatório técnico, 3GPP.

3rd Generation Partnership Project (3GPP) (2021), Evolved universal terrestrial radio access (e-utra); physical layer procedures, Relatório Técnico TS 36.213.

Agarwal, Bharat, Mohammed Amine Togou, Marco Ruffini & Gabriel-Miro Muntean (2022), 'A Comprehensive Survey on Radio Resource Management in 5G HetNets: Current Solutions, Future Trends and Open Issues', *IEEE Communications Surveys & Tutorials* **24**(4), 2495–2534.
URL: *https://ieeexplore.ieee.org/document/9896125/*

Ahmad, Ishtiaq, Sajjad Hussain, Sarmad Nozad Mahmood, Hala Mostafa, Ahmed Alkhayyat, Mohamed Marey, Ali Hashim Abbas & Zainab Abdulateef Rashed (2023), 'Co-Channel Interference Management for Heterogeneous Networks Using Deep Learning Approach', *Information* **14**(2), 139.
URL: *https://www.mdpi.com/2078-2489/14/2/139*

Alam, Mohammed Jaber, Md Rahat Hossain, Salahuddin Azad & Ritesh Chugh (2023), 'An overview of LTE/LTE-A heterogeneous networks for 5G and beyond', *Transactions on Emerging Telecommunications Technologies* **34**(8), e4806.

Alzubaidi, Osamah Thamer Hassan, Mhd Nour Hindia, Kaharudin Dimyati, Kamarul Ariffin Noordin, Amelia Natasya Abdul Wahab, Faizan Qamar & Rosilah Hassan (2022), 'Interference Challenges and Management in B5G Network Design: A Comprehensive Review', *Electronics* **11**(18), 2842.
URL: *https://www.mdpi.com/2079-9292/11/18/2842*

Benni, Nirmalkumar S. & Sunilkumar S. Manvi (2022), Clustering Algorithm To Mitigate Intra And Inter Cell Interference In 5G Backhaul Wireless Mesh Networks, *em* '2022 IEEE 19th India Council International Conference (INDICON)', IEEE, Kochi, India, pp. 1–8.
URL: *https://ieeexplore.ieee.org/document/10039818/*

Burhanuddin, Liyana Adilla Binti, Xiaonan Liu, Yansha Deng, Maged Elkashlan & Arumugam Nallanathan (2023), 'Inter-Cell Interference Mitigation for Cellular-Connected UAVs Using MOSDS-DQN', *IEEE Transactions on Cognitive Communications and Networking* **9**(6), 1596–1609.
URL: *https://ieeexplore.ieee.org/document/10227360/*

Busoniu, Lucian, Robert Babuska & Bart De Schutter (2006), Multi-Agent Reinforcement Learning: A Survey, *em* '2006 9th International Conference on Control, Automation, Robotics and Vision', IEEE, Singapore, pp. 1–6.
URL: *http://ieeexplore.ieee.org/document/4150194/*

Cavalcanti, F. Rodrigo P., Tarcisio F. Maciel, C. Freitas Jr. Walter & Yuri C. B. Silva (2018), *Comunicaçao movel celular*, 1ª edição, Elsevier.

Cavalcanti, F. Rodrigo P., Tarcisio F. Maciel, Walter C. Freitas Jr. & Yuri C. B. Silva (2018), *Comunicação Móvel Celular*, 1ª edição, Elsevier, Rio de Janeiro. Inclui bibliografia e índice.

Chen, Shuaifei, Jiayi Zhang, Jing Zhang, Emil Björnson & Bo Ai (2022), 'A survey on user-centric cell-free massive MIMO systems', *Digital Communications and Networks* **8**(5), 695–719.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S2352864821001024*

Chikha, Wassim Ben, Marie Masson, Zwi Altman & Sana Ben Jemaa (2024), 'Radio Environment Map Based Inter-Cell Interference Coordination for Massive-MIMO Systems', *IEEE Transactions on Mobile Computing* **23**(1), 785–796.
URL: *https://ieeexplore.ieee.org/document/9954187/*

Cisco Systems, Inc. (2023), Cisco converged core: Enabling 5g opportunities, White paper, Cisco. Cisco Public, Document ID: C11-3964490-00, Published 11/2023.
URL: *https://www.cisco.com/c/en/us/solutions/service-provider/5g-network-architecture.html*

CTIA, Accenture (2022), 'Spectrum Allocation in the United States'.

de O. Nóbrega, Luciana F., Iago D. do Rêgo & Vicente A. de Sousa Jr. (2018), Análise de desempenho de técnicas icic usando o ns-3, *em* 'Anais do Encontro de Ciência e Tecnologia do Oeste Potiguar (ECOP 2018)', Universidade Federal Rural do Semi-Árido, Pau dos Ferros, RN, Brasil, pp. 105–112. ISSN 2526-7574.
URL: *https://periodicos.ufersa.edu.br/index.php/ecop*

de Santana, Pedro M., Vicente A. de Sousa Jr., Fuad M. Abinader Jr. & José M. de C. Neto (2019), 'DM-CSAT: a LTE-U/Wi-Fi coexistence solution based on reinforcement learning', *Telecommunication Systems* **1**, 1–12.

do Rego, Iago Diógenes & Vicente A. de Sousa Jr. (2021), 'Solution for interference in hotspot scenarios applying q-learning on ffr-based icic techniques', *Sensors* **21**(23), 7899.

do Rêgo, Iago D., Luciana F. de O. Nóbrega & Vicente A. de Sousa Jr. (2018), Análise paramétrica do strict frequency reuse utilizando o ns-3, *em* 'XXXVI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT)', Campina Grande, PB, Brazil. In Portuguese.

Du, D. R., J. Li, Z. Y. Ding, L. Q. Wu, L. Li & S. J. Huang (2022), 'A Novel ICIC Scheme Combining 3D ML-SFR and CoMP', *Radioengineering* **31**(1), 85–93.
URL: *https://www.radioeng.cz/fulltexts/2022/22_01_0085_0093.pdf*

Fahim Raouf, Amir Hossein, Sung Joon Maeng, Ismail Guvenc, Özgür Özdemir & Mihail Sichitiu (2023), Spectrum Monitoring and Analysis in Urban and Rural Environments at Different Altitudes, *em* '2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)', IEEE, Florence, Italy, pp. 1–7.
URL: *https://ieeexplore.ieee.org/document/10200994/*

Gulia, Sandeep, Anwar Ahmad, Sandeep Singh & Mangal Deep Gupta (2022), 'Interference management in backhaul constrained 5G HetNets through coordinated multipoint', *Computers and Electrical Engineering* **100**, 107982.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0045790622002476*

Guo, Wanying, Bojun Wang, Lian Zhao & Isma Farah Siddiqui (2024), 'Resource allocation-aware efficient interference management technique for ultra-dense Femto environment', *Computer Communications* **225**, 120–125.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0140366424001877*

Hamza, Abdelbaset S., Shady S. Khalifa, Haitham S. Hamza & Khaled Elsayed (2013), 'A Survey on Inter-Cell Interference Coordination Techniques in OFDMA-Based Cellular Networks', *IEEE Communications Surveys & Tutorials* **15**(4), 1642–1670.
URL: *http://ieeexplore.ieee.org/document/6476063/*

Hasan, Ayesha & Bilal Muhammad Khan (2022), 'Coexistence Management in Wireless Networks-A Survey', *IEEE Access* **10**, 38600–38624.
URL: *https://ieeexplore.ieee.org/document/9750069/*

Haykin, Simon S. (2009), *Neural networks and learning machines*, 3ª edição, Pearson Education, Upper Saddle River, NJ.

Instituto Metrópole Digital / UFRN (2025*a*), 'NPAD's hardware information'. Accessed: 2025-01-01.
URL: *https://npad.ufrn.br/npad/hardware*

Instituto Metrópole Digital / UFRN (2025*b*), 'Núcleo de Processamento de Alto Desempenho'. Accessed: 2025-01-01.
URL: *http://npad.ufrn.br/*

ITU-R (2015), 'RECOMMENDATION ITU-R V.431-8 - Nomenclature of the frequency and wavelength bands used in telecommunications'.
URL: *https://www.itu.int/dms_pubrec/itu-r/rec/v/R-REC-V.431-8-201508-I!!PDF-E.pdf*

Kassam, Joumana, Daniel Castanheira, Adão Silva, Rui Dinis & Atílio Gameiro (2023), 'A Review on Cell-Free Massive MIMO Systems', *Electronics* **12**(4), 1001.
URL: *https://www.mdpi.com/2079-9292/12/4/1001*

Kihato, Peter, Stephen Musyoki & Antony Onim (2022), 'Fractional Frequency Reuse Optimal SINR Threshold Selection Using NIR and ISODATA', *Telecom* **3**(3), 433–447.
URL: *https://www.mdpi.com/2673-4001/3/3/23*

Kimura, Tatsuaki & Masaki Ogura (2021), 'Distributed 3D Deployment of Aerial Base Stations for On-Demand Communication', *IEEE Transactions on Wireless Communications* **20**(12), 7728–7742.
URL: *https://ieeexplore.ieee.org/document/9454276/*

Konishi, Satoshi (2013), 'Comprehensive Analysis of Heterogeneous Networks with Pico Cells in LTE-Advanced Systems', *IEICE Transactions on Communications* **E96.B**(6), 1243–1255.
URL: *https://www.jstage.jst.go.jp/article/transcom/E96.B/6/E96.B_1243/_article*

Kousias, Konstantinos, Mohammad Rajiullah, Giuseppe Caso, Ozgu Alay, Anna Brunstorm, Luca De Nardis, Marco Neri, Usman Ali & Maria-Gabriella Di Benedetto (2022), Coverage and performance analysis of 5G non-standalone deployments, *em* 'Proceedings of the 16th ACM Workshop on Wireless Network Testbeds, Experimental evaluation & Characterization', ACM, Sydney NSW Australia, pp. 61–68.

Kumbhar, Abhaykumar, Farshad Koohifar, Ismail Guvenc & Bruce Mueller (2017), 'A Survey on Legacy and Emerging Technologies for Public Safety Communications', *IEEE Communications Surveys & Tutorials* **19**(1), 97–124.
URL: *http://ieeexplore.ieee.org/document/7573003/*

Lam, Sinh Cong, Bach Hung Luu, Nam Hoang Nguyen & Trong Minh Hoang (2023), 'Performance of Modified Fractional Frequency Reuse in Nakagami-m Fading Channel', *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* **E106.A**(7), 1016–1019.
URL: *https://www.jstage.jst.go.jp/article/transfun/E106.A/7/E106.A_2022EAL2101/_article*

Lam, Sinh Cong, Nam Hoang Nguyen, Vu Thien Tran & Viet-Cuong Ta (2022), Power Control in a modified Strict Frequency Reuse Algorithm Utilizing Q-Learning, *em* '2022 IEEE Ninth International Conference on Communications and Electronics (ICCE)', IEEE, Nha Trang, Vietnam, pp. 48–52.
URL: *https://ieeexplore.ieee.org/document/9852036/*

Lam, Sinh Cong & Xuan Nam Tran (2021), 'Fractional Frequency Reuse in Ultra Dense Networks', *Physical Communication* **48**, 101433.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1874490721001701*

Liang, Jiaxin, Haotian Miao, Kai Li, Jianheng Tan, Xi Wang, Rui Luo & Yueqiu Jiang (2025), 'A Review of Multi-Agent Reinforcement Learning Algorithms', *Electronics* **14**(4), 820.
URL: *https://www.mdpi.com/2079-9292/14/4/820*

Liu, Qiaoshou, Jianwen Zou & Yingchen Gu (2022), 'An Analysis of Distributed Joint Spectrum Resource Allocation With Power Control in Ultra-Dense Cellular Networks', *IEEE Transactions on Vehicular Technology* **71**(10), 11201–11215.
URL: *https://ieeexplore.ieee.org/document/9826382/*

Luu, B.-H., S.-C. Lam, N.-H. Nguyen & T.-M. Hoang (2024), 'Performance of the User in the TDD NOMA Cellular Networks Enabling FFR', *Radioengineering* **33**(2), 312–321.
URL: *https://www.radioeng.cz/fulltexts/2024/24_02_0312_0321.pdf*

Malekzadeh, Mina (2023), 'Enabling efficient and reliable IoT deployment in 5G and LTE cellular areas for optimized service provisioning', *The Journal of Supercomputing* **79**(2), 1926–1955.

Marinescu, Andrei, Irene Macaluso & Luiz A. DaSilva (2018), A Multi-Agent Neural Network for Dynamic Frequency Reuse in LTE Networks, *em* '2018 IEEE International Conference on Communications Workshops (ICC Workshops)', IEEE, Kansas City, MO, USA, pp. 1–6.
URL: *https://ieeexplore.ieee.org/document/8403663/*

Ming, Yang, Ziyuan Sha, Shuang Jin & Yuhan Dong (2022), Fairness-Aware Soft Frequency Reuse for Multi-Cell OFDMA Networks with Statistical CSI, *em* 'GLOBECOM 2022 - 2022 IEEE Global Communications Conference', IEEE, Rio de Janeiro, Brazil, pp. 3593–3598.
URL: *https://ieeexplore.ieee.org/document/10001086/*

Nafees, Muhammad, John Thompson & Majid Safari (2021), 'Multi-Tier Variable Height UAV Networks: User Coverage and Throughput Optimization', *IEEE Access* **9**, 119684–119699.
URL: *https://ieeexplore.ieee.org/document/9521913/*

ns-3 (2025), ns-3 documentation, Página na internet, ns-3.
URL: `https://www.nsnam.org/documentation/`

ns-3 Project (2016*a*), 'Attribute list documentation (ns-3.26)', `https://www.nsnam.org/docs/release/3.26/doxygen/_attribute_list.html`. Accessed: 2025-03-16.

ns-3 Project (2016*b*), 'Ltefrstrictalgorithm class reference (ns-3.26)', `https://www.nsnam.org/docs/release/3.26/doxygen/classns3_1_1_lte_fr_strict_algorithm.html`. Accessed: 2025-03-16.

ns3 (2025*a*), 'LTE Design Documentation'. Accessed: 2025-01-01.
URL: *https://www.nsnam.org/docs/release/3.36/models/html/lte-design.html*

ns3 (2025*b*), 'ns-3 Object Model Documentation'. Accessed: 2025-01-01.
URL: *https://www.nsnam.org/docs/manual/html/object-model.html*

ns3 (2025*c*), 'ns-3 Overview - What is ns-3?'. Accessed: 2025-01-01.
URL: *https://www.nsnam.org/about/*

Onim, Antony, S. Musyoki & P. Kihato (2022), 'Selection of optimal SINR threshold in fractional frequency reuse by comparing Otsu's and entropy method', *Heliyon* **8**(11), e11736.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S2405844022030249*

Pan, Meng-Shiuan & Ming-Yang Wu (2024), 'A genetic algorithm based flow control scheme for LTE-NR dual connectivity networks', *Wireless Networks* **30**(1), 437–452.

Richard S. Sutton, Andrew G. Barto (2018), *Reinforcement Learning: An Introduction*, MIT Press.

Rinaldi, Federica, Helka-Liina Maattanen, Johan Torsner, Sara Pizzi, Sergey Andreev, Antonio Iera, Yevgeni Koucheryavy & Giuseppe Araniti (2020), 'Non-Terrestrial Networks in 5G & Beyond: A Survey', *IEEE Access* **8**, 165178–165200.
URL: *https://ieeexplore.ieee.org/document/9193893/*

Rochman, Muhammad Iqbal, Vanlin Sathya, Damian Fernandez, Norlen Nunez, Ahmed Ibrahim S., William Payne & Monisha Ghosh (2023), 'A comprehensive analysis of the coverage and performance of 4G and 5G deployments', *Computer Networks* **237**, 110060.

Safdar, Hashim, Rahat Ullah & Zubair Khalid (2022), 'Game Theoretic based Distributed Dynamic Power Allocation in Irregular Geometry Multicellular Network', *International Journal of Computer Science and Network Security* **22**(7), 199–205.
URL: *https://doi.org/10.22937/IJCSNS.2022.22.7.24*

Sallakh, Usama, Stephen S. Mwanje & Andreas Mitschele-Thiel (2014), Multi-parameter Q-Learning for downlink Inter-Cell Interference Coordination in LTE SON, *em* '2014 IEEE Symposium on Computers and Communications (ISCC)', IEEE, Funchal, Madeira, Portugal, pp. 1–6.
URL: *http://ieeexplore.ieee.org/document/6912599/*

Seo, Minyoung, Seok-Ho Chang, Jong-Man Lee, Ki-Hun Kim, Hyun Park & Sang-Hyo Kim (2023), 'Optimal Coverage of Full Frequency Reuse in FFR Networks in Relation to Power Scaling of a Base Station', *Sensors* **23**(21), 8925.
URL: *https://www.mdpi.com/1424-8220/23/21/8925*

Sesia, Stefania, Issam Toufik & Matthew Baker (2011), *LTE – The UMTS Long Term Evolution: From Theory to Practice*, 2ª edição, Wiley.

Shahid, Husnain, Miguel Ángel Vázquez, Xavier Artiga, Matilde Sánchez-Fernández & Álvaro Callejas-Ramos (2024), Performance Evaluation of Fractional Frequency Reuse in Multi-Beam Satellite System, *em* '2024 IEEE 25th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)', IEEE, Lucca, Italy, pp. 641–645.
URL: *https://ieeexplore.ieee.org/document/10694407/*

Su, Yuhan, Zhibin Gao, Xiaojiang Du & Mohsen Guizani (2023), 'User-centric base station clustering and resource allocation for cell-edge users in 6G ultra-dense networks', *Future Generation Computer Systems* **141**, 173–185.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0167739X22003727*

Sun, Qian, Huiliang Liu, Fei Peng, Fan Jiang, Xiaoyue Bao & Xinyu Zhang (2022), Interference Analysis Method for C-band Spaceborne SAR System to Ground-based Radars, *em* '2022 International Conference on Microwave and Millimeter Wave Technology (ICMMT)', IEEE, Harbin, China, pp. 1–4.
URL: *https://ieeexplore.ieee.org/document/10023232/*

Susanto, Misfa, Soraida Sabella & Fx Arinto Setyawan (2024), 'Reducing interference effects in distributed D2D communication underlaying multicell cellular communication network using Soft Fractional Frequency Reuse', *International Journal of Electronics and Telecommunications* pp. 901–908.
URL: *https://journals.pan.pl/dlibra/publication/152076/edition/133216/content*

Tani, Junya & Kenichi Higuchi (2022), 'Online Probabilistic Activation Control of Base Stations Utilizing Temporal System Throughput and Activation States of Neighbor Cells for Heterogeneous Networks', *IEICE Transactions on Communications* **E105.B**(11), 1458–1466.
URL: *https://www.jstage.jst.go.jp/article/transcom/E105.B/11/E105.B_2021EBT0005/_article*

Trabelsi, Nessrine, Lamia Chaari Fourati & Chung Shue Chen (2024), 'Interference management in 5G and beyond networks: A comprehensive survey', *Computer Networks* **239**, 110159.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1389128623006047*

Ullah, Rahat, Abdullah Gani, Muhammad Shiraz, Imran Khan Yousufzai & Khalid Zaman (2022), 'Auction Mechanism-Based Sectored Fractional Frequency Reuse for Irregular Geometry Multicellular Networks', *Electronics* **11**(15), 2281.
URL: *https://www.mdpi.com/2079-9292/11/15/2281*

Ullah, Rahat, Zubair Khalid, Fargham Sandhu & Imran Khan (2021), 'Hungarian Mechanism based Sectored FFR for Irregular Geometry Multicellular Networks', *EMITTER International Journal of Engineering Technology* **9**(2), 313–325.
URL: *http://emitter.pens.ac.id/index.php/emitter/article/view/627*

Watkins, Christopher (1989), Learning From Delayed Rewards, Tese de doutorado, King's College, London.

Wei, Haichao, Na Deng & Martin Haenggi (2022), 'Performance Analysis of Inter-Cell Interference Coordination in mm-Wave Cellular Networks', *IEEE Transactions on Wireless Communications* **21**(2), 726–738.
URL: *https://ieeexplore.ieee.org/document/9494282/*

Więcek, Dariusz, Marcin Mora & Igor Michalski (2024), 'Coexistence of 5G/6G and Digital Terrestrial Television Networks', *Applied Sciences* **14**(13), 5756.
URL: *https://www.mdpi.com/2076-3417/14/13/5756*

Yang, Lihua, Teng Joon Lim, Junhui Zhao & Mehul Motani (2021), 'Modeling and Analysis of HetNets With Interference Management Using Poisson Cluster Process', *IEEE Transactions on Vehicular Technology* **70**(11), 12039–12054.
URL: *https://ieeexplore.ieee.org/document/9547787/*

Ying, Mingjun & Shuyu Wang (2022), Capacity Analysis and Hybrid Power Allocation for Multi-cell 5G Networks, *em* '2022 11th International Conference on Communications, Circuits and Systems (ICCCAS)', IEEE, Singapore, Singapore, pp. 192–197.
URL: *https://ieeexplore.ieee.org/document/9824531/*

Zhou, Yiqing, Ling Liu, Hongyan Du, Lin Tian, Xiaodong Wang & Jinglin Shi (2014), 'An overview on intercell interference management in mobile cellular networks: From 2G to 5G', *IEEE* **1**, 217–221.