Received: February 13, 2024. Revised: April 7, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Reinforcement Learning-based Joint Interference Mitigation and Resource Allocation in Dense Beyond 5G Heterogeneous Networks

Samar Farhan[1]* **Mohammed Aal-nouman[1]**

[1]*College of Information Engineering, Al-Nahrain University, Baghdad, Iraq*
* Corresponding author's Email: samar_ie@yahoo.com

Abstract: The deployment of small cells (SCs) in heterogeneous beyond 5G (B5G) networks holds immense promise in meeting the ever-growing demand for data rates and ensuring the desired Quality-of-Service (QoS) for cell-edge users equipment (CEUEs) in densely populated B5G networks. In our pursuit of jointly maximizing the sum-rate while minimizing interference for CEUEs, we present an optimization-based approach with a specific focus on fulfilling the minimum QoS requirements of CEUEs. Our innovative two-step algorithm begins with a reinforcement learning (RL)-based matching process among UEs and available resources, followed by optimal power allocation to UEs based on the matched pairs. At the first step, by leveraging a Q-learning-based method, our algorithm identifies the optimal UEs-resources pairing. The learning process utilizes the achieved sum-rate of each pairing among UEs and resources at each step and converges to a sub-optimal pairing among them. In the second step, power allocation for the selected pairing of the first step is solved in optimal manner. Achieving optimal power allocation is facilitated by exploiting the difference of concaves form of the objective function and harnessing the majorization-minimization (MaMi) technique considering the minimum required QoS of CEUEs.Our numerical results showcase the effectiveness of the proposed scheme, demonstrating near-optimal performance. The results show the employed RL approach effectively converges to the near-optimal pairing among UEs and resources in a dense environment. Additionally, it is evident that the optimal power allocation not only maximizes the sum-rate but also minimizes interference for CEUEs. Considering different values of macro cell (MC) transmission power and SC radius, the proposed schemes achieve a sum-rate enhancement of at least 10% and 25% compared to other existing matching and power allocation methods, respectively.

Keywords: Reinforcement learning, Q-Learning, Beyond-5G networks, Dense networks, Interference mitigation, Optimization.

1. Introduction

In response to the ever-increasing data rate demands of cellular users within traditional cellular networks, new communication paradigms have emerged. These paradigms incorporate innovative methods and architectures, including heterogeneous networks (HetNets)[1], multiple-input-multiple-output (MIMO) techniques, device-to-device (D2D) communication [2], and the dense deployment of users. These developments have ushered in a new era of connectivity solutions, addressing the growing needs of modern wireless communication. HetNets, in particular, have emerged as a promising approach in this landscape by introducing the concept of utilizing different network tiers to cater to varying classes of UEs. Given the dense deployment of UEs in today's wireless landscape, SCs have also gained prominence as a valuable addition to HetNets [3]. These SCs play a pivotal role in enhancing network capacity and improving UE's experiences in the face of escalating data demands.

In the context of dense B5G networks, where the frequency reuse factor approaches unity, the utilization of the same spectrum across different cells is a common practice. Moreover, network

Received: February 13, 2024. Revised: April 7, 2024.

densification strategies, particularly in HetNets with distinct tiers catering to various classes of UEs, are employed to meet the escalating data rate demands [4]. Consequently, each UE in close proximity to MC or SC BSs benefits from strong desired signal power and diminished interference. Whereas CEUEs contend with weaker desired signal strength and considerably heightened interference levels in comparison to cell-center UEs (CCUEs) [5]. Ensuring the minimum QoS for these CEUEs represents a paramount challenge in densely populated B5G networks. This predicament has been extensively explored in the existing literature, with numerous methodologies attempting to address it. Many of these techniques involve spectrum partitioning or user rearrangement strategies. Yet, a substantial portion relies on suboptimal approaches that cannot maintain a frequency reuse factor of one, a crucial factor in the context of B5G networks. Furthermore, several of these methods adopt heuristic approaches lacking a robust mathematical foundation.

In the context of CEUEs, the imperative need for robust and efficient methods is evident. Solutions founded on optimization principles and analytically solvable models tend to outperform heuristic approaches, making them a compelling choice [6]. Therefore, employing a resource allocation approach coupled with mathematical modelling emerges as a potent strategy for mitigating interference experienced by CEUEs while simultaneously enhancing and guaranteeing minimum QoS requirements. These methods typically operate under the assumption of achieving a reuse factor close to one, which is a fundamental requirement in the pursuit of reliable, robust, and nearly optimal solutions for the challenges faced by CEUEs. In addition to that, leveraging learning-based methods can significantly expedite the process of achieving optimal or near-optimal solutions. These approaches offer practicality and versatility, rendering the algorithm applicable across a broader range of scenarios. The adaptability and efficiency associated with learning-based techniques make them a valuable asset in addressing complex optimization challenges efficiently and effectively [7].

This paper delves into an investigation of the downlink transmission within dense HetNets featuring multiple SCs that harness the full extent of their cellular resources. Each UE within this network stipulates a minimum QoS requirement in terms of Signal-to-Interference-plus-Noise Ratio (SINR). The inherent challenge in such a densely populated environment lies in the substantial interference generated by the multitude of UEs. To address this issue, we formulate the problem as a sum-rate maximization task for the UEs, taking into account their respective minimum QoS requirements. The ensuing optimization problem takes on the form of a mixed-integer non-linear optimization challenge. To tackle this, we propose an innovative two-step algorithm. The first step employs RL techniques to establish a close-to-optimal matching between cellular resources and UEs. Leveraging a Q-learning-based approach and guided by the output of the second step, this algorithm progressively converges to a near-optimal matching. The second step focuses on power allocation, leveraging the matching from the first step to transform the initial mixed-integer non-linear problem into a non-linear one. By exploiting the concave nature of the objective function, we approximate the non-convex objective as a convex function. Additionally, we convert non-convex constraints into affine ones, employing an iterative interior point method to achieve an optimal solution to the optimization problem. The paper also presents the algorithmic implementation of the entire process, including the first and second step algorithms. Simulation results demonstrate that the second step allocates power to UEs in a dense B5G network optimally. The first step converges to a near-optimal matching and comprehends the dynamics of UEs within a cellular network. These simulations underscore the effectiveness of the proposed scheme in guaranteeing QoS for CEUEs in dense B5G networks. The main contributions of this paper are:

1- We present a mixed-integer non-linear optimization problem aiming at solving the problem of joint interference mitigation and resource allocation in B5G networks.

2- Our introduced two step algorithm is an effective method to maximize sum-rate and minimize the interference caused to CEUEs simultaneously. The algorithm is designed properly such that the output of the second step is used as an input for the first step such that the overall algorithm has a near-optimal performance.

3- The first step utilizes reinforcement learning-based methods to converge to a near optimal matching among UEs and cellular resources. The proposed scheme can effectively learn the dynamics of the cellular network.

4- A modified Q-learning algorithm defined with a novel action set has been proposed. The proposed action set of the algorithm significantly reduces the size of the Q-Matrix and the complexity of the algorithm.

5- We have proposed an optimal approach for the power allocation step related to the matching of the first step using the MaMi technique.
6- We have derived a lower bound for the sum-rate maximization step and maximize the lower bound iteratively in order to achieve the optimal point of the second step.and the appropriate corresponding method called Majorize-Minimization (MaMi) is used to give the optimal point.

The rest of the paper is structured as follows: section 2 explores the related work, section 3 describes the system model, section 4 explains the proposed scheme, the results and discussion are given in section 5 and finally the conclusion of this paper is discussed in section 6.

2. Related works

Given the paramount significance of CEUEs within the landscape of densely populated future cellular networks, the perpetual endeavour to diminish CEUEs interference remains a central concern. Literature has explored an array of interference mitigation techniques, with fractional frequency reuse (FFR) being among the notable contenders. FFR's modus operandi involves spectrum partitioning and allocation to user groups, albeit at the cost of substantially reducing spectral efficiency (SE), a critical consideration in the context of B5G networks. Although this approach manages to meet the minimum required QoS for UEs, it is inherently inefficient. In pursuit of heightened SE, alternative methodologies rooted in soft frequency reuse (SFR) [8] have emerged. SFR-based strategies strive to achieve performance levels approaching a frequency reuse factor of 1 while concurrently mitigating interference. These approaches partition the spectrum accordingly. Nonetheless, they typically result in CCUEs receiving signals with notably higher SINRs compared to their CEUEs counterparts due to heightened interference levels and path loss at the cell peripheries—a formidable challenge for CEUEs QoS assurance in future cellular networks. A recent effort by the authors of [9] leverages cell partitioning to tackle intercell interference and augment the SFR approach. However, this approach hinges on a heuristic cell partitioning scheme, lacking the optimality required for robust performance. Furthermore, many of these methods rely predominantly on heuristic techniques, lacking the solid mathematical underpinnings necessary for comprehensive problem-solving. The proliferation of UEs in densely populated environments has led to the emergence of HetNets. To attain high spectral efficiency (SE), opting for a reuse factor of one represents an optimal choice, but it brings forth the pressing concern of interference mitigation. The authors of [8] introduced a mechanism enabling macro cells and fixed/mobile SCs to dynamically allocate transmit power to their respective serving BSs. The mechanism focused on mitigating dynamic downlink interferences stemming from the mobility of both SCs and UEs within the network. The authors presented the Cell-User Mobility (CUM) model to analyze the mobility patterns of cells and users. However, their approach hinges primarily on the location of UEs, which inherently represents a suboptimal strategy. Further enhancements to their algorithm could be realized by incorporating Channel State Information (CSI) as a superior metric in contrast to the location-based method. In their work [1], the authors introduced an innovative SFR scheme to reduce interference and enhance network throughput. This scheme achieved its objectives by partitioning the cellular region into two distinct zones: center and edge, and explored two alternative shapes for the center zone of SCs: circular and irregular, then obtained the optimal radius value that maximized the throughput of the network. Additionally, this scheme relied on switching on /off the SCs based on their interference contribution value, which efferently reduced power consumption in 5G HetNets. However, the primary focus of this approach is not specifically geared towards minimizing interference caused to UEs and cannot guarantee their minimum QoS requirements. The authors of [9] used joint transmission coordinated multi-point to improve the performance of users in the cell expansion area (CEA), which suffer from interference and receive SINR less than 0.Multiple BSs collaborated to enhance SINR and overall throughput. Unlike traditional per-tier biasing, they employed particle swarm optimization (PSO) to balance load among SC BSs and maximize system throughput. However, the primary focus of this approach is load balancing rather than guaranteeing QoS for CEUEs, with its emphasis on UEs located in the CEA. In their study [10], the authors mitigated the interference in ultra-dense HetNets by employing a resource allocation-based approach in conjunction with cell partitioning and SFR schemes. The network consisted of three macro cells and inside each the femto cells added incrementally. The first femto cell utilized the same resources of the second or third macro cell based on a lowest interference and according to the frequency reuse factor of 3, then the neighbor femto cannot use the same resources. Therefore, the femtocells are grouped based on the different resources assigned to them cooperatively with macro cell. However, there is room for enhancing the approach's performance by defining a more robust objective function for optimization and providing corresponding solutions. Additionally, ensuring QoS for CEUEs remains a critical aspect that needs to be addressed within this framework. The authors of [11] addressed dense cellular networks and introduced a near-optimal matching mechanism between UEs and resources. They employed optimization-based techniques to maximize the cell's sum-rate. However, their approach did not assume HetNets and did not take into account the concept of CEUEs. The authors of [12] introduced an SFR algorithm called Load-Driven SFR, which dynamically adapts resource allocation parameters, specifically BS bandwidth assignment, based on the network's load distribution. This intelligent adjustment improves interference mitigation, and Load-Driven SFR demonstrated superior performance when compared to several implementations of the standard SFR algorithm that relied on fixed bandwidth allocation. However, it's worth noting that their approach did not utilize mathematical expressions, such as rate equations, and predominantly relied on heuristic concepts like SFR. In their work [13], the authors presented a resource allocation scheme aimed to enhance the performance of CEUEs within the context of Long Term Evolution-Advanced (LTE-A) systems. Their proposed algorithm focused on the optimal assignment and allocation of Carrier Components Resource Blocks (RBs), and Modulation and Coding Scheme indices to UEs. This allocation is determined based on QoS requirements, ensuring that the on-demand service requests are met efficiently. However, it's important to note that the study predominantly emphasizes RBs, did not assume a dense B5G network scenario, and did not specifically investigate interference mitigation for CEUEs.

Given the recent advancements in artificial intelligence (AI), the integration of learning-based approaches holds significant promise in addressing interference mitigation challenges. The authors of [14] proposed a power control scheme based on RL aimed at mitigating downlink inter-cell interference and conserving energy in ultra-dense SC deployments. This innovative scheme empowered BSs to efficiently schedule downlink transmit power, even without prior knowledge of the interference distribution and the channel states of neighbouring SCs. This scheme relied on state compromised of the density of user distribution in the cell, SINR of each user, and power gains of their channels. To specify further, the Q value determined BS transmission power and is updated according to the Bellman equation. The proposed optimized BS functionality in enhancing SINR with less energy consumption and lower interference through trial and error in the dynamic process of interference management. Furthermore, a deep RL is used to speed up the learning process when the number of users is `large. However, it's worth noting that the approach did not assume a mathematical expression for the rate metric. Instead, it relied on a heuristic combination of rate, energy consumption, and interference considerations.

The authors of [15] tackled the challenge of optimizing beamforming, power control, and interference coordination in a joint manner. They frame this complex problem as a non-convex optimization task with the goal of maximizing the SINR. To solve this intricate problem, they employed deep RL, leveraging the advantageous nature of deep Q-learning (DQL) to estimate the future rewards of various actions. This approach took into account the reported SINR and coordinates of the UE within the network every millisecond without the need to know the CSI.This approach is applied to the Barriers of voice and data to enhance the received data rate and reduce retransmissions. It’s worth noting that while the study explored the notion of serving a UE with multiple BSs and harnessed the capabilities of MIMO systems, it did not extensively consider the concept of CEUEs and their associated QoS requirements. In their work [16],the authors introduced a self-optimization algorithm designed to simultaneously manage energy-saving and interference coordination mechanisms within HetNets. Their approach utilized online learning framework to achieve these objectives. the two-stage framework consists of a global controller responsible for multiple macro cells and local controller for each macro cell is used. The global controller used an algorithm with the ability for learning, and its constraints are energy consumption and QoS. The algorithm converged on its Predicted convergence time. the global controller learned the control actions, and the local controller translated them into local decision. However, While the method mitigated interference, it primarily did in a heuristic manner, with its primary focus on energy conservation as the main objective. The authors of [17] presented an algorithm in their study that leverages DQL for intelligent interference mitigation. This approach focused on power control and aimed to solve a non-convex optimization problem to maximize the SINR using DQL techniques without the need for knowledge of CSI. Instead, each user needs to send its SINR and coordinates to BS. However, it's important to note that the channel models and the proposed approach are tailored to non-terrestrial networks, and the consideration of CEUEs and their QoS requirements is not addressed within this framework .in [18], the authors proposed a heuristic and centralized resource allocation algorithm for dense SCs environment .it is formulated as an NP-hard problem aimed to minimize interference, achieving maximum spectrum utilization and fairness between the users, which are classified according to their priority to ensure satisfying their QoS requirement. The proposed allowed to the individual entities to observe the resource allocation process to avoid problems of central node failure. However, the used method is heuristic and it is harder to adapt with varying network conditions.

The authors of [19] introduced a joint power control and beamforming algorithm to coordinate interference problem. They discussed the challenges of RL suggestions to address this problem. The implemented suggestions are decentralized multi-agent structures where each agent represented an independent BS, decentralized methods have another challenge for dynamic multi-agent networks which require to exchange information between agents which increases overhead. By proposing the RL method of a single agent, the demand for the traditional reward function and sharing information is eliminated and replaced with an efficient example to lead the learning process. However, guide learning by example cannot guarantee the optimal performance due the dynamic nature of the network. In [20] proposed multi-agent deep RL to manage the interference in multicell network. The approach maximized network spectral efficiency by optimizing the beamforming vectors and transmit power relying on users' locations and without CSI sharing. However, deploying multi-agent increases the complexity of converging to optimal solution since the state space is large.

The above literature showed that some existing approaches [1,9,10,12,14,16,18] tend to propose heuristic methods that lack a strong mathematical foundation. Some studies [1,9,10,17] have overlooked the crucial aspect of ensuring the minimum QoS for CEUEs. While certain researchers [15] have delved into complex MIMO systems and tackled the intricate problem of beamforming, there is still room for investigations in scenarios involving single-antenna systems. Furthermore, the approaches [8,19,20] give the suboptimal solution due to the infeasibility of the global one. Additionally, some studies like [11] have yet to harness the potential of learning-based approaches to unveil the intricate nature of cellular networks, and those that have may not have effectively addressed the needs of CEUEs. Considering the various approaches mentioned earlier and their drawbacks, the exploration of joint interference mitigation and the sum-rate optimization for CEUEs according to minimum QoS in dense B5G networks using learning-based methods remains a promising research avenue. Many researchers have focused on CEUEs, but not necessarily within the context of B5G networks. Hence, our research aims to delve into this relatively underexplored realm and put forth novel approaches to address these challenges effectively by defining a new state space for the modified Q-learning algorithm, we improve the performance of the learning method. In addition to that, we utilize the MaMi technique by finding a novel lower bound for the objective function. which has not been considered previously, according to the best of our knowledge.

3. System model

Within this section, we delve into the network model and the underlying assumptions. Following that, we provide a comprehensive explanation of variable definitions, mathematical modelling, and the formulation of the problem, which encompasses both joint interference mitigation and the maximization of the sum-rate for CEUEs in dense B5G networks.Fig.1 represent the system model of proposed dense B5G network.

3.1 Network model

We focus on the downlink spectrum of a Time Division Multiple Access/Frequency Division Multiple Access (TDMA/FDMA) 5G dense cellular network. This network comprises one MC and *M*

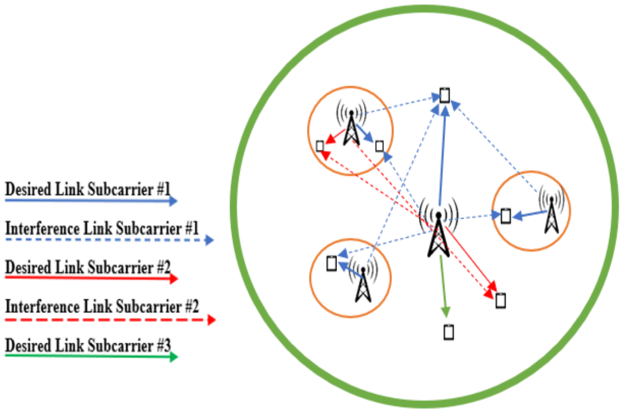
Desired Link Subcarrier #1
Interference Link Subcarrier #1
Desired Link Subcarrier #2
Interference Link Subcarrier #2
Desired Link Subcarrier #3

Figure. 1 System model

SCs where j-th SC is denoted by s_j. Within this network, there are N available cellular resources corresponding to the N subcarriers of an OFDM-based system in a 5G network.

These subcarriers are orthogonal in frequency and experience independent fading. Each SCUE or MCUE can utilize one subcarrier, and each subcarrier is utilized by one SCUE in each SC and one MCUE in the MC. Given the assumption of densely deployed UEs, it is assumed that $M_j \leq N$ UEs are served by each j-th SC and also exactly N UEs are served by each MC. The k-th UE of j-th SC and the i-th UE of the MC are denoted as $s_{j,k}$ and m_i, respectively. Each SCUE and MCUE requests a minimum QoS in terms of rate from the network. The minimum required rate of m_i and the minimum required rate of $s_{j,k}$ are denoted as R_i^{min} and $R_k^{j,min}$, respectively. The transmission power of MCBS on n-th subcarrier and the transmission power of j-th SC on n-th subcarrier are denoted as p^n and p_j^n, respectively. The maximum transmission power of MCBS on each subcarrier and the maximum transmission power of each SCBS on each subcarrier are denoted as p_{MC}^{max} and p_{SC}^{max}, respectively.

To characterize the channel gains within this network, it's essential to note that we assume Rayleigh fading for all channels. It's worth mentioning that each channel gain not only relies on the transmitter and receiver of the specific link but also depends on the cellular resource being used. This is due to the fact that different cellular resources experience independent fading effects, contributing to the dynamic nature of the channel gains across the network. The channel gain value from MCBS to m_i and from the SCBS of s_j to SCUE $s_{j,k}$ on n-th subcarrier are denoted as h_i^n and $h_k^{n,j}$, respectively. These two-channel gains correspond to desired links. The channel gain among the MCBS and SCUE $s_{j,k}$ and among SCBS of s_f and SCUE $s_{j,k}$ on n-th subcarrier when $f \neq j$ are denoted by $g_k^{n,j}$ and $g_k^{n,f,j}$, respectively. The channel gain between SCBS of s_j and MCUE m_i on n-th subcarrier is also denoted by $q_i^{n,j}$. These three channel gains correspond to interference links.

3.2 Problem formulation

In order to formulate the optimization problem, some resource sharing indicators and resource sharing variables should be defined. A resource sharing indicator verifies whether a particular UE is utilizing a specific subcarrier or not, while a resource sharing variable determines if a specific UE is using any subcarrier at all or not. If MCUE m_i is using n-th subcarrier, its corresponding resource sharing indicator denoted as ρ_i^n will be equal to one ($\rho_i^n = 1$) and $\rho_i^n = 0$, otherwise. If $s_{j,k}$ is using n-th subcarrier, its corresponding resource sharing indicator denoted as $\rho_k^{n,j}$ will be equal to one ($\rho_k^{n,j} = 1$) and $\rho_k^{n,j} = 0$, otherwise. The resource sharing variable of MCUE m_i denoted as ζ_i is equal to one if m_i is able to utilize one of the subcarriers and is equal to zero otherwise. The SCUE resource sharing indicator of $s_{j,k}$ denoted as ζ_k^j is equal to one is $s_{j,k}$ is able to utilize one of the subcarriers and is zero otherwise. The resource sharing variable of MCUE m_i and SCUE $s_{j,k}$ can be formulated as

$$\zeta_i = \sum_{n=1}^{N} \rho_i^n \quad (1)$$

and

$$\zeta_k^j = \sum_{n=1}^{N} \rho_k^{n,j} \quad (2)$$

respectively. The MCUE m_i receives a desired signal from the MC on a specific subcarrier and an interference signal from all SCBSs on the same subcarrier. The received power of the desired signal at MCUE m_i ,which is transmitted from the MCBS, can be formulated as

$$T_i = \sum_{n=1}^{N} \rho_i^n \, p^n h_i^n \quad (3)$$

using the resource sharing indicators. The power of the received interference signal at MCUE m_i from all SCBSs can be expressed as:

$$I_i = \sum_{n=1}^{N} \sum_{j=1}^{M} \rho_i^n \, p_j^n q_i^{n,j} \quad (4)$$

using the resource sharing indicators. As a result, the achievable data rate of MCUE m_i denoted by R_i can be written as

$$R_i = \log_2 \left(1 + \frac{T_i}{\sigma^2 + I_i} \right) \quad (5)$$

where σ^2 denotes the variance of additive white Gaussian noise (AWGN).

The SCUE $s_{j,k}$ receives a desired signal on a specific subcarrier from its corresponding SCBS which is s_j. The signals transmitted from all other SCBSs on the same subcarrier are interference signals for this SCUE. In addition to that the signal transmitted from the MCBS on that specific subcarrier is also an interference signal. The received power of the desired signal at SCUE $s_{j,k}$ denoted by T_k^j can be formulated as

$$T_k^j = \sum_{n=1}^{N} \rho_k^{n,j} \, p_j^n \, h_k^{n,j} \tag{6}$$

The interference signal power received at SCUE $s_{j,k}$ caused by SCBSs denoted by $I_{k,2}^j$ can be expressed as

$$I_{k,2}^j = \sum_{n=1}^{N} \sum_{\substack{f=1 \\ f \neq j}}^{M} \rho_k^{n,j} p_f^n \, g_k^{n,f,j} \tag{7}$$

The interference signal power received at SCUE $s_{j,k}$ caused the MCBS denoted by $I_{k,1}^j$ can be written as

$$I_{k,1}^j = \sum_{n=1}^{N} \rho_k^{n,j} p^n \, g_k^{n,j} \tag{8}$$

As a result, the achievable data rate of $s_{j,k}$ denoted by R_k^j can be formulated as

$$R_k^j = \log_2\left(1 + \frac{T_k^j}{\sigma^2 + I_{k,1}^j + I_{k,2}^j}\right) \tag{9}$$

In scenarios characterized by dense UE deployments, the effective mitigation of interference plays a pivotal role in maximizing the sum-rate. The presence of high interference levels significantly constrains the network's capacity. Therefore, when the sum-rate is effectively maximized, it implies that interference has been successfully mitigated. The accomplishment of this sum-rate maximization problem signifies not only the enhancement of overall network performance but also the specific mitigation of interference for CEUEs while adhering to predefined QoS constraints. As a result, we aim to introduce the objective function and the constraint of the problem. The objective function is sum-rate, which is the summation of the achievable data rate of all UE and denoted by R which can be expressed as

$$R = \sum_{i=1}^{N} \zeta_i \, R_i + \sum_{j=1}^{M} \sum_{k=1}^{M_j} \zeta_k^j \, R_k^j \tag{10}$$

using resource sharing variables. The mixed-integer non-convex sum-rate maximization problem can be formulated as

$$\max_{P,\rho} \; R \tag{11}$$

s.t.

$$R_i \geq \zeta_i \; R_i^{min} \qquad \forall i \tag{11a}$$

$$R_k^j \geq \zeta_k^j \, R_k^{j,min} \qquad \forall j, \forall k \tag{11b}$$

$$0 \leq p^n \leq p_{mc}^{\max} \qquad \forall n \tag{11c}$$

$$0 \leq p_j^n \leq p_{sc}^{\max} \qquad \forall n, \forall j \tag{11d}$$

$$\rho_k^{n,j} \in \{0,1\} \qquad \forall n, \forall j, \forall k \tag{11e}$$

$$\rho_i^n \in \{0,1\} \qquad \forall n, \forall i \tag{11f}$$

$$\zeta_i \in \{0,1\} \qquad \forall i \tag{11g}$$

$$\zeta_k^j \in \{0,1\} \qquad \forall k, \forall j \tag{11h}$$

where equations (11a) and (11b) correspond to the minimum QoS constraint of MCUE m_i and SCUE $s_{j,k}$, respectively, in constraints (11a) and (11b), resource sharing variables play a crucial role. These variables are utilized to ensure that the achievable rate of each UE exceeds its minimum required QoS if the UE is actively utilizing any subcarrier ($\zeta_i = 1$ or $\zeta_k^j = 1$). If a UE is unable to utilize any subcarrier, its achievable rate naturally becomes zero, and the constraints are satisfied accordingly. These resource sharing variables help in modelling the QoS requirements of UEs effectively within the optimization framework. Constraints (11c) and (11d) represent the minimum and maximum transmission powers on all subcarriers for the MCBS and SCBSs, respectively. Constraints (11e) and (11f) represent the binary structure of resource sharing indicators. Constraints (11g) and (11h) express that resource sharing variables are also binary, meaning that each SCUE or MCUE can utilize at most one subcarrier. The optimization variables are the set of all transmission powers denoted by P and the set of all resource sharing indicators denoted by ρ.

Table 1. Notations

Symbol	Description	Symbol	Description
N	The number of OFDM subcarriers	I_i	The received interference signal at m_i from the all SCBSs
N_t	The total number of UEs in the network	T_k^j	The received power of the desired signal at $s_{j,k}$
M	The number of SCs	$I_{k,1}^j$	The interference signal power received at SCUE $s_{j,k}$from MCBS
s_j	The index of SC	$I_{k,2}^j$	The interference signal power received at SCUE $s_{j,k}$ caused by SCBSs
M_j	The number of UEs are served by each s_j	R_k^j	The achievable data rate of $s_{j,k}$
$s_{j,k}$	The k-th UE of j-th SC	R	The achievable data rate of all UEs
m_i	The i-th UE of MC	$\tilde{I}_n$	The received interference from SCBSs to the MCUE on n-th subcarrier
R_i^{min}	The minimum rate requirement of m_i	$\tilde{R}_n$	the achievable rate of the MCUE on n-th subcarrier
$R_k^{j,min}$	The minimum rate requirement of $s_{j,k}$	$\tilde{T}_n^l$	The received power of the desired signal from its corresponding SCBS to the SCUE on n-th subcarrier
p^n	The transmission power of MCBS on n-th subcarrier	$\tilde{I}_{n,2}^l$	The received interference from MCBS to the SCUE on n-th subcarrier
p_j^n	The transmission power of j-th SC on n-th subcarrier	$\tilde{I}_{n,1}^l$	The received interference from other SCBSs to the SCUE on n-th subcarrier
p_{MC}^{max}	The maximum transmission power of MCBS.	$\tilde{R}_n^l$	The achievable data rate of SCUE on n-th subcarrier
p_{SC}^{max}	The maximum transmission power of each SCBS.	R_n^t	The sum rate of all UEs utilize subcarrier n
h_i^n	The channel gain from MCBS to m_i on n-th subcarrier	$\tilde{p}^n$	The vector of powers of UEs utilize subcarrier n
$h_k^{n,j}$	The channel gain value from the SCBS of s_j to SCUE $s_{j,k}$	$\gamma_n^{\min}$	The minimum SINR for MCUEs
$g_k^{n,j}$	The channel gain among the MCBS and SCUE $s_{j,k}$ on n-th subcarrier	$\gamma_n^{l,min}$	The minimum SINR for SCUEs
$g_k^{n,f,j}$	The channel gain among SCBS of s_f and SCUE $s_{j,k}$when $f \neq j$ on n-th subcarrier	a_i	The SDV of MCUE m_i
$q_i^{n,j}$.	The channel gain between SCBS of s_j and MCUE m_i on n-th subcarrier	a_k^j	The SDV of SCUE s_k^j
ρ_i^n	. The resource sharing indicator of m_i on n-th subcarrier	L^n	The number of SCUEs that are reusing subcarrier n
ζ_i	The resource sharing variable of m_i	b_1^n	The SCII of n -th subcarrier which corresponds to the index of the SC of the UE that utilizes n -th subcarrier
$\rho_k^{n,j}$	The resource sharing indicator of $s_{j,k}$ on n-th subcarrier	b_2^n	The UEII which corresponds to index the UE that utilizes n -th subcarrier inside its own SC
ζ_k^j	The resource sharing variable $s_{j,k}$	b_3^n	The MCUEII of n -th subcarrier which corresponds to the MCUE that utilizes n -th subcarrier
T_i	The received power of the desired signal at m_i	σ^2	The variance of AWGN

4. Proposed scheme

In this section, we present our effective and novel proposed scheme. Firstly, the idea and the approach are discussed. After that, the proposed learning method for finding the matching among UEs and cellular resources is described. Then, the optimal power allocation based on optimization-based approaches is presented. Finally, overall resource allocation algorithm is described in an algorithmic manner. Table 1 includes the notations used in this research paper.

4.1 Approach and general discussion

The current problem is a mixed-integer nonlinear optimization problem, with the binary variables representing resource sharing indicators. These indicators essentially determine which UE should utilize each subcarrier, forming a matching between UEs and available cellular resources. Our proposed resource allocation scheme consists of two steps: the first step focuses on finding the optimal matching between UEs and cellular resources, which involves determining all resource sharing indicator values. In the second step, we assume a specific matching exists, and the goal is to allocate power to all UEs, essentially addressing power allocation.

For the first step, we employ a RL-based approach using Q-learning. This step selects a matching between UEs and cellular resources and passes it to the second step. By utilizing the output of the second step, which is the sum-rate of all UEs, the learning step adapts and learns which cellular resources are better suited for specific UEs. Through multiple iterations, the learning algorithm converges to a near-optimal matching. The learning approach goes through different phases, including exploration and exploitation, which will be discussed in greater detail

In the second step, once the matching is known, the mixed-integer nonlinear problem can be transformed into a nonlinear problem, as all binary variables become known and can be eliminated. We reformulate the remaining nonlinear constraints into affine ones and employ the difference of concave form of the objective function to derive an upper bound, facilitating the optimal solution through an iterative approach. To solve the power allocation problem, we address the power allocation for UEs on each subcarrier in parallel. We consider two different cases: case one corresponds to a scenario where one of the MCUEs is using the subcarrier, while case two assumes that none of the MCUEs are utilizing that specific subcarrier. These distinct cases are essential for optimizing the power allocation across the network.

4.2 Learning-based matching among uEs and cellular resources

In this step, we present our proposed Q-learning approach to identify a sub-optimal matching among UEs and cellular resources. Q-learning is a widely utilized reinforcement learning algorithm employed for solving problems where agents interact with an environment to learn optimal actions in order to maximize cumulative rewards. In our context, each UE functions as an agent capable of selecting its action, representing its choice of subcarrier. The environment in our Q-learning framework represents the cellular network within which the Q-learning algorithm operates to learn optimal actions for UEs and cellular resource allocation. This environment encapsulates the network's dynamics, interference patterns, resource availability, and the interactions between UEs and base stations, allowing the Q-learning algorithm to adapt and make informed decisions to maximize the network's performance, particularly in terms of sum-rate and interference mitigation for CEUEs. However, due to the constraint that UEs within the same SC or MC cannot share the same subcarrier, we need to make certain adjustments regarding the reward and actions.

The fundamental concept behind Q-learning involves the utilization of a Q-matrix to denote the quality or utility of taking a specific action in a particular state. To implement this, the Q-matrix needs to be designed appropriately, and the state structure of UEs should be carefully defined. Each UE has $(N+1)$ possible actions, where actions 1 to N correspond to selecting subcarriers 1 to N, while action $(N+1)$ signifies that the UE is not using any of the subcarriers, implying it will not receive service from the network. This situation often arises when a UE requests a QoS from the network that exceeds the network's capacity. The total number of UEs in the network is N_t and can be expressed as

$$N_t = N + \sum_{j=1}^{M} M_j \qquad (12)$$

which represents that N of UEs are MCUEs and M_j of them are the UEs of the j-th SC. Consequently, we define the current state vector as $S \in B^{N_t \times 1}$ which can be formulated as

$$S = \{S_1, \dots, S_N, S_{1,1}, \dots, S_{1,M_1}, \dots, S_{M,M_M}\}$$

where S_i is the state of MCUE m_i and $S_{i,j}$ is the state of SCUE $s_{i,j}$. Each element of S corresponds to the state of one of the UEs and can be an integer value from the set $B = \{1, 2, \dots, N+1\}$ indicating the subcarrier used by that UE. This state representation enables us to apply Q-learning to find the sub-optimal matching effectively. Initially, all values of the Q-Matrix are set to zero.

In our Q-learning framework, the agent interacts with the environment by selecting actions based on the current state. To train the Q-learning algorithm and derive the final Q-matrix, we divide the learning process into N_e episodes, each consisting of N_s steps. During each step, all agents employ exploration strategies, such as epsilon-greedy, to determine their actions. ϵ -greedy entails exploring different actions with a certain probability (exploration) and selecting actions with the highest Q-values with the remaining probability (exploitation). We control the exploration-exploitation trade-off with the parameter ϵ, which initially sets the probability of exploration. For a predefined number of episodes, ϵ remains constant, indicating that the exploration probability does not change significantly, and agents select actions randomly with that probability. After this phase, we gradually reduce the ϵ value after each episode until it reaches a minimum threshold ϵ_{min}. In this later phase, the probability of exploration decreases, and agents rely more on the learned Q-values. It's important to note that the ϵ value does not reach zero, allowing for a slight degree of randomness even as the number of episodes increases. This approach ensures a balance between exploration and exploitation throughout the training process.

In our Q-learning framework, we employ a Q-matrix with $N_R = N^{N_t}$ rows and $N_C = N_t$ columns, where N_R represents the maximum number of possible states and N corresponds to the number of possible actions available to each agent. It's important to note that not all states are feasible due to the constraints of the optimization problem. Consequently, the Q-values for rows corresponding to infeasible states are initialized to the minimum possible Q-value, which is zero. These values are then held constant throughout the algorithm's execution. To determine the feasibility of a state, we introduce the function $f(S)$, which evaluates whether a given state S is possible in terms of resource sharing indicators. The function can be expressed as:

$$f(S) = \begin{cases} 1, & \text{if } S \text{ is feasible} \\ 0, & \text{if } S \text{ is not feasible}' \end{cases}$$

where a value of one indicates that the input state S is feasible based on resource sharing indicators, while a value of zero signifies that the state is infeasible. This function plays a crucial role in determining which states are valid and helps guide the Q-learning process by focusing on feasible state-action pairs.

In our Q-learning framework, each agent, representing a UE, makes decisions on which action to take in the current state based on the exploration-exploitation strategy. Once an action is chosen, the agent carries it out by requesting the network to change its subcarrier according to the action, leading to a transition to the next state. After performing the selected action and observing the resulting state and the reward received, the Q-values in the Q-table are updated following the Q-learning update rule. In this context, the reward corresponds to the achieved sum-rate, which is determined during the second step of our proposed scheme, specifically the power allocation step. This reward reflects the effectiveness of the chosen action in maximizing the sum-rate of UEs and is used to guide the learning process.

In our Q-learning framework, the update rule plays a pivotal role in the learning process. By denoting the current action as A, the update rule is formulated as

$$Q_{new}(S, A) = \\ Q(S, A) + \alpha \, [\, R_e + \gamma \, S^M - Q(S, A)] \tag{13}$$

where $Q_{new}(S, A)$represents the updated Q value, $Q(S, A)$ represents the Q-value for the current state-action pair, which we aim to update A. The parameters α and γ denote the learning rate and the discount factor, respectively. The learning rate modulates the extent of Q-value updates during each step. The discount factor weighs the importance of future rewards. It ranges from 0 to 1 and adjusts for the trade-off between immediate and long-term rewards. R_e is the immediate reward received after taking a specific action in the current state and corresponds to the achievable sum-rate of the optimal power allocation step. The value S^M represents the maximum Q-value among all possible actions in the next state and can be expressed as

$$S^M = \max_{\text{A}} Q(S_n, A) \tag{14}$$

where S_n is the next state. The Q-value update rule essentially recalibrates the Q-value for the current state-action pair based on the observed reward and the estimate of the maximum future expected reward. The agent continues to interact with the environment iteratively, selecting actions and updating Q-values. The ultimate goal is to determine the optimal policy, which consists of a set of actions that maximize the expected cumulative reward over time. Through ample exploration and learning, the Q-values gradually converge toward their optimal values, which signify the best actions to take in each state to maximize cumulative rewards. The reason behind updating each Q-value with the maximum value from the next state lies in the principle of optimizing future expected rewards. This approach ensures that the agent learns to make decisions that maximize its long-term reward. This fundamental aspect of Q-learning encourages the agent to explore and exploit actions that lead to higher expected rewards in the future, ultimately driving the convergence toward an optimal policy.

4.3 Optimal power allocation

Following the successful execution of the matching process in the first step, the resource sharing indicators are determined using the proposed Q-learning algorithm. Consequently, the binary variables, which represent the matching of UEs to available cellular resources, can be eliminated from the optimization problem. This simplifies the problem to the form

$$\max_{P,\rho} \; R \tag{15}$$

s.t.

$$R_i \;\geq\; \zeta_i \; R_i^{min} \qquad \forall i \tag{15a}$$

$$R_k^j \geq \zeta_k^j \, R_k^{j,min} \qquad \forall j, \forall k \tag{15b}$$

$$0 \leq p^n \leq p_{mc}^{\max} \qquad \forall n \tag{15c}$$

$$0 \leq p_j^n \leq p_{sc}^{\max} \qquad \forall n, \forall j \tag{15d}$$

where the optimization variables exclusively involve the transmission powers of MCBS and SCBSs on different subcarriers. The second step in the process of resource allocation revolves around power allocation and aims to solve it in an optimal manner. To address this, we can tackle the problem in parallel, as the presence of each UE utilizing a specific subcarrier does not introduce interference to or receive interference from UEs using other subcarriers. Consequently, the power allocation problem for each subcarrier can be treated independently, effectively transforming the overall power allocation problem into N parallel optimization subproblems, each corresponding to a distinct subcarrier. This approach streamlines the optimization process and facilitates efficient power allocation across the network. In light of this approach, we introduce new subcarrier demonstrator variables (SDV), where each variable associates a MCUE or a SCUE with the its corresponding subcarrier. Specifically, we denote the SDV of MCUE m_i as a_i , representing the subcarrier employed by the MCUE. If the MCUE cannot utilize any subcarriers, then $a_i = 0$. Similarly, we define the SDV of SCUE s_k^j as a_k^j, indicating the subcarrier used by the SCUE. In a case where the SCUE cannot access any subcarriers, $a_k^j = 0$.

These SDVs help streamline the representation of subcarrier allocation and usage within the network optimization framework. SDVs bear a close relationship to the resource sharing indicators used in the optimization problem. Specifically, we can express the presence or absence of non-zero subcarrier demonstrator variables for MCUEs as:

$$0 = a_i \rightarrow \rho_i^n = 0 \rightarrow \zeta_i = 0 \tag{16}$$

And

$$0 < a_i \leq N \;\rightarrow \begin{cases} \rho_i^{a_i} = 1 \\ \rho_i^n = 0 \end{cases} \rightarrow \zeta_i = 1 \tag{17}$$

respectively. Similarly, for SCUEs, we can represent the zero and non-zero subcarrier demonstrator variables as:

$$a_k^j = 0 \rightarrow \rho_k^{j,n} = 0 \rightarrow \; \zeta_k^j = 0 \tag{18}$$

And

$$0 < a_k^j \leq N \;\rightarrow \begin{cases} \rho_k^{j,a_j^k} = 1 \\ \rho_k^{j,n} = 0 \end{cases} \rightarrow \zeta_k^j = 1 \tag{19}$$

respectively. These associations between SDVs and the optimization parameters streamline the mathematical representation of the problem and its constraints, aiding in the efficient solution of the power allocation problem in the second step of our proposed approach. To transform the power allocation problem into parallel optimizations, we need to formulate two distinct cases separately. The first case arises when a particular subcarrier is actively utilized by a specific MCUE, and this can be mathematically represented as

$$\exists \mathrm{i} \in \{1, \dots, \mathrm{N}\}, \qquad \rho_i^n = 1 \rightarrow a_i \neq 0 \qquad (20)$$

The second case pertains to situations where a specific subcarrier remains unallocated and is not being used by any MCUEs, and this can be expressed mathematically as

$$\forall i \in \{1, \dots, N\} \rightarrow \rho_i^n = 0 \rightarrow a_i = 0 \qquad (21)$$

By segregating these cases, we can effectively address the power allocation problem in a parallel manner, simplifying the optimization process and ensuring efficient resource utilization in our proposed approach. To enhance the readability and clarity of our proposed scheme for readers, we will discuss our approach for the first and second cases in separate subsections. This organizational structure will allow us to provide a comprehensive and detailed explanation of our methodology for each case, ensuring that readers can easily follow and understand the intricacies of our proposed solution.

4.4 Optimal power allocation for case 1

In this specific scenario, multiple SCUEs from different SCs are concurrently utilizing the same subcarrier (assume subcarrier n), where this subcarrier is also assigned to a particular MCUE. To effectively express the mathematical formulations for this situation, we introduce three essential indicator variables: the SC index indicator (SCII), the UE index indicator (UEII), and the MCUE index indicator (MCUEII). Let's denote the number of SCUEs that are reusing subcarrier n as L^n. The SCII represented as $b_1^n \in D^{L^n \times 1}$ depends on the subcarrier number and yields a vector of length L^n, where $D = \{0, 1, \dots, N\}$. The l-th element of b_1^n corresponds to the SC number of the l-th UE reusing subcarrier n.

Similarly, the UEII, denoted as $b_2^n \in D^{L^n \times 1}$ also depends on subcarrier n and results in a vector of length L^n. The l-th element of this vector represents the UE number within the $b_1^n(l)$-th SC that also utilizes the shared subcarrier. The MCUEII represented as $b_3^n \in D^{1 \times 1}$ depends on subcarrier n and corresponds to the number of the MCUE that is utilizing subcarrier n.The received power of the desired signal from the MCBS to the MCUE that is utilizing subcarrier n can be expressed as

$$\tilde{T}_n = T_{b_3^n} = p^n \tilde{h}_n \qquad (22)$$

where $\tilde{h}_n$ is the channel gain from MCBS to the MCUE. The interference caused by SCBSs to the MCUE on subcarrier n can also be formulated as

$$\tilde{I}_n = \sum_{l=1}^{L^n} \tilde{p}_l^n \tilde{q}_l^n \qquad (23)$$

where

$$\tilde{p}_l^n = p_{b_2^n(l)}^n \qquad (24)$$

$$\tilde{q}_l^n = q_{b_3^n}^{n, b_2^n(l)} \qquad (25)$$

Hence, the achievable rate of the MCUE can be expressed as

$$\tilde{R}_n = \log_2(1 + \frac{\tilde{T}_n}{\sigma^2 + \tilde{I}_n}) \qquad (26)$$

Considering l-th SCUE, which is using subcarrier n, the received power of the desired signal from its corresponding SCBS can also be expressed as

$$\tilde{T}_n^l = T_{b_{1(l)}^n}^{b_2^n(l)} = \tilde{p}_l^n \tilde{h}_n^l \qquad (27)$$

where

$$\tilde{h}_n^l = h_k^{n,j} \qquad (28)$$

The received power of the interference signal received at l-th SCUE which uses subcarrier n from the MCBS on subcarrier n can also be expressed as

$$\tilde{I}_{n,1}^l = p^n \tilde{g}_n^l \qquad (29)$$

Where

$$\tilde{g}_n^l = g_{b_1^n(l)}^{n, b_2^n(l)} \qquad (30)$$

The received power of the interference signal from other SCBSs that are using subcarrier n to the l-th SCUE which is using subcarrier n can also be expressed as

$$\tilde{I}_{n,2}^l = \sum_{\substack{f=1 \\ f \neq l}}^{L^n} \tilde{p}_f^n \tilde{g}_n^{l,f} \qquad (31)$$

Where

$$\tilde{g}_n^{l,f} = g_{b_n^1(l)}^{n,f,b_n^2(l)} \qquad (32)$$

Hence, the achievable data rate of l-th SCUE which is using subcarrier n can be expressed as

$$\tilde{R}_n^l = \log_2(1 + \frac{\tilde{T}_n^l}{\sigma^2 + \tilde{I}_{n,1}^l + \tilde{I}_{n,2}^l}). \tag{33}$$

As each UE utilizing a specific subcarrier contributes to interference for other UEs sharing the same subcarrier, the overall power allocation problem can be effectively decomposed into n parallel subproblems, each dedicated to a specific subcarrier. Consequently, the power allocation for each subcarrier can be formulated as a separate optimization problem, allowing for a more focused and tractable approach, and can be formulated as

$$\underset{\tilde{P}^n}{\mathrm{Max}}\, \tilde{R}_n + \sum_{l=1}^{L^n} \tilde{R}_n^l \tag{34}$$

s.t.

$$\tilde{R}_n \geq \tilde{R}_n^{min} \tag{34a}$$

$$\tilde{R}_n^l \geq \tilde{R}_n^{l,min}, \forall l \tag{34b}$$

$$0 \leq \tilde{p}^n \leq p_{mc}^{\max} \tag{34c}$$

$$0 \leq \tilde{p}_l^n \leq p_{sc}^{\max} \quad \forall l \tag{34d}$$

$$\tilde{R}_n^{min} = \tilde{R}_{b_n^3}^{min} \tag{35}$$

$$\tilde{R}_n^{l,min} = R_{b_n^1(l)}^{b_n^2(l),min} \tag{36}$$

where $\tilde{P}^n = \{\tilde{p}^n, \tilde{p}_1^n, \dots, \tilde{p}_{L_n}^n\}$. Constraints (34a) and (34b) are QoS constraints of the MCUE and SCUEs utilizing subcarrier n, respectively. Constraints (34c) and (34d) are the transmission power limits of the MCUE and SCUEs utilizing subcarrier n, respectively.

The QoS constraints pose a notable challenge in the optimization problem, as they are inherently non-convex and complex to handle. Specifically, the QoS constraints for the MCUE and l-th SCUE utilizing subcarrier n can be formulated as

$$\log_2\left(1 + \frac{\tilde{T}_n}{\sigma^2 + \tilde{I}_n}\right) \geq \tilde{R}_n^{min} \tag{37}$$

And

$$\log_2(1 + \frac{\tilde{T}_n^l}{\sigma^2 + \tilde{I}_{n,1}^l + \tilde{I}_{n,2}^l}) \geq \tilde{R}_n^{l,min} \tag{38}$$

respectively. These minimum QoS requirements can be expressed in terms of achievable data rate, reflecting the minimum data rate that must be ensured. Conversely, the same QoS requirement can also be characterized in terms of SINR, serving as an alternative but equivalent expression for QoS, albeit in different terminology. Therefore, the QoS specifications for both the MCUE and SCUEs using subcarrier n can be expressed as

$$\frac{p^n \tilde{h}_n}{\sigma^2 + \sum_{l=1}^{L^n} \tilde{p}_l^n \tilde{q}_l^n} \geq \gamma_n^{\min} \tag{39}$$

and

$$\frac{\tilde{p}_l^n \tilde{h}_n^l}{\sigma^2 + p^n \tilde{g}_n^l + \sum_{\substack{f=1 \\ f \neq l}}^{L^n} \tilde{p}_f^n \tilde{g}_n^{l,f}} \geq \gamma_n^{l,min} \quad \forall l \tag{40}$$

respectively. However, these expressions are still non-convex. Using a simple transformation, these constraints can be transformed into convex constraints. As a result, the QoS constraints for the MCUE and l-th SCUE utilizing subcarrier n can be expressed in a convex form as

$$p^n \tilde{h}_n \geq \gamma_n^{\min} \sigma^2 + \gamma_n^{\min} \sum_{l=1}^{L^n} \tilde{p}_l^n \tilde{q}_l^n \tag{41}$$

and

$$\tilde{p}_l^n \tilde{h}_n^l \geq \gamma_n^{l,min} \sigma^2 + \gamma_n^{l,min} p^n \tilde{g}_n^l + \gamma_n^{l,min} \sum_{\substack{f=1 \\ f \neq l}}^{L^n} \tilde{p}_f^n \tilde{g}_n^{l,f} \quad \forall l \tag{42}$$

respectively. At this stage of the optimization process, the power allocation problem for each subcarrier n has convex constraints, yet the objective function remains non-convex. To address this challenge and make the objective function more amenable to optimization, we employ a specific form of rate function. The rate of a UE can be expressed as the difference between two concave functions. Specifically, considering the rate of the MCUE, it can be represented as

$$\tilde{R}_n = f_n^1 - f_n^2 \tag{43}$$

where both f_n^1 and f_n^2 are concave functions and can be expressed as

$$f_n^1 = \log_2\left(\sigma^2 + \tilde{I}_n + \tilde{T}_n\right) \tag{44}$$

And

$$f_n^2 = \log_2(\sigma^2 + \tilde{I}_n) \tag{45}$$

respectively. Similarly, the rate of a l-th SCUE using subcarrier n can be formulated as

$$\tilde{R}_n^l = f_{n,l}^3 - f_{n,l}^4 \quad (46)$$

where $f_{n,l}^3$ and $f_{n,l}^4$ are also concave functions and can be written as

$$f_{n,l}^3 = \log_2(\sigma^2 + \tilde{I}_{n,1}^l + \tilde{I}_{n,2}^l + \tilde{T}_n^l) \quad (47)$$

and

$$f_{n,l}^4 = \log_2\left(\sigma^2 + \tilde{I}_{n,1}^l + \tilde{I}_{n,2}^l\right) \quad (48)$$

respectively. Consequently, the sum-rate of all UEs utilizing subcarrier n can be expressed as

$$R_n^t = f_n^1 - f_n^2 + \sum_{l=1}^{L^n} f_{n,l}^3 - f_{n,l}^4 \quad (49)$$

which can be further simplified as

$$R_n^t = f_n^1 + \sum_{l=1}^{L^n} f_{n,l}^3 - (f_n^2 + \sum_{l=1}^{L^n} f_{n,l}^4) \quad (50)$$

The overall sum-rate can then be expressed as

$$R^t = f_n^5 - f_n^6 \quad (51)$$

where F5 represents the summation of all concave terms with positive signs, constituting another convex function, and F6 corresponds to the summation of all concave terms with negative signs, also forming a concave function which is formulated as

$$f_n^5 = f_n^1 + \sum_{l=1}^{L^n} f_{n,l}^3 \quad (52)$$

and

$$f_n^6 = (f_n^2 + \sum_{l=1}^{L^n} f_{n,l}^4) \quad (53)$$

respectively. Therefore, the sum-rate expression ultimately boils down to the difference between two concave functions, facilitating a more tractable objective for optimization. By utilizing the concept of difference of concave functions, we can derive a lower bound for the objective function by applying a first-order Taylor expansion around an initial point. This lower bound is obtained by preserving the concave function f_n^5 and approximating the convex function $(-f_n^6)$ as a linear function. The $(-f_n^6)$ function can be effectively approximated by a linear function through Taylor expansion. Importantly, a line derived from this Taylor expansion serves as a strict minimizer for the $(-f_n^6)$ function due to its convex form. Consequently, the lower bound can be expressed as

$$R_n^t\left(\tilde{P}^n\right) \geq f_n^5\left(\tilde{P}^n\right) + \nabla f_n^6\left(\tilde{P}_0^n\right)\left(\tilde{P}^n - \tilde{P}_0^n\right) \quad (54)$$

where $\tilde{P}^n$ is a vector of powers of UEs utilizing subcarrier n. $\tilde{P}_0^n$ is an initial point representing a specific amount of power values of UEs utilizing subcarrier n. $\nabla f_n^6\left(\tilde{P}_0^n\right)$ is the gradient of the f_n^6 and point $\tilde{P}^n$ which can be computed effectively since all the functions are analytically described. This lower bound facilitates the optimization process by providing an approximation of the objective function that is more amenable to mathematical optimization techniques. The application of the lower bounds for each point in the optimization process enables the utilization of MaMi technique. This approach is particularly valuable when dealing with optimization problems featuring non-convex objective functions. By employing MaMi, a maximization problem with a non-convex objective function can be systematically and optimally solved through iterative steps. The key insight lies in estimating a lower bound function at each iteration. Since we have derived a general lower bound at each step, we can effectively maximize the objective function while adhering to the constraints. This maximization yields the next point corresponding to the optimal solution within the current iteration. Subsequently, this newly found point is utilized to generate another lower bound, initiating another iteration in the optimization process. This iterative cycle continues until the problem is ultimately solved, converging towards an optimal solution. By integrating the methodologies described above, the complex sum-rate optimization problem can be effectively transformed into a parallel form, where individual subcarriers are addressed as separate, parallel optimization problems. These parallel problems feature non-convex constraints, which, through meticulous mathematical reformulation, can be converted into convex constraints. Furthermore, the objective function itself can be elegantly expressed as the difference between two concave functions. The critical component of this approach is the derivation of a lower bound for the objective function, facilitating the application of the MaMi technique. By leveraging this technique, the optimization problem can be tackled iteratively utilizing the lower bound. Consequently, each step of the optimization problem is transformed into a convex problem written as

$$\max_{\tilde{P}^n} f_n^5\left(\tilde{P}^n\right) + \nabla f_n^6\left(\tilde{P}_0^n\right)\left(\tilde{P}^n - \tilde{P}_0^n\right) \quad (55)$$

s.t.

$$34c, 34d, 41, 42$$

which can be effectively and efficiently solved using interior-point methods. This systematic

Received: February 13, 2024. Revised: April 7, 2024.

approach ensures that the challenging sum-rate optimization problem can be efficiently addressed for each subcarrier in a dense cellular network.

4.5 Optimal power allocation for case 2

In the second case, we encounter a scenario where the MCUE is unable to receive any service, resulting in its transmission power being set to zero. This situation arises when the MCUE demands a minimum QoS that surpasses the network's capacity, leading the BS to make the decision not to serve it. In this particular case, the mathematical formulations and expressions undergo slight adjustments to accommodate this unique scenario, which we will delve into to ensure a comprehensive understanding of the power allocation problem. Let's consider the scenario where we aim to analyze the UEs that are utilizing subcarrier n. In this specific case, no MCUEs are utilizing the subcarrier, which can be mathematically expressed as

$$\forall i \in \{1, \dots, N\} \rightarrow \rho_i^n = 0 \rightarrow a_i = 0 \tag{56}$$

As a consequence of this configuration, the received desired signal power for all MCUEs becomes zero. Consequently, the achievable rate for all MCUEs is also reduced to zero, formulated as

$$\tilde{T}_n = \tilde{R}_n = 0 \tag{57}$$

This situation forms a unique subset within the power allocation problem, which we will address separately to provide a clear understanding of the mathematical expressions and calculations involved.

Within this assumption, the SCUEs do not encounter interference from any MCUE. This is because the transmission power of the MC on subcarrier n has been reduced to zero, as no MCUEs are utilizing that specific subcarrier. Consequently, the desired signal power received at the l-th SCUE using subcarrier n can be denoted as

$$\tilde{T}_n^l = T_{b_{1(l)}^n}^{b_2^n(l)} = \tilde{p}_l^n \tilde{h}_n^l \tag{58}$$

The interference originating from the MC and the interference originating from other SCs can be expressed as

$$\tilde{I}_{n,2}^l = \sum_{\substack{f=1 \\ f \neq l}}^{L} \tilde{p}_f^n \tilde{g}_n^{l,f} \tag{59}$$

and

$$\tilde{I}_{n,1}^l = 0 \tag{60}$$

respectively. As a result, the achievable data rate of l-th SCUE on subcarrier n can be formulated as

$$\tilde{R}_n^l = \log_2\left(1 + \frac{\tilde{T}_n^l}{\sigma^2 + \tilde{I}_{n,2}^l}\right) \tag{61}$$

Similar to the previous step, the overall power allocation problem can be effectively decomposed into n parallel subproblems, each dedicated to a specific subcarrier. The power allocation problem for subcarrier n can be formulated as

$$\max_{\tilde{P}^n} \sum_{l=1}^{L^n} \tilde{R}_n^l \tag{62}$$

s.t.

$$\tilde{R}_n^l \geq \tilde{R}_n^{l,min} \quad \forall l \tag{62a}$$

$$0 \leq \tilde{p}_l^n \leq p_{sc}^{\max} \quad \forall l \tag{62b}$$

Where $\tilde{P}^n = \{\tilde{p}_1^n, \dots, \tilde{p}_{L_n}^n\}$. The QoS constraints are also non-convex and pose a notable challenge in the optimization problem. The QoS constraints for the l-th SCUE utilizing subcarrier n can be formulated as

$$\log_2\left(1 + \frac{\tilde{T}_n^l}{\sigma^2 + \tilde{I}_{n,2}^l}\right) \geq \tilde{R}_n^{l,min} \tag{63}$$

Similar to the previous section, the QoS requirement can also be characterized in terms of SINR, serving as an alternative but equivalent expression for QoS. Therefore, the QoS specifications for l-th SCUE using subcarrier n can be expressed as

$$\frac{\tilde{p}_l^n \tilde{h}_n^l}{\sigma^2 + \sum_{\substack{f=1 \\ f \neq l}}^{L^n} \tilde{p}_f^n \tilde{g}_n^{l,f}} \geq \gamma_n^{l,min} \quad \forall l \tag{64}$$

This non-convex expression can transform into a convex form expressed as

$$\tilde{p}_l^n \tilde{h}_n^l \geq \gamma_n^{l,min} \sigma^2 + \gamma_n^{l,min} \sum_{\substack{f=1 \\ f \neq l}}^{L^n} \tilde{p}_f^n \tilde{g}_n^{l,f} \quad \forall l \tag{65}$$

In the scenario where no MCUE is utilizing any subcarriers, the constraints are currently expressed in a convex form. However, the objective function remains non-convex and can be expressed as the difference between two convex functions.

Consequently, the rate of the l-th SCUE in this case can be formulated as

$$\tilde{R}_n^l = f_{n,l}^7 - f_{n,l}^8 \tag{66}$$

where both $f_{n,l}^7$ and $f_{n,l}^8$ are convex functions and can be expressed as

$$f_{n,l}^7 = \log_2(\sigma^2 + \tilde{I}_{n,2}^l + \tilde{T}_n^l) \tag{67}$$

and

$$f_{n,l}^8 = \log_2(\sigma^2 + \tilde{I}_{n,2}^l) \tag{68}$$

respectively. Thus, the sum-rate of UEs utilizing subcarrier n denoted as R_n^t can be expressed as

$$\tilde{R}_n^l = f_n^9 - f_n^{10} \tag{69}$$

where f_n^9 represents the summation of all convex functions with positive signs, forming another convex function, and f_n^{10}corresponds to the summation of all convex functions with negative signs, also resulting in a convex function which can be formulated

$$f_n^9 = \sum_{l=1}^{L^n} f_{n,l}^7 \tag{70}$$

And

$$f_n^{10} = \sum_{l=1}^{L^n} f_{n,l}^8 \tag{71}$$

respectively. Applying a similar rationale to that presented in the preceding section, an upper bound for the sum-rate of the UEs utilizing subcarrier n can be derived, expressed as

$$R_n^t(\tilde{P}^n) \geq f_n^9(\tilde{P}^n) + \nabla f_n^{10}(\tilde{P}_0^n)(\tilde{P}^n - \tilde{P}_0^n) \tag{72}$$

where ∇f_n^{10} represents the gradient of the convex function f_n^{10} . Employing analogous explanations and motivations as in the previous scenario, the MaMi technique can be effectively employed to optimally solve the power allocation problem iteratively. Specifically, each iteration within the MaMi technique corresponds to a maximization problem, written as

$$\max_{\tilde{P}^n} f_n^9(\tilde{P}^n) + \nabla f_n^{10}(\tilde{P}_0^n)(\tilde{P}^n - \tilde{P}_0^n) \tag{73}$$

s.t.

$$62b, 65$$

which is inherently a convex optimization problem. Consequently, this convex problem can be efficiently solved using interior-point methods. This iterative approach ensures the optimal allocation of power to UEs on a specific subcarrier, effectively managing interference and enhancing the overall network's performance.

Wherever Times is specified, Times Roman of Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 fonts are preferred.

5. Numerical results

We assume a dense heterogeneous B5G network consisting of one MC and multiple SCs in the network. There exist multiple UEs in each SC. Rayleigh fading with log-normal slow fading and unit mean exponentially distributed fast fading is assumed. Table.1 introduces the parameters of the network as well as that of the fading model.

In our comprehensive evaluation, we conducted an extensive comparison between various aspects of our proposed scheme and state-of-the-art methods, clearly demonstrating the superior performance of our approach. To provide context for our evaluations,

We employed a well-defined system model to simulate a network configuration. This network consisted of an MC, two SCs, CEUEs, CCUE, and SCUEs. This model served as the foundation for our performance assessments, allowing us to draw meaningful comparisons and draw conclusions about the effectiveness of our proposed solution. Fig. 2 serves as a visual representation of the effectiveness of our proposed Q-learning method. In the initial episodes of the learning procedure, random matching between UEs and subcarriers is conducted, resulting in a sum-rate with an alternating pattern. However, as more episodes are

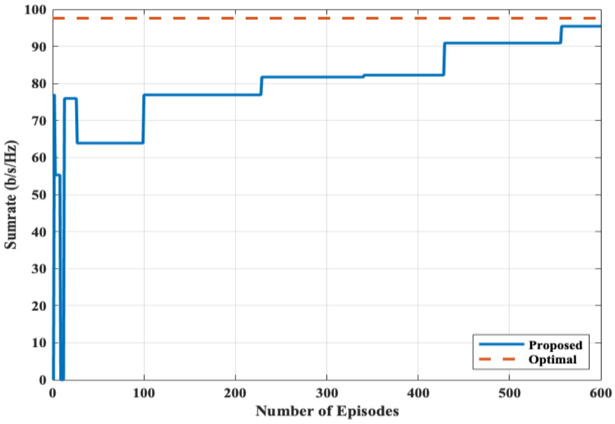

Figure. 2 Convergence steps of the proposed learning scheme toward the optimal matching among UEs and available cellular resources

Table 2. System Parameters

Parameter	Value
Physical Type	Downlink
Cell Radius	500 (m)
Number of Subcarriers	4
Number of CEUEs	2
Number of CCUEs	2
Center frequency	4.7 GHz
Bandwidth	$150 KHz$
Path-loss exponent	4
Path-loss constant	10^{-2}
Shadowing standard deviation	6 dB
Noise spectral density	-174 dBm/Hz
Number of SCs	2
Number of UEs of SCs	[2, 1]

completed, we linearly diminish the reliance on random matching while exploitation becomes more prominent in a linear fashion. Furthermore, the optimal matching is determined through an exhaustive search and is also depicted in the figure.

It is evident that the proposed learning algorithm, when combined with optimal power allocation, gradually converges to the optimal matching. This demonstrates the effectiveness of our proposed learning method in finding near-optimal solutions as it learns and adapts over time.To assess the performance of our proposed scheme, we focus on two key aspects. The first part of our investigation pertains to the impact of channel allocation during the matching step.

We conduct a comparative analysis of our proposed scheme against two other methods: random matching and a distance-based matching method. The "DistanceAlloc" method employs a matching algorithm inspired by the approach presented in [8], which relies on the distances among UEs, followed by our optimal power allocation algorithm. On the other hand, the "RandomAlloc" method utilizes random matching between UEs and available cellular resources, followed by optimal power allocation.

Fig. 3 provides a comparison of the sum-rate performance between our proposed scheme and the "RandomAlloc" and "DistanceAlloc" methods with respect to the maximum transmission power of the MC. It is evident from the plot that allowing the MC to transmit at higher power levels leads to an increase in the network's sum-rate. Furthermore, our proposed scheme consistently outperforms the other methods, showcasing superior sum rate performance. This implies that our approach can significantly minimize interference for CEUEs, ultimately improving the network's overall performance.

Fig. 4 delves into the impact of altering the SC radius on the network’s sum-rate. The plot in Fig. 4 provides a comparative analysis of the sum-rate performance between our proposed scheme, the "RandomAlloc," and the "DistanceAlloc" methods as the SC radius varies.

Notably, it becomes evident that as the SC radius increases, the sum-rate experiences a decline. This observation underscores the trade-off between coverage area and network efficiency. Interestingly, the plot further reaffirms the superiority of our proposed scheme across different SC radii, indicating that our approach consistently outperforms alternative methods under varying network conditions.

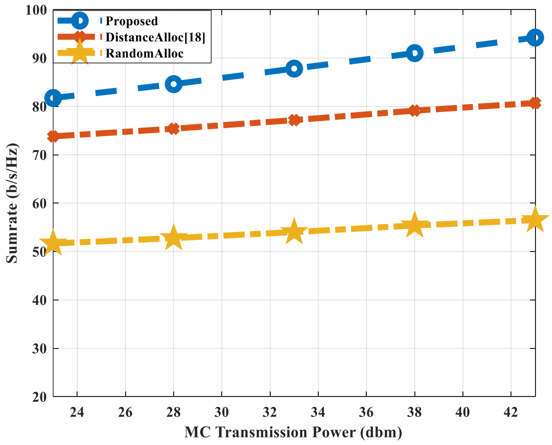

Figure. 3 Sum-rate of the network versus the MC maximum transmission power of the proposed scheme compares to other existing methods that use other matching methods

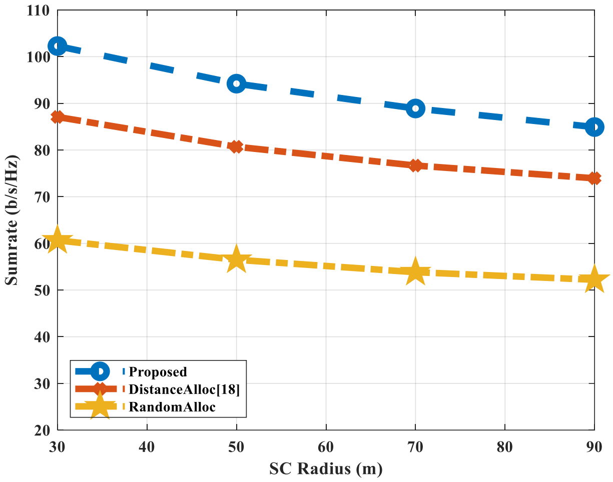

Figure. 4 Sum-rate of the network versus the SC radius of the proposed scheme compares to other existing methods that use other matching methods

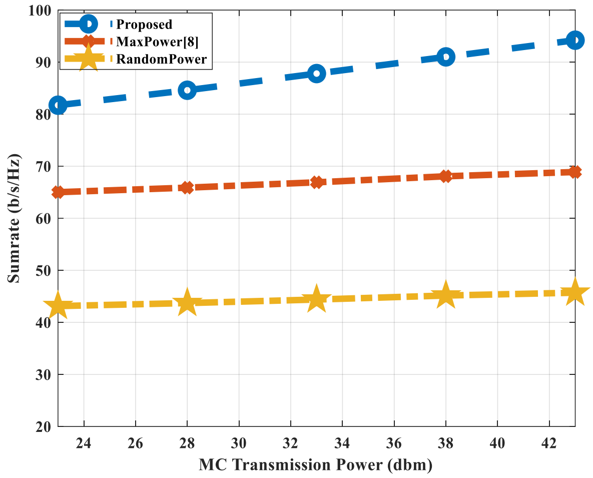

Figure. 5 Sum-rate of the network versus the MC maximum transmission power of the proposed scheme compares to other existing methods that use other power allocation methods

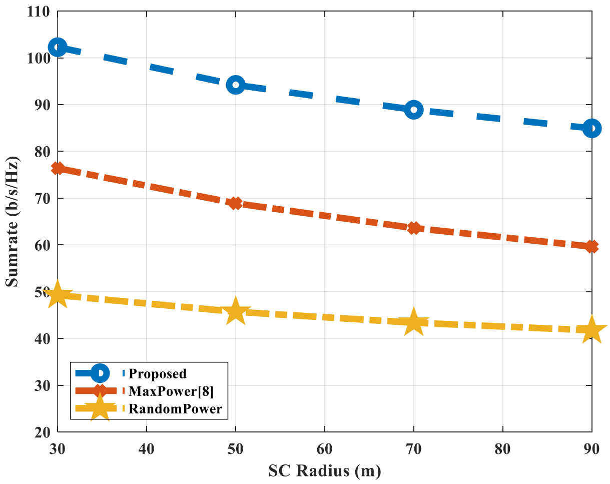

Figure. 6 Sum-rate of the network versus the SC radius of the proposed scheme compares to other existing methods that use other power allocation methods

In addition to the previous comparisons, we can also explore the impact of our proposed optimal power allocation scheme on the network's sum-rate performance. To investigate this aspect, we compare the sum-rate performance of our proposed scheme with that of the "RandomPower" and "MaxPower" methods. In the "RandomPower" method, our proposed matching algorithm is employed, followed by a random power allocation process that satisfies the QoS constraints.

Different random power levels are assigned iteratively until the QoS requirements are met. On the other hand, the "MaxPower" method incorporates our proposed learning-based matching into the approach from [18]. Fig. 5 illustrates the relationship between the network's sum-rate and the maximum transmission power of the MC. As expected, increasing the MC's maximum transmission power leads to an enhancement in the network's sum-rate. The plot confirms the consistent superiority of our proposed scheme over other methods. Fig. 6 further explores the impact of the power allocation method employed in our proposed scheme in comparison to "RandomPower" and "MaxPower," this time with respect to the SC radius.

The plot illustrates the variation in the sum-rate as the SC radius is increased. It becomes evident that there is a decreasing trend in the sum-rate as the SC radius grows larger. performance. Additionally, the proposed scheme compared to other previous studies in terms of contributions and drawbacks in Table 3.

6. Conclusion

In this paper, we tackled the complex challenge of interference mitigation and sum-rate maximization for CEUEs operating within the densely packed landscape of B5G HetNets, where multiple SCs coexist. To address this problem, we converted the problem into multiple parallel resource allocation tasks, meticulously modelling the utilized subcarriers. In addition, we proposed to decouple the problem into two distinct steps: a matching problem and a power allocation problem. Our novel two-step algorithm is designed for maximum efficacy. In the first step, we harnessed RL techniques, observing the sum-rate outcomes of the second step to determine a near-optimal matching

between UEs and available cellular resources. The second step is an iterative process, leveraging the MaMi methodology to estimate a lower bound on the sum-rate function and subsequently maximize the sum-rate of the cell. Crucially, our approach excels in maximizing the sum-rate for CEUEs and effectively mitigating interference, a pivotal aspect of network performance.

The simulation results vividly illustrate the efficiency and effectiveness of our proposed schemes where the proposed matching and power allocation algorithms achieve enhancement of 10% and 25 % compared to existing methods. It Showcases its potential to enhance the performance of CEUEs in dense B5G HetNets significantly.

Conflicts of Interest

The authors declare no conflicts of interest and no financial support.

Table 3. comparison of our proposed method with that of other state-of-the-art existing methods

Work	Method	Contribution	Drawbacks
M. Osama, and et al, in 2021 [1]	SFR-based on / off switching	Used ICR concept with SFR considering the irregular shape of SCs.	It mainly focused on reduce energy consumption more interference and cannot guarantee the minimum QoS requirement
M. Susanto, and et al, in 2021 [10]	dynamic resource allocation	Mitigate Co -tier and cross-tier interference based on partition the cells into sectors and inner and outer regions	The method did not present its objective function and the QoS requirement of UEs
O. T. Asak and et al, in 2021 [12]	Load -Driven SFR	Allocate resources based on the effect of number, demand, and location of users	Used heuristic approach which lacks the robust mathematical foundation.
F. B. Mismar, and et al, in 2019 [15]	Deep RL framework to Jointly optimizing power control, beamforming and interference coordination	Maximize SINR and sum rate using near optimal policy	Despite the exploiting capabilities of MIMO for enhancing end users' rate, there is no consideration for their QoS requirements
J.S. SHEU and et al, in 2023 [19]	RL for Joint power control, beamforming optimization	Remove the need for information exchange and traditional reword	Need huge number of examples to adapt with the variations of cellular network
M.Dahal and et al, in 2023 [20]	Multi-agent deep RL for optimizing beamforming vector and power control	It does not require for exchanging CSI	Suffers from sub-optimal performance due the high dimensional state space
Proposed method	Two step joint sum-rate maximization and interference mitigation based on enhanced q-learning and optimal power allocation	Mitigates co-tier and cross-tier interference while satisfying QoS requirement for all UEs.	

Author Contributions

The first author responsible for Conceptualization, methodology, software, validation formal analysis, investigation, resources, data curation, writing, original draft preparation, writing, review editing, and visualization, and the second author for supervision and project administration.

References

[1] M. Osama, S. El Ramly, and B. Abdelhamid, "Interference mitigation and power minimization in 5G heterogeneous networks", *Electronics*, Vol. 10, No. 14, pp. 1723, 2021.

[2] M. H. Bahonar and M. J. Omidi, "Centralized QoS-Aware Resource Allocation for D2D Communications with Multiple D2D Pairs in One Resource Block", In: *Proc. of Iranian Conference on Electrical Engineering (ICEE)*, pp. 643-648, 2018.

[3] J. Tanveer, A. Haider, R. Ali, and A. Kim, "An overview of reinforcement learning algorithms for handover management in 5G ultra-dense small cell networks", *Applied Sciences*, Vol. 12, No. 1, pp. 426, 2022.

[4] M. Kountouris, "Performance limits of network densification", *IEEE Journal on Selected Areas in Communications*, Vol. 35, No. 6, pp. 1294-1308, 2017.

[5] M. Koolivand, M. H. Bahonar, and M. S. Fazel, "Improving energy efficiency of massive MIMO relay systems using power bisection allocation for cell-edge users", In: *Proc. of 2019 27th Iranian Conference on Electrical Engineering (ICEE)*, pp. 1470-1475, 2019.

[6] I. Lee and D. K. Kim, "Decentralized Multi-Agent DQN-based Resource Allocation for Heterogeneous Traffic in V2X Communications", *IEEE Access*, 2024.

[7] H. Fourati, R. Maaloul, N. Trabelsi, L. Chaari, and M. Jmaiel, "An efficient energy saving scheme using reinforcement learning for 5G

and beyond in H-CRAN", *Ad Hoc Networks,* pp. 103406, 2024.

[8] J. Borah and J. Bora, "Dynamic and location-based power allocation mechanism for inter-cell interference mitigation in 5G heterogeneous cellular network", *International Journal of Communication Systems*, Vol. 33, No. 15, pp. e4548, 2020.

[9] T. M. Shami, D. Grace, A. Burr, and J. S. Vardakas, "Load balancing and control with interference mitigation in 5G heterogeneous networks", *EURASIP Journal on Wireless Communications and Networking,* Vol. 2019, No. 1, pp. 1-12, 2019.

[10] M. Susanto, S. N. Hasim, and H. Fitriawan, "Interference Management with Dynamic Resource Allocation Method on Ultra-Dense Networks in Femto-Macrocellular Network", *Jurnal Rekayasa Elektrika,* Vol. 17, No. 4, 2021.

[11] M. H. Bahonar, M. J. Omidi, and H. Yanikomeroglu, "Low-complexity resource allocation for dense cellular vehicle-to-everything (C-V2X) communications", *IEEE Open Journal of the Communications Society,* Vol. 2, pp. 2695-2713, 2021.

[12] O. T. Asaka, A. Adejo, N. Salawu, A. J. Onumanyi, H. Bello-Salau, and F. T. Oluwamotemi, "Load-driven resource allocation for enhanced interference mitigation in cellular networks", In: *Proc. of 2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS),* pp. 1-6, 2021.

[13] R. Gatti and Shivashankar, "Improved resource allocation scheme for optimizing the performance of cell-edge users in LTE-A system", *Journal of Ambient Intelligence and Humanized Computing,* Vol. 12, No. 1, pp. 811-819, 2021.

[14] L. Xiao, H. Zhang, Y. Xiao, X. Wan, S. Liu, L. Wang, and H. V. Poor, "Reinforcement learning-based downlink interference control for ultra-dense small cells", *IEEE Transactions on Wireless Communications,* Vol. 19, No. 1, pp. 423-434, 2019.

[15] F. B. Mismar, B. L. Evans, and A. Alkhateeb, "Deep reinforcement learning for 5G networks: Joint beamforming, power control, and interference coordination", *IEEE Transactions on Communications,* Vol. 68, No. 3, pp. 1581-1592, 2019.

[16] J. A. Ayala-Romero, J. J. Alcaraz, A. Zanella, and M. Zorzi, "Online learning for energy saving and interference coordination in HetNets", *IEEE Journal on Selected Areas in Communications,* Vol. 37, No. 6, pp. 1374-1388, 2019.

[17] A. Warrier, S. Al-Rubaye, D. Panagiotakopoulos, G. Inalhan, and A. Tsourdos, "Interference mitigation for 5G-connected UAV using deep Q-learning framework", In: *Proc. of 2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC),* 2022: IEEE, pp. 1-8.

[18] A. Pratap, R. Misra, and S. K. Das, "Resource allocation to maximize fairness and minimize interference for maximum spectrum reuse in 5G cellular networks", In: *Proc. of 2018 IEEE 19th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM),* 2018: IEEE, pp. 1-9.

[19] M. Dahal and M. Vaezi, " Multi-agent Deep Reinforcement Learning for Multi-Cell Interference Mitigation", In: *Proc. of 2023 IEEE 2023 57th Annual Conference on Information Sciences and Systems (CISS),* 2023.

[20] J.S. Sheu, C.K. Huang, and C.L. Tsai. "Joint Beamforming, Power Control, and Interference Coordination: A Reinforcement Learning Approach Replacing Rewards with Examples", *IEEE Access*, 2023.