IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 11, 2022, accepted January 30, 2022, date of publication February 11, 2022, date of current version February 24, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3151090

Improving the QoS in 5G HetNets Through Cooperative Q-Learning

MUHAMMAD USMAN IQBAL, EJAZ AHMAD ANSARI, SALEEM AKHTAR, AND ALI NAWAZ KHAN, (Member, IEEE)

Department of Electrical and Computer Engineering, COMSATS University Islamabad (CUI), Lahore Campus, Lahore 54000, Pakistan

Corresponding author: Muhammad Usman Iqbal (usmaniqbal@cuilahore.edu.pk)

ABSTRACT Heterogeneous networks are an integral part of the 5G cellular networks as they are one of the important enabling technologies for increased coverage and capacity. However, interferences in multi-tiered architecture bottleneck its performance. Although multiple schemes have been proposed for efficient radio resource management to handle the interferences in heterogeneous networks but provision of quality of service to macrocell and small cell user equipment simultaneously, is still an open research problem. Intelligent schemes for radio resource management in heterogeneous networks have proved their effectiveness due to their self-optimization capabilities. In this research article, a cooperative Q-Learning, algorithm is proposed for efficient joint radio resource management in ultra-dense heterogeneous networks to handle interferences by adaptive power allocation to small cell base stations while considering the minimum quality of service requirements. In this proposed cooperative Q-Learning algorithm, small cell base stations interacts with the neighboring small cell base stations to exchange information and performs self-optimization based on a joint reward function. The proposed solution not only provided significant improvement in the capacity of macrocell and small cell user equipment as compared to other state of art Q-Learning based radio resource management schemes but also ensure the provision of quality of service to all macrocell and small cell user equipment simultaneously in the cluster of 16 small cells. The proposed solution provided a minimum capacity of 2 b/s/Hz to macrocell and small cell user equipment which is 100% higher than the minimum quality of service requirements defined in literature where none of recently proposed solution could meet minimum quality of service requirements. The results analysis shows that cooperation among the small cells yields a significant improvement of 48% in capacity of small cell user equipment at the cost of a slight increase in computational time as compared to independent learning.

INDEX TERMS Heterogeneous networks, cooperative learning, 5G.

I. INTRODUCTION

Wireless communication technologies have evolved very rapidly from 1G to 5G in the last three decades to meet the demands of exponentially growing cellular network users in terms of higher throughput, data rate, capacity, and coverage while reducing the latency to zero. However, after the emergence of 4G and 5G, the social/ industrial applications are becoming more and more data-centric, data-dependent, and automated. The development of 1G to 5G is not only limited to improvement in throughput, coverage, and capacity, but some other key performance indicators (KPIs) like interference, scalability, energy efficiency (EE), spectral efficiency (SE), and compatibility with previous networks are also a challenge in the design and development of new mobile technologies. Therefore, the dream of the 5G cellular networks (CN), which are expected to connect billions of devices cannot be achieved through simple improvements in 4G due to its peculiar and very stringent requirements. Very high throughput, less than 1ms latency, massive connectivity, EE, SE, better quality of service (QoS), and Quality of Experience (QoE) are some of the prominent features of the 5G CN. The requirements of the 5G are summarized in [1]–[7]. There are many challenges that need to be addressed by the 5G networks as mentioned in [1]–[5], [8] to fulfill the above-mentioned requirements of 5G CN. The 5G CN made significant progress to fulfill the stringent requirements by adding additional features like millimeter-wave (mmW) communication, massive multiple inputs and multiple outputs (MIMO), software defined networking (SDN), network

The associate editor coordinating the review of this manuscript and approving it for publication was Weisi Guo.

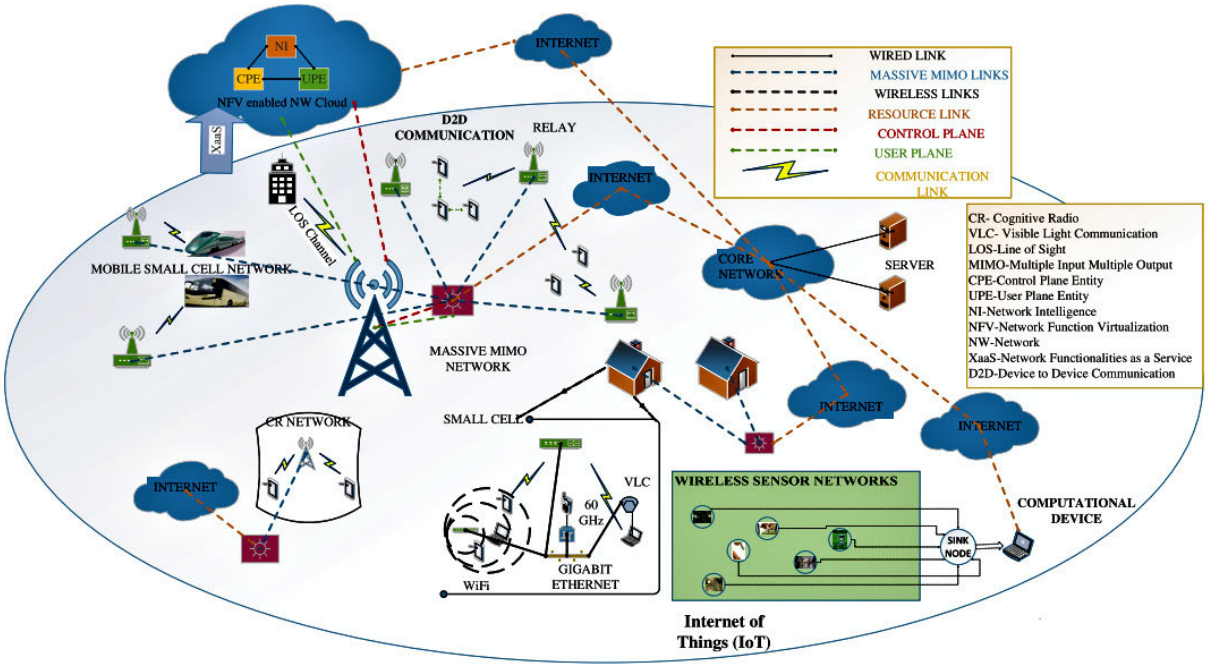

FIGURE 1. Multi-tiered HetNets architecture for 5G CN [8].

function virtualization (NFV), and a complete redesign of the core network. However, these developments could not raise data rate in the order of terabits per second, a latency of hundreds of microseconds, and 10^7 connections per km^2 in very rapidly developing data-centric societies and internet of things (IoT) based automated processes [3], [5], [8], [9]. Although massive MIMO and mmW communication, are referred to as an integral part of the 5G and 6G CN, ultra-dense small cell (SC) Heterogeneous networks (HetNets) and Self Organizing Networks (SON) are the ones that have the potential to solve the problems of high throughput, zero latency, high EE, improved coverage and capacity [3], [5]–[7] but it results in new challenges for the researchers in form of co-tier interference (CoI), cross-tier interference (CrI), and efficient radio resource management (RRM). Interferences due to multi-tier HetNets architecture severely degrade EE, QoS, and QoE. Therefore, to ensure the EE, QoS, and QoE in 5G SC HetNets, effective interference management is vital [7], [10].

Several approaches for interference mitigation have been proposed in the literature based on efficient spectrum utilization, antenna patterns, adaptive power control, and a combination of these schemes. However, a detailed literature review of interference mitigation in 5G CN reveals that the performance of cognition enabled or intelligent interference mitigation is good but the provision of QoS to macrocell user equipment, UE^m, and small cell user equipment, UE^s, simultaneously is still a challenge [7]–[15]. To improve the QoS of UE^m and UE^s simultaneously in ultra-dense SC HetNets, a machine learning (ML) technique based on cooperative learning (CL) is proposed and analyzed in this article.

A. MOTIVATION

The ultra-densification, using different types of SCs based on the number of user equipment (UEs), cell radius and transmit power, in multi-tiered HetNet architecture, is promising solution to meet the explosive data rate and capacity requirements of 5G and 6G CN [5]–[7]. The ultra-densification efficiently offloads traffic among the network tiers to support the exponentially growing UEs with increased QoS, data rates, and EE [2].

Although the deployment of SCs results in numerous benefits; their initial cost, reliability of the complete system, and interferences due to multi-tiered architecture are open challenges [3]. A multi-tiered HetNets architecture for 5G and future CN is presented in Fig. 1.

Recently, researchers have proposed multiple solutions to optimize the reliability, throughput, QoS, QoE, coverage and capacity in SC HetNets by mitigating the CoI and CrI by exploiting SON features defined in LTE 3GPP TS 36.300 [16] and introducing intelligence in the network by ML either through independent learning (IL) [17]–[19], or both IL and CL [20]–[27].

A fundamental limitation of the ML and SON-based schemes is the failure to provide QoS to both UE^m and UE^s simultaneously in ultra-dense SC HetNets by coping with the interferences due to the density of SCs. Furthermore, recently proposed schemes in literature for QoS in SON-based HetNets either utilized IL or CL to optimize the learning process. However, still there is a need to explore an optimal learning strategy. Despite many efforts to provide QoS to UE^m and UE^s simultaneously, current research lacks the crucial features of SON such as either working cooperatively or

independently for autonomous adaptability to the dynamic ultra-dense HetNets conditions while considering minimum QoS requirements, computational time, complexity, signaling overhead and EE.

B. RELATED WORK

Recently many solutions are proposed in context of each of the 5G enabling solutions like mmW, massive MIMO, SDN, NFV and ultradensification to make 5G dream come true. These solutions are focused on SE, EE, optimal resources allocations in mmW based ultra-dense HetNets and data security. Along with these additional tracts of 5G, effective RRM is a vital part of the 5G and future CN to efficiently utilize the radio resources (RR) to ensure QoS and improved network performance. The RMM functions in HetNets like power control, load management, and handover are performed in a distributed way by the base station (BS), user equipment (UE), and other network elements. Recently, authors in [28] proposed optimal power allocation in linearly coded network to improve the SE and reduce the outage probability by exploiting cooperative communication. The proposed solution successfully improved the performance of the system. In another recently proposed solution, authors proposed SDN based solution for effective implementation of ultra-dense HetNets using mmW aiming to reduce the signaling overhead and computational complexity [29]. Although, all the enabling solutions are vital for realization of 5G CN dream but in this article we focused on the RRM for interference mitigation through optimal power allocation in ultra-dense HetNets by expoliting the SON and ML integration.

The integration of SON functionalities in 5G HetNets provides a platform for automatic performance improvement through optimal utilization of RRM in terms of improved coverage and capacity, QoS, profitability for the operators, and a significant decrease in deployment and operational cost [30], [31].

SON was introduced as a 3GPP standard in LTE 3GPP TS 36.300 [16] and 3GPP TS 32.500 [32] which was market-driven as the cellular operators found it a viable solution to many fundamental issues in the development and deployment of LTE and future CN [33]. The benefits of deployment of SON in LTE and future CN in terms of improved interference mitigation and throughput at the reduced cost with high profitability are summarized in [31], [34]. The SON requires cognition/ intelligence to perform SON functionalities to adapt according to the dynamic network conditions in HetNets. Therefore, cognitive radio based RRM solutions were succeeded by more efficient AI/ML based schemes for provision of SON functionalities in HetNets.

Among the ML techniques, reinforcement learning (RL) falls in the category of unsupervised learning that makes it a suitable option for RRM in dynamic communication networks like HetNets in 5G where the network conditions are changing continuously. These features are model-free implementation and less computational complexity. Q-Learning (QL), deep Q Networks (DQN) [35] and Deep Deterministic Policy Gradient (DDPG) [36] are techniques of RL implementation. However, QL, an algorithm using "Dynamic Programming Methods" (DPMs), is a perfect choice for dynamic HetNets as being model-free and less computationally complex as compared to DQN and DDPG. QL may provide robustness, computational efficiency, and scalability to the 5G HetNets [30]. QL can be easily implemented in real-time scenarios in either cooperative or distributed manner as it requires low level processing unit [30], [37]. Therefore, QL as a potential solution to solve the self-configuration and self-optimization problem in SC HetNets for optimal RRM is an area of interest since the last decade. However, for efficient implementation of QL, the design of an appropriate, effective reward function (RF) and learning technique is crucial which considers the constraints in the optimization problem and cooperation among the small cell base stations, BS^s, in the 5G SC HetNets.

QL can be implemented either through IL or CL [38]. An extensive literature review of the QL based RRM techniques reveals that authors in [17]–[19] utilized IL-based QL whereas CL was utilized in [24], [26], [27] for QL. Conversely authors in [20]–[23], [25] utilized both learning paradigms in QL to optimize the RR and compared the IL and CL paradigms. In [17], authors proposed a SON functionality-based transmit power optimization of femtocell base station (BS^f) to manage CrI due to co-channel deployment mode in HetNtes using QL in IL paradigm. Despite, an attractive solution for RRM in HetNets, the proposed solution could not prove superiority against other state-of-the-art schemes. Authors in [20], [21] proposed a similar solution for adaptive power allocation for HetNets based on distributive and cooperative QL for cognitive femtocells (FC) to mitigate the CrI and improve sum capacity of the FC using both learning paradigms, IL and CL. Authors in [20], [21], established that CL is superior to IL in terms of improvement in aggregate FC capacity at the cost of signaling overhead. Despite the detailed theoretical background and multiple improved RFs, the authors did not provide comprehensive results in terms of UE^m capacity and minimum QoS requirements of UE^m and UE^s.

In [18], authors further improved the RF design for QL by considering the distance of the neighboring BS^f and allocate power adaptively to BS^f and reduce the CoI. However, the proposed reward function which was applied using the IL was biased to UE^m and hence did not provided the minimum required QoS to femtocell user equipment (UE^f). The work in [18] has been extended in [22] and utilized the CL for the same RF to improve the learning speed and showed significant improvement in convergence as compared to IL.

An improvement in the RF presented in [18] was proposed in [24] in CL paradigm. Although the proposed RF in [24], handled the bias of RF presented in [18] to some extent but failed to ensure the minimum QoS requirements for both UE^f and UE^m in ultradense HetNets. Later on, authors of [24] extended their work, [23], [25], in the context of SON and

TABLE 1. Summary of Q-Learning based RRM Techniques for HetNets.

Ref	Research Objectives	Learning Mode	Distributed/ Centralized	Optimization Parameters	Key Performance Indicators (KPIs)	Advantages	Limitations
[17]	Interference mitigation	Independent	Distributed	• Transmit Power of FBSs • SINR Threshold for Macro Users • FUE's Average Capacity	• MUE's Capacity • Sum Capacity of System	Reduced Cross-tier Interfereces	• No QoS for FUEs and MUEs • No Co-Tier Interference Mitigation • No Comparative Analysis
[18]	Proximity Reward Function Design	Independent	Distributed	• MUE Capacity • FUE Capacity	• Sum Capacity of Network • QoS for MUE	Impproved QoS for MUE	• No QoS for FUEs • No Co-Tier Interference Mitigation • No Comparative Analysis
[19]	• Resource Scheduling • Capacity Optimization	Independent	Distributed	• MUE Capacity • FUE Capacity	• Throughput • Drop rate • Faireness	Improved throughput for Cell Edge Users	• No Co-Tier interference Mitigation • No QoS for FUES
[20]	Cross Tier Interference Mitigation	Independent/ Cooperative	Distributed	• Transmit Power of FBSs • QoS for Macro Users • Sum Capacity for FUEs	• MUE's Capacity • FUE's Average Capacity • Jain's Fairness Index	• Improved FUE's Sum Capacity	• Low performance of IL • Overhead in CL • No QoS for FUEs • No Co-Tier interference Mitigation
[21]	Interference Mitigation in Co-channel Deployment	Independent/ Cooperative	Centralized	• Transmit Power of FBSs • QoS for Macro Users • Sum Capacity for FUEs	• FUE's Sum Capacity • Robustness • Scalability • Reaction to Dynamic Conditions	• Improved FUEs Sum Capacity • Real-time Implementation	• No QoS Parameters Evaluation • No Comparative Analysis
[22]	Resource Allocation	Independent/ Cooperative	Distributed	• MUE Capacity • FUE Capacity	• Sum Capacity of Network • QoS for MUE • QoS forFUE • Convergence of Q-Learning	• Impproved QoS for MUE • Improved Convergence • CL outperform IL	• No QoS for FUEs • No QoS for MUEs in high density • No Comparative Analysis
[23]	• Power Allocation • Self-organisation	Cooperative	Distributed	• Distributed Power allocation	• QoS Maintenance	• Scalability • Self-organizing Capabilities • Reduced Complexity	• Only FUEs are considered. • No QoS for MUEs • No Cross-tier Interference Mitigation
[24]	• Power Allocation • Interference Mitigation • QoS	Cooperative	Distributed	• Sum Capacity of FUEs • Transmit Power of FBSs • QoS for FUEs and MUEs	• MUE's Capacity • FUE's Capacity • Sum Capacity • Fairness	Improved FUE Capacity and Sum FUE's Capacity	Failed to provide QoS to both MUEs and FUEs in desne HetNets
[25]	• Power Allocation • Self-organization	Independent/ Cooperative	Distributed	• Power allocation • MUE Capacity • FUE Capacity	• MUE's Capacity • FUE's Capacity • Sum Capacity • Sum Power	• Comparison of IL and CL • Improved QoS for MUE	• No QoS for FUEs
[26]	Interference Mitigation in Co-channel Deployment	Cooperative	Distributed	• FUEs Sum Capacity • MUEs Sum Capacity	• FUEs Sum Capacity • MUEs Sum Capacity • Energy Efficiency	• Improved Sum Capacity • Neural Network Based Implementation • CL outperforms IL	• No QoS for FUEs • No QoS for MUEs
[27]	Resource allocation	Cooperative	Distributed	• Power Allocation • QoS for MUE and FUE	• FUE's Sum Capacity • MUE Capacity • Number of Served FUEs	• Improved Convergence for CL	• No QoS for FUEs • No Comparative Analysis

mmW and proposed new improved RFs which were implemented in both CL and IL. The results of [23]–[25] proved superiority of cooperative Q-Learning (CQL) implementation. A summary of recently proposed QL based solution of RRM in HetNets is presented in Table.1.

Although, many recently proposed solutions for optimal RRM in HetNets to implement SON functionalities and interference mitigation using QL are deployed either in distributed or cooperative manner based on IL or CL respectively. However, in the above-cited solutions, the selection of RFs and learning paradigm was not formulated to handle the density and dynamic network conditions in SC HetNets and therefore could not provide a minimum required QoS to UEs either through IL or CL. Furthermore, a proper comparison of CL and IL paradigm in implementation of QL using the same RF and simulation conditions has not been explored in terms of QoS, computational complexity, and other related KPIs. In this paper, we investigated the impact of CL on RRM through QL for maximizing throughput while maintaining the minimum QoS requirements for UE^m and UE^s by mitigating CoI and CrI simultaneously and compared the performance against the IL-based QL algorithms.

C. CONTRIBUTIONS

To mitigate the CoI and CrI simultaneously, in multi-tiered 5G HetNets, we proposed a self-adaptive framework by considering each BS^s as an agent in the MDP in a distributed manner in our previous work [10]. To provide the minimum required SINR to UE^m and UE^s, we systematically developed a RF to optimally allocate transmission power to each BS^s in the HetNets and successfully achieved QoS requirements by effectively mitigating interferences in and among the tiers. However, the distributed implementation of the proposed QL scheme utilized IL for effective RRM through SON functionalities of cognitive BS^s.

In this paper, we have investigated the cooperative implementation of the QL algorithm proposed in our previous work [10] by utilizing the CL paradigm. Contributions of the paper are summarized below:

- To handle the CoI and CrI in the ultra-dense SC HetNets where SCs are equipped with cognition and SON functionalities, a cooperative adaptive power allocation scheme based on QL using CL is proposed. We utilized QL based model of the SCs HetNets as multi-agent MDP where each of the SC's base station, BS^s, acts as the agent in the network and explored the CQL framework in the context of the SON.
- We propose an adaptive power allocation algorithm for SC HetNets in a cooperative manner using CQL to provide minimum required capacity (b/s/Hz) to UE^m and UE^s in ultra-dense HetNets to meet the QoS requirements. The cooperation among the SCs and CQL algorithm for RF maximization is also presented in detail.
- The proposed CQL algorithm for adaptive power allocation to SCs in multi-tiered HetNets is validated in multiple standard interference scenarios by various KPIs related to the QoS requirements which include UE^m capacity, minimum UE^s capacity, and sum capacity of the UE^s, sum power of UE^s, computational time and Jain's fairness index.

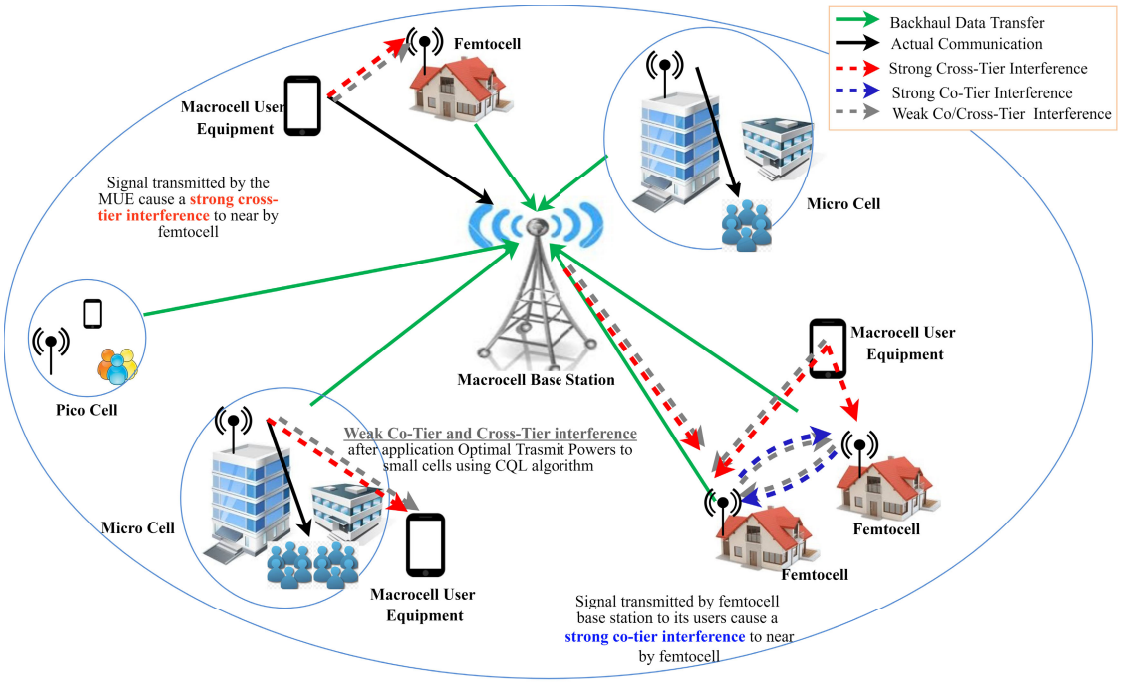

FIGURE 2. System model comprised of different types of SCs overlaid under the MC.

- Results of Monte-Carlo simulations of the proposed solution in various standard interference scenarios based on 3GPP TR36.872 [39], show that the proposed solution successfully provided QoS to both UE^m and UE^s simultaneously in ultra-dense HetNets and also prove its superiority in terms of reduced transmit power and computational time.

The paper is organized as follows: in section II, the system model for exploring CQL for adaptive power allocation in HeNets is presented followed by the problem formulation in section III. In section IV, the RL based RRM using CQL is discussed to model the SC HetNets as the multi-agent MDP. The proposed CQL and RF are presented in section V followed by the simulation setup and parameters for evaluation of CQL in section VI. The results of Monte-Carlo simulations in various standard interference scenarios are presented in section VII whereas the conclusion of the paper is presented in section VIII.

II. SYSTEM MODEL

We employed a system model, presented in Fig.2, composed of the multi-tiered ultra-dense SC HetNets where the SCs are deployed under the overlaid MC in the co-channel deployment mode, which is similar to the one presented in our previous research [10], and [25]–[27]. The system model is based on the Scenario2b in the 3GPP TR 36.872 which is a standard simulation scenario for the evaluation of interference mitigation and QoS enhancement techniques for SCs in HetNets [39]. In Fig.2, ultradensification severely degrades the QoS and QoE for both UE^m and UE^s due to the strong CoI and CrI indicated as the blue and red arrows, respectively. However, an effective interference mitigation scheme will reduce the severity of the interference, indicated with gray arrows in Fig.2, will consequently improve QoS related parameters of all UEs in HetNets.

In this article, we have explored the improvement in QoS of complete cluster of SCs, $\mathcal{C}$, by employing CL-based ML technique for interference mitigation through adaptive power allocation in the downlink of the ultra-dense SC HetNets. We expolite the SON features defined in 3GPP TR 32.500 [32], for CL among the cluster of SCs, $\mathcal{C}$, to provide the required minimum SINR to the both UE^m and UE^s, $\urcorner_{\mathbb{M}}$, and $\urcorner_{\mathbb{C}}$, respectively.

In the system model presented in Fig.2, we consider a single MC of 5G HetNets operating over a set of orthogonal subbands, $\boldsymbol{\beta}$, where $\boldsymbol{\beta} = \{1, 2, 3, \ldots .B\}$, in the downlink transmission. The MC is composed of macrocell base station, BS^m, and UE^m where the BS^m is deployed at the center of the MC and UE^m are located near/ inside the cluster of SCs, $\mathcal{C}$, or at a random location in the coverage area of the MC as as per Scenario 2b in the 3GPP TR 36.872 [39].

A cluster of SCs, $\mathcal{C}$, where $\mathcal{C} = \{1, 2, 3, \ldots .C\}$, is deployed in the coverage area of the MC. All the SCs and their related UE^s deployment is indoor [39]. Each SC in the cluster, $\mathcal{C}$, selects a subband $b \in \boldsymbol{\beta}$, randomly and provide services to one or more related UE^s in the co-channel deployment mode. The QoS related parameters are defined by the operator in the self-configuration process of SCs, in terms of a minimum average SINR, $\urcorner_{\mathbb{M}}$, and $\urcorner_{\mathbb{C}}$ by the BS^m and BS^s respectively. We assumed

that power is equally divided by BS^m and BS^s among their related UEs [40].

In downlink, the SINR at i^{th} UE^m, UE_i^m, where $i = \{1, 2, 3, \ldots, \mathcal{I}\}$ operating on the subband $b \in \boldsymbol{\beta}$, is impacted by the CrI from cluster of SCs, $\boldsymbol{\mathcal{C}}$ where $\boldsymbol{\mathcal{C}} = \{1, 2, 3, \ldots .C\}$. The SINR at the UE_i^m, ς_i^m, can be calculated as

$$\varsigma_i^m = \frac{p_i^m |h_{m,i}^m|^2}{\underbrace{\sum_{c \in \boldsymbol{\mathcal{C}}} p_c^s |h_{m,i}^c|^2}_{\text{CrI}} + N_o} \tag{1}$$

where p_i^m and p_c^s are the transmitted power by BS^m to UE_i^m, and BS^s of c^{th} SC respectively, $h_{m,i}^m$ and $h_{m,i}^c$ are the channel gains from the BS^m and BS^s of c^{th} SC to the UE_i^m respectively. N_o represents the variance, σ^2, of the additive white Gaussian noise (AWGN).

Unlike (1), the SINR at k^{th} UE^s of c^{th} SC, $UE_{c,k}^s$ where $k = \{1, 2, 3, \ldots \mathcal{K}\}$ in the downlink operating on the subband $b \in \boldsymbol{\beta}$, is impacted by CrI from BS^m, CoI from the neighboring BS^s and thermal noise. The SINR at the $UE_{c,k}^s$, $\varsigma_{c,k}^s$, is obtained as

$$\varsigma_{c,k}^s = \frac{p_{c,k}^s |h_{c,k}^c|^2}{\underbrace{\left[p^m |h_{c,k}^m|^2\right]}_{\text{CrI}} + \underbrace{\left[\sum_{j \in \boldsymbol{\mathcal{C}}, j \neq c} p_j^s |h_{c,k}^j|^2\right]}_{\text{CoI}} + N_o} \tag{2}$$

where p^m, p_j^s and $p_{c,k}^s$ are the transmitted power by BS^m, BS^s of j^{th} SC, and BS^s of c^{th} SC to $UE_{c,k}^s$ respectively. $h_{c,k}^c$ $h_{c,k}^j$, and $h_{c,k}^m$ are the channel gains from the BS^s of c^{th} SC, j^{th} SC, and BS^m to the $UE_{c,k}^s$ of c^{th} SC respectively.

Finally, the normalized capacities at the UE_i^m and $UE_{c,k}^s$, C_i^m and $C_{c,k}^s$, respectively, based on (1) - (2) are given below:

$$C_i^m = log_2(1 + \varsigma_i^m) \tag{3}$$

$$C_{c,k}^s = log_2(1 + \varsigma_{c,k}^s), \tag{4}$$

The minimum capacities for providing QoS to UE^m and UE^s, ξ_m and ξ_c, respectively, can be calculated using (3) and (4) by inserting the minimum required SINR of UE^m and UE^s for QoS, i.e. $\urcorner_{\mathbb{M}}$ and $\urcorner_{\mathbb{C}}$. However, these values are network operator defined in the self-configuration process according to the 3GPP in the 3GPP TR 36.300 [16], 3GPP TR 36.814 [41], and 3GPP TR 36.902 [33]. The C_{sum}^s is accumulated value of capacities of all UE^s and defined as follows:

$$C_{sum}^s = \sum C_{c,k}^s \; \forall \; k \; in \; c \; \& \; c \in \mathcal{C} \tag{5}$$

III. PROBLEM FORMULATION

The problem defined in this research is analogous to our previous work [10] and many other recently proposed schemes for optimal resource allocation and interference mitigation in 5G HetNets [20]–[22], [24], [27]. However, the fundamental difference of the optimization problem lies in the optimization function and conditions.

TABLE 2. List of Used Symbols/ Notations.

Symbol	Description
$\boldsymbol{\beta}$	A set of B subbands of frequency where $\boldsymbol{\beta} = \{1, 2, ..., B\}$
$\boldsymbol{\mathcal{C}}$	A Set of SC operating under MC $\boldsymbol{\mathcal{C}} = \{1, 2, ..., C\}$
$\urcorner_{\mathbb{M}}$	Minimum average SINR at UE^m for QoS over each $b \in \boldsymbol{\beta}$
$\urcorner_{\mathbb{C}}$	Minimum average SINR at UE^s of c^{th} SC, $c \in \boldsymbol{\mathcal{C}}$ for QoS over each $b \in \boldsymbol{\beta}$
ς_i^m	SINR at the i^{th} UE^m
$\varsigma_{c,k}^s$	SINR at the k^{th} UE^s of c^{th} SC
p_i^m	Downlink transmission Power of BS^m to to UE_i^m
$p_{c,k}^s$	Downlink transmission Power of BS^s of c^{th} SC to its k^{th} UE^s
$h_{m,i}^m$	BS^m to UE_i^m channel gain
$h_{m,i}^c$	BS^s of c^{th} SC to UE_i^m channel gain
$h_{c,k}^c$	BS^s of c^{th} SC to its $UE_{c,k}^s$ channel gain
$h_{c,k}^j$	BS^s of j^{th} SC to $UE_{c,k}^s$ channel gain
$h_{c,k}^m$	BS^m to $UE_{c,k}^s$ of c^{th} SC channel gain
N_o	Noise Variance, σ^2 of AWGN
C_i^m	Capacity of UE_i^m
$C_{c,k}^s$	Capacity of $UE_{c,k}^s$ of c^{th} SC
C_{sum}^s	accumulated value of capacities of all UE^s
ξ_m	Minimum UE^m capacity for QoS
ξ_c	Minimum UE^s capacity for QoS
$\mathbb{P}$	A set of transmission powers of the BS^s of c^{th} SC $\mathbb{P} = \{p_1, p_2, ...p_{max}\}$
π^*	Optimal policy
$V^*(x)$	Optimal value function
$Q^*(x, a)$	Optimal Q-function
s_t^i	State of SC at time t
α	Learning Rate
$\urcorner$	Discount Factor
$\mathcal{R}_k^t$	Proposed reward function
ζ_m	Absolute difference of C_m and ξ_m
ζ_c	Absolute difference of C_c and ξ_c
K	Total number of episodes
H	Number of steps per episode

In this research, the objective of the optimization problem (OP) is to maximize C_i^m, $C_{c,k}^s$, and C_{sum}^s through an effective intelligent interference mitigation scheme to keep C_i^m and $C_{c,k}^s$ above the minimum required capacity thresholds, ξ_m and ξ_c, which guarantee to provide the minimum QoS requirements to UE_i^m and $UE_{c,k}^s$. Adaptive power allocation to BS^s through intelligent interference mitigation scheme can effectively handle the CoI and CrI and thus improve SINR for all UE^m and UE^s which results in improved minimum capacity thresholds ξ_m and ξ_c.

By assuming that the BS^s of c^{th} SC, where $c \in \boldsymbol{\mathcal{C}}$, operating over a subband, $b \in \boldsymbol{\beta}$, can select a transmit power, p_c^s from the available set of powers, $\mathbb{P} = \{p_1, p_2, \ldots p_{max}\}$, the adaptive power allocation problem is presented as follow:

$$\max_{\mathbb{P}} \quad C_i^m, C_{c,k}^s, C_{sum}^s \tag{6a}$$

$$\text{subject to} \quad p_1 \leq p_c^s \leq p_{max}, \; c \in \boldsymbol{\mathcal{C}} \; \& \; \boldsymbol{\mathcal{C}} = \{1, 2, \ldots .C\} \tag{6b}$$

$$C_i^m \geq \xi_m, \; i = \{1, 2, 3, \ldots, \mathcal{I}\} \tag{6c}$$

$$C_{c,k}^s \geq \xi_c, \; c \in \boldsymbol{\mathcal{C}} \; \& \; k = \{1, 2, 3, \ldots, \mathcal{K}\} \tag{6d}$$

where p_1 and p_{max} are the minimum and maximum transmit powers which any BS^s in the system may select.

The objective function of the OP, presented in (6a), maximize C_i^m, $C_{c,k}^s$, and C_{sum}^s whereas the constraints, (6b), (6c), and (6d) of the (6a), describe limits of p_c^s for each $c \in \mathcal{C}$, C_i^m and $C_{c,k}^s$. The constraints defined in (6c) and (6d), ensure minimum QoS provision to UE^s and UE^m in the ultradense SC HetNets. Constraining the objective function of the OP, (6a), with minimum QoS requirement for UE^s is in line with the [24], [27]. OP in (6a)- (6d) has been discussed in detail in our previous work [10]. By treating the OP in (6a) - (6d) as black-box, we propose to solve it through learning based solution by relating the p_c^s of SCs, $c \in \mathcal{C}$ to the C^m and C_c while constraining over ξ_m and ξ_c. In the next section, the required learning framework to solve the optimization problem in (6a) - (6d) is presented.

IV. REINFORCEMENT LEARNING BASED RADIO RESOURCE MANAGEMENT IN HetNets

Physical layer specifications, simulation scenarios and several key performance indicators (KPIs) for QoS and QoE are defined by the 3GPP in the 3GPP TR 36.300 [16], 3GPP TR 36.814 [41], and 3GPP TR 36.902 [33] for LTE and future CN which are studied through auto-tuning of the parameters by integration of SON features in HetNets for joint RRM (JRRM) in either distributive or cooperative manner. The SON functionalities in LTE and future CN are discussed in detail in [10], [30], [31], [34]. The scope of this article is limited to the capacity optimization under the self-configuration and self-optimization under the SON functionalities in LTE.

The self-configuration, defined in [16], is a pre-operational process which is initialized by powering up an BS^s of SC until the RF transmitter of BS^s is functional. Therefore, during the self-configuration, a new SC configures its hardware and software which include automatic neighbor discovery (AND), transmit power, QoS parameters, and other radio parameters.

In the operational state of an BS^s, the self-optimization process, defined in [16], may auto-tune the initially configured parameters like transmit power in accordance to the defined QoS parameters. However, the self-optimization process can be an independent or cooperation based solution. The self-optimization process in HetNets communication networks is a control process which is usually difficult to design due dynamic conditions in ultra-dense SC HetNets. However, an effective optimization process can be designed through independent or cooperative learning. Therefore, an optimal controller for self-optimization to perform JRRM can be designed through a ML technique known as "Reinforcement Learning" (RL) [37]. RL is non-supervised, a model-free learning technique which satisfies Markov Property and are therefore called as "Markov Decision Process" (MDP). The detailed discussion about RL is presented in [10] and [37].

The SON in HetNets introduced the concept of SCs acting as the single or multi-agent. According to 3GPP, BS^s in SC are capable of SON functionalities by acting as the agent and can share the sensed information with other neighboring BS^s to perform self-configuration and self-optimization. In the multi-agent system, the agents of HetNets, i.e. BS^s in SCs, can utilize sensed information to optimize the resource allocation. Detailed description of SC HetNets as MDP is presented in [10].

V. PROPOSED QL BASED POWER ALLOCATION ALGORITHM IN HetNets AND REWARD FUNCTION

An RL implementation through the QL algorithm is based on the iterative interaction of QL agents and the environment. Three fundamental elements of QL iteration are *(i)* a set of possible actions for QL agents, *(ii)* a set of states of QL agents to be selected after an appropriate action, and *(iii)* a reward for QL agent after taking an action and change in state accordingly. In the RL, an agent strive for a maximum cumulative reward by adopting an optimal policy, $\pi*$ which can be found through the following Bellman optimality equation:

$$V^*(x) = \max_{a \in A} Q^*(x, a) \tag{7}$$

where

$$Q^*(x, a) = \sum_{x'} P_x^a x' \left[R_x^a x' + \gamma \max_a Q^*(x', a') \right] \tag{8}$$

However, finding $\pi*$ is an iterative process of improving the selected policy, found in (7). The (7) can be solved easily through dynamic programming methods (DPM), however, agents should have prior knowledge of their environments. In case no prior information of the environment as in dynamic SC HetNets, (6) can also be solved through the temporal difference method [37]. Therefore, $Q^t(x, a)$ at time t can be found through iteratively updating the following equation.

$$\begin{aligned} &Q^{t+1}(x_t, a_t) \\ &\quad = (1 - \alpha) Q^t(x_t, a_t) + \alpha \{ R_{t+1} + \underbrace{\gamma \max_{a'} Q^t(x_{t+1}, a')}_{R_f} \} \end{aligned} \tag{9}$$

where α represents the learning rate of agent, R_{t+1} is the reward in the current state, R_f is the an approximation of future reward, and γ is the discount factor. The value function is then defined as

$$V^t(x) = \max_{b \in A} Q^t(x, b) \tag{10}$$

The optimal value of the action which maximize $Q^t(x, b)$ the for each state can be computed using the following relation

$$a = arg \max_{b \in A} Q^t(x, b) \tag{11}$$

At any time, t, the action, a_t, is selected based on the following exploration/ exploitation policy (EEP) function [37]:

$$a_t = \begin{cases} arg \max_{a \in A} Q^t(x, b) & exploitation \\ rand_{a \in A}(a) & exploration \end{cases} \tag{12}$$

In the (12), EEP is applied using the "$\varepsilon-$ greedy" policy where exploitation and exploration have probabilities as ε and $1-\varepsilon$ respectively.

In the subsequent subsections we model SC HetNets as the MDP to apply RL for RRM and provide details of proposed CQL algorithm, learning paradigms and proposed RF.

A. SC HetNets AS MDP

In the 5G CN, the RRM and interference mitigation can be considered as a π in the MDP. To model the SC HetNets as the MDP, followings are the basic constituents of MDP in context of HetNets where the BS^s are the agents of multi-agent MDP:

1) ACTIONS

In context of the SC HetNets and above mentioned π, the actions of the agents, $a_c \in A$, are a set of transmission powers, $\mathbb{P}$, of BS^s, where $\mathbb{P} = \{p_1, p_2, \ldots p_{max}\}$.

2) STATES

In RL, state of an agent is its current situation. We have defined, the state of an agent, BS^s, in SC HetNets based on its current distance region from the BS^m and UE^m, D_{BS^m} and D_{UE^m}, respectively. The number of radial distance regions for D_{BS^m} and D_{UE^m} are defined as follows:

$$D_{BS^m} = \{1, 2, \ldots, N_1\}$$
$$D_{UE^m} = \{1, 2, \ldots, N_2\}$$

The each distance region in D_{BS^m} and D_{UE^m}, is based on radius $d^m = \{d_1^m, d_2^m, \ldots, d_{N_1}^m\}$ and $d^s = \{d_1^s, d_2^s, \ldots, d_{N_2}^s\}$, respectively.The number of distance regions, N_1 and N_2, and their corresponding radii in d^m and d^s are operator-defined for the agents, BS^s.

Therefore, at any time t, the state, $x_c^t \in X$ is defined as follows:

$$x_c^t = \{D_{BS^m}, D_{UE^m}\} \tag{13}$$

where X a set of all possible combinations of D_{BS^m} and D_{UE^m}

3) Q-TABLE

A table comprised of all combinations of actions, $a_c \in A$, and states $x_c^t \in X$ is called a Q-Table (QT). In the QT, a_c and x_c are presented in column and rows respectively. The size of the QT depends on the size of the set A and X.

4) REWARD

A reward is a value obtained after an agent performs an action in any state. However, an RF which maximizes the objective function of the OP results in the successful implementation of RL. In this research, the objective function of OP is to maximize the capacity of the UEs while considering the minimum QoS requirement in SC HetNets.

Proposed CQL algorithm for RRM in SC HetNets and RF for the underlying research is presented in the subsequent subsection.

B. PROPOSED CQL ALGORITHM

Based on the rationale of RL and SC HetNets as MDP in previous subsections, we have proposed CQL algorithm, presented in Algorithm 1. Proposed CQL algorithm is based on the definitions of SC HetNets as MDP and is initialized with arbitrary x_c^t and no entry in QT at $t = 0$. In the QL iteration, an action, a_c^t, can be either selected randomly based on exploration or exploitation in EEP (12). The exploitation in EEP involves the learning paradigms, i.e. CL and IL.

After, selection of an appropriate action a_c^t at time t, reinforcement, R^{t+1} is done followed by selection of new state, x_c^{t+1} then QT is updated using the (9). Each agent in the system shares rows of the updated QT with other agents in the system. The execution of the proposed CQL algorithm is presented through the flow chart in Fig.3. The success of CQL lies in the appropriate design of the RF in the (9). The proposed an RF to solve the OP (6a-6d) is presented in the subsequent subsections. Finally the state x_c^t is updated with x_c^{t+1}. The details of IL and CL is presented in the following subsection.

C. INDEPENDENT LEARNING (IL) VS COOPERATIVE LEARNING (CL)

In the SC HetNets as MDP, BS^s act as agents and interact with the environment repeatedly in order to learn an optimal policy, π^*, for optimal RRM to ensure the QoS while striving for maximum the capacity of UE^m and UE^s simultaneously. The agents, BS^s of SCs, can learn from the environment interaction either independently or cooperatively. Details of both learning paradigms are given below:

1) INDEPENDENT LEARNING

In this paradigm, each BS^s in the HetNet learn independently from the environment and consider all other BS^s and their actions as a part of the environment. In the IL, BS^s do not share any QL related information which include sensed information, QT or episodic experience, with other neighboring BS^s. Despite, IL has proved success in wireless communication networks but there could be convergence and oscillation problem. In our previous research [10], we utilized IL for learning of BS^s in 5G SC HetNets and successfully mitigated CoI and CrI simultaneously to provide QoS requirements.

2) COOPERATIVE LEARNING

Although QL algorithms perform well in the IL paradigm, each agent of SC HetNets has to learn itself without any prior information about the environment, therefore, requires more time to learn an optimal policy, π^*. Furthermore, all agents learn an optimal policy, π^*, individually in the IL regardless of its impact on the neighboring agents. In contrast to the IL, a cluster of SCs cooperate by exchanging the information among them in the CL paradigm. This provides prior information to the new agent entering the system by sharing their QT to quickly learn an optimal policy, π^*. The cooperation

Algorithm 1 Cooperative QL (CQL)

Number of QL agents in the system, i.e. SCs, $\mathcal{C}$
For each agent $c \in \mathcal{C}$, Define
A set of states of agents $x_c^t \in X$
A set of possible actions by agents $a_c \in A$
Initialize Q-Table arbitrarily i.e. $Q_t(x_c^t, a_i)$
At $t = 0$,
if $c > 1$ **then**
  Initialize Q-Table i.e. $Q^0(x_c^0, a_c)$
  Update Q-Table with prior information if available i.e Shared Q-Table Rows
else
  Initialize Q-Table i.e. $Q^0(x_c^0, a_c)$
end
for *All* $c \in \mathcal{C}$ *in system, Run CQL in parallel* **do**
  for *Iterations* $\leq N_{itrations}$ **do**
    initialize state x_c^t as x_c^0
    for *Step* $\leq N_{step}$ **do**
      if *rand*$<\epsilon$ **then**
        Select action $a_c^t \in A$ randomly
      else
        if *Learning* == *Cooperative* **then**
          Share $Q_i^t(x_c^t, :)$ with cooperating agents, j,
          Collect $Q_j^t(x_j^t, :)$ from cooperating agents, j,
          $a_i^t \leftarrow arg\max_a \sum_{c\in\mathcal{C},c=1}^{C} Q_k^t(x_c^t, a^t)$
        else
          % Learning == Independent
          $a_c^t \leftarrow arg\max_a Q_c^t(x_c^t, a^t)$
        end
      end
      Perform action a_c^t
      Reinforcement R^{t+1}
      new state x_c^{t+1}
      update Q-Table
      $Q^{t+1}(x^t, a^t) = (1-\alpha)Q^t(x^t, a^t) + \alpha\{R^{t+1} + \gamma V^t(x^{t+1})\}$
      set $x_c^t \leftarrow x_c^{t+1}$ set $t \leftarrow t+1$
    end
  end
  Share the rows of the updated Q-Table with all agents in the system
end

among the QL agents also helps the agents consider the surrounding agents in learning optimal policy, π^* in such a way that it does not negatively impact other agents. Therefore, CL can further reduce the co-tier interference in case of SC HetNets and convergence time for new agent in the system.

The CL is a step a head of the IL paradigm where no information is shared with the neighboring SCs. The cooperation in SCs can be done by sharing information in three different ways, i.e. *i)* instantaneously sensed information, *ii)* episodic information and *iii)* learned policies [38]. In this research, each BS^s shares a portion of its QT with all other cooperating BS^s to cooperatively learn an optimal policy to adaptively allocate BS^s power to handle interferences and improve the capacity of the SCs while considering minimum required QoS parameters as proposed in [20], [21], [25].

CL is performed as follows: a BS^s shares the row of its QT corresponding to its current state with the other neighboring BS^s in its range. Then it selects its actions according to the following equation:

$$a_i^t \leftarrow arg\max_a \sum_{c\in\mathcal{C},c=1}^{C} Q_k^t(x_c^t, a^t) \tag{14}$$

The fundamental concept of the CL lies in the value that is called "Global Q-Function" (GQF) i.e. $Q(x, a)$. The GQF is the QF of the whole multi-agent MDP system. In terms of single and multi-agent, GQF is a combination of QF of all individual BS^s. Therefore in this context when an individual BS^s maximizes its QF, the GQF also increases. However, GQF is not always an optimal solution for any BS^s in the system, but it maximizes the aggregate capacity of the 5G SC HetNets.

D. COMPLEXITY OF PROPOSED CQL ALGORITHM

The complexity of an RL algorithm depends on three fundamental factors, i.e. *i)* the state space size, *ii)* the structure of states, and *iii)* the primary knowledge of the agents [25]. If priori knowledge is not available to an agent or if the environment changes and the agent has to adapt, the search time can be excessive. Considering the above, decreasing the effect of state-space size on learning rate and providing agents with priori knowledge has been a subject of significant research as discussed in related work. Due to the nature of Q-iteration being linear, the complexity of the approach increases in line with the number of states and actions. However, the cooperative approach decreases the number of iterations leading to a reduction in computational complexity. The computational complexity of a model-free QL algorithm is presented below [42], [43]:

$$\begin{aligned} &\text{Regret} && : \Omega(min\{KH, A^{H/2}\}) \\ &\text{Time Complexity} && : \mathcal{O}(KH) \\ &\text{Space Complexity} && : \mathcal{O}(XAH) \end{aligned}$$

where

$$\begin{aligned} K &= \textit{Total number of episodes} \\ H &= \textit{Number of steps per episode} \\ A &= \textit{Number of actions, } a_c \textit{ in } A \end{aligned}$$

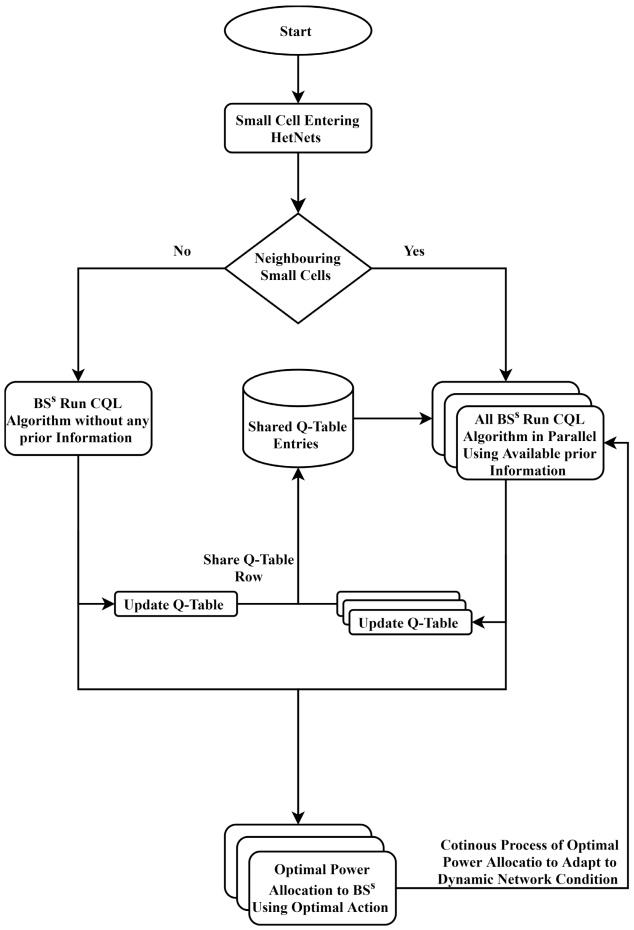

FIGURE 3. Flow chart for execution of proposed CQL algorithm in HetNets.

The maximum number of episodes, K, and number of steps per episode are constants for the proposed CQL algorithm. Therefore, the time complexity of the proposed algorithm is linear.

E. PROPOSED REWARD FUNCTION

An efficiently designed RF is the fundamental requirement of the QL based interference mitigation scheme through adaptive power allocation of BS^s. Despite there is no specific technique or algorithm to derive an efficient RF but in our previous work [10], we elaborated the approach to design the RF and also compared the designed RF with other recently proposed approached for adaptive power allocation based on QL. In this article we utilized our previously propose RF, $\mathcal{R}_c^t$, [10] which is in line with system model presented in section II to solve the solve the OP presented in (6a)-(6d) through QL at BS^s at any time t, defined as a function of $(C_i^m, C_{c,k}^s, \xi_m, \xi_c)$ is given below:

$$\mathcal{R}_c^t(C_i^m, C_{c,k}^s, \xi_m, \xi_c) = \upsilon \underbrace{(C_i^{m,t})^n C_{c,k}^{s,t}}_{\mathfrak{A}} - \upsilon^{-2} \underbrace{\{\zeta_m + \zeta_c\}}_{\mathfrak{B}} \tag{15}$$

TABLE 3. Simulation Parameters.

Simulation Parameter	Value
MC and SC Parameters	
Number of BS^m	1
Number of SCs	16 (Scenario 1-3), 32 (Scenario 4)
Number of UE^m	1 (Scenario 1-3), 2 (Scenario 4)
Number of UE^s in each SC	1 (Scenario 1-3), 2(Scenario 4)
Coverage Area of BS^m	350 m
Coverage Area SCs	10 m
Total Transmit Power of BS^m, p^m	50 dBm
Total Transmit Power of BS^s, p_c^s	-15 to 15 dBm
Number of Power Levels N_{Power}	32
MC Operating Frequency	2.0 GHz
Number of D_{BS^m} Regions, N_1	3
Number of D_{UE^m} Regions, N_2	3
$d^m = \{d_1^m, d_2^m, ..., d_{N_1}^m\}$	50,150,250
$d^s = \{d_1^s, d_2^s, ..., d_{N_2}^s\}$	15,25,40
QoS Parameters	
Minimum UE^m Capacity, ξ_m	1 b/s/Hz
Minimum UE^s Capacity, ξ_c	1 b/s/Hz
QL Parameters	
Learning Rate, α	0.5
Discount Factor ⅂	0.9
Number of Iterations	75000
Channel and Traffic Model	
Channel Model	3GPP TR 36.814 Dual Strip Model [41]
Traffic Model	3GPP TR 36.814 Full Buffer [41]

where

$$\zeta_m = \{C_i^{m,t} - \xi_m\}^2$$
$$\zeta_c = \{C_{c,k}^{s,t} - \xi_c\}^2$$
$$\upsilon = \frac{D_{BS^s - UE^m}}{d_{th}}$$

The proposed RF (15) is a function of two operator provided constants ξ_m and ξ_c, and two variables, C_i^m, and $C_{c,k}^s$. The proposed RF, $\mathcal{R}_c^t$, in (15) is composed of two major parts $\mathfrak{A}$ and $\mathfrak{B}$. The part $\mathfrak{A}$ encourages the system for maximum reward based on C_i^m and $C_{c,k}^s$. Increase in the reward is directly proportional to C_i^m and $C_{c,k}^s$. A more contribution to the reward by the UE$_i^m$ is due to its role as the primary user (PU) in the system. Therefore, a small improvement in C_i^m results in a significant improvement in reward. Any value of n, where $n \geq 2$ can be chosen according the system and priority of the UEs.

The second part, $\mathfrak{B}$, of $\mathcal{R}_c^t$ in (15) guarantees to meet the minimum QoS requirements for UE$_i^m$ and UE$_{c,k}^s$ by incorporating the deviation of C_i^m and $C_{c,k}^s$ from the ξ_m and ξ_c respectively in terms of ζ_m and ζ_c. The deviation from ξ_m and ξ_c are subtracted from the capacity maximizing part, $\mathfrak{A}$, of the reward.

A multiplier υ, based on the distance of the UE^m from nearby BS$_c^s$ and a defined as distance threshold, d_{th} is used a balancing factor between $\mathfrak{A}$ and $\mathfrak{B}$. In the SON based HetNets, the value of the d_{th} in υ is operator-dependent parameter. However, a value between the 15-25 has been proven effective in simulations.

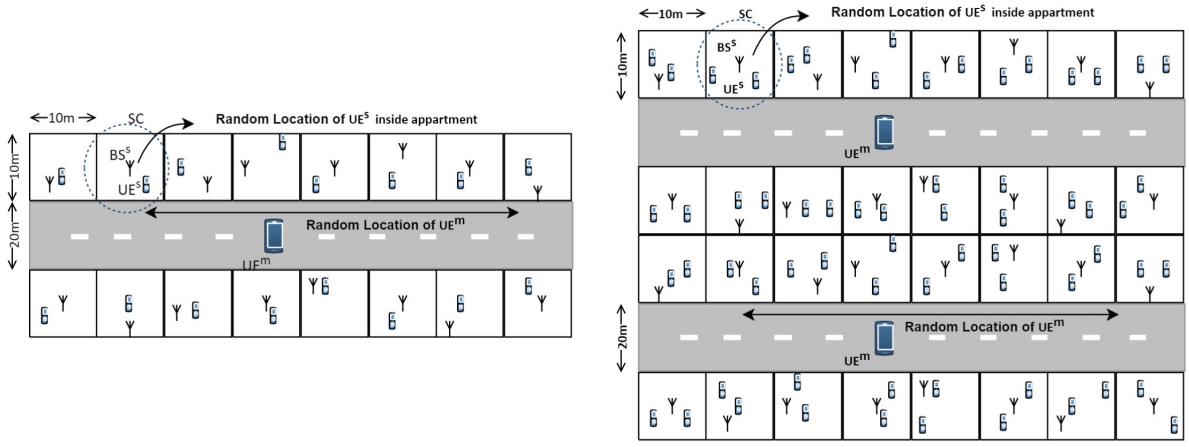

FIGURE 4. Simulation environment (a) Single apartment strip, UE_i^m where $i = 1$ & $UE_{c,k}^s$ where $k = 1$ [39] (b) Two apartment strips, UE_i^m where $i = 2$ & $UE_{c,k}^s$ where $k = 2$ [10].

VI. SIMULATION SETUP AND PARAMETERS

To validate the proposed CQL algorithm for interference mitigation in ultra-dense SC HetNets using adaptive power allocation to BS^s in a cluster of SCs, we employed the standard simulation setup defined by the 3GPP for evaluation of SON in LTE and LTE-A for interference mitigation algorithms [39] and developed it in MATLAB 2020a on Corei7, 16 GB memory machine. We created several interference scenarios based on variation in CoI and CrI, the density of SCs, and the number of UE^m and UE^s. The Scenario 2b (sparse) and Scenario 2b (dense) as prescribed in the 3GPP TR36.872 and 36.814 [39], [41] based on the urban dual strip model are employed as the simulation setup in this article and [10]. However, to further increase the density of SCs and number of UE^s and UE^m, we developed another simulation setup in [10] by increasing the number of apartment strips and UEs by two-fold in comparison to the Scenario 2b (sparse) and Scenario 2b (dense) in the 3GPP TR 36.872 [39]. The simulation setups namely single apartment strip and dual apartment strips are shown in Fig. 4a and Fig. 4b respectively where the UE^s and UE^m may have random positions inside the apartment and on the road respectively.

We developed four different simulation scenarios shown in Fig. 5, based on the simulation setups of Fig.4. Simulation scenario 1 - 3, presented in Fig. 5a-5c are based on the single strip apartment, Fig.4a, whereas the Fig.5d is based on the dual apartment strips, Fig.4b. The location of SCs cluster and UE^m is varied to create a different combinations of CoI and CrI in scenario 1 - 4.

The simulation parameters of MC and SC have been adapted according to the 3GPP TR 36.872 [39]. The minimum required capacity thresholds for UE^m and UE^s, ξ_m and ξ_c, are assumed to be both 1(b/s/Hz). The assumption of these values of the thresholds are in line with the [20], [22], [24]–[27].

To simulate in line with the system model and simulation setup presented in the section II and Fig.4, respectively, a channel model according to 3GPP TR 36.814 is employed [41] whereas traffic model is the full buffer based on the specification provided in 3GPP TR 36.814 [41]. Summary of the simulation parameters is provided in Table 3.

VII. RESULTS

The simulation results were obtained by considering initially one SC in the system and then adding more SCs after convergence of the CQL. The initially obtained parameters after convergence, are used for learning by the particular SC and the newly added SC. After the addition of new SCs, each SC runs CQL individually however it cooperate with nearby SCs by sharing the information to collectively optimize their transmit powers. All the SCs learn and operate in parallel but utilize prior information from other SCs for fast learning. In simulations, sixteen and thirty-two SCs were simulated for simulation scenarios 1-3 and 4 respectively. All the related results are evaluated in terms of the number of SCs in the system. The results of the proposed solution are analyzed in three ways, firstly, if the CL-based proposed solution, CQL, can effectively handle interference in highly dense SC HetNets to provide minimum QoS requirements for both UE^m and UE^s, secondly, the performance comparison of the proposed solution with the other recently proposed solutions in literature [25]–[27] in terms of C^m, C^c, C^s_{sum}, T_c, and Jain's Fairness Index (JFI) and thirdly, the analysis of CL-based proposed solution and our previously IL-based solution, IQL, [10] in terms of various KPIs. The results of Mote-Carlo simulations in terms of different QoS parameters are presented in the subsequent subsections. We initially conducted 500 Mote-Carlo simulations and calculated an optimal number of Mote-Carlo simulations using the technique presented in [44] for a confidence interval of 95%. The statistical

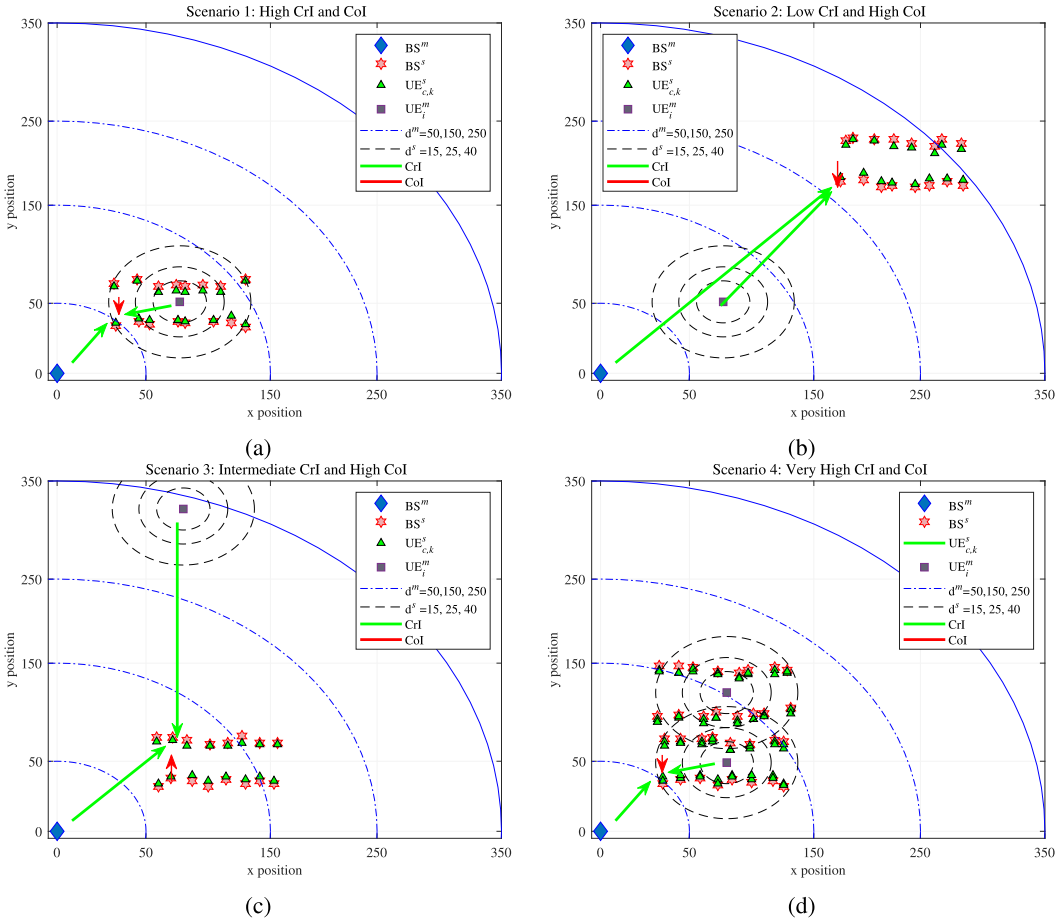

FIGURE 5. Simulation scenarios based on the Fig.4 (a) scenario 1, high CrI and CoI, (b) scenario 2, low CrI and high CoI, (c) scenario 3, intermediate CrI and high CoI, and (d) scenario 4, very high CrI and CoI.

TABLE 4. Statistical Data for Mont-Carlo Simulations.

Statistical Parameter	Value
Total Number of Initial Simulations	500
Standard Deviation, s	1.5
Confidence Interval	95%
z-statistics for 95% confidence interval, z	1.96
Level of Precision, ϕ	0.1
Optimal Number of Mote-Carlo Simulation, n, Using Equation : $\frac{\phi}{s\sqrt{n}} = z$, [44]	864.36
Total Number of Mote Carlo Simulations Conducted	875

data for calculation of optimal number of simulations is presented in the Table 4.

1) CAPACITY OF UE^m

The UE_i^m capacity, C_i^m, is one of the fundamental KPI in the ultra-dense SC HetNets in 5G CN due to its direct relationship to the density of SCs, c. Although C_i^m is a decreasing function with respect to c, it should not fall below ξ_m to ensure the minimum required QoS to UE_i^m irrespective of the c in 5G SC HetNets. C_i^m was measured in all four simulation scenarios of Fig.5 using the proposed solution, recently proposed solutions in literature [25]–[27], and non-adaptive greedy power allocation for BS^s. The results for C_i^m with respect to c, are presented in Fig.6. The minimum threshold capacity for UE_i^m, ξ_m, is represented using a turquoise color line in Fig.6.

In simulation scenario 1, which is a case of high CoI and CrI, UE_i^m, where $i = 1$, is affected by high CrI from the nearby SCs due to presence in the middle of the SCs cluster. Simulation results in Fig.6a shows that for a small number of SCs, c, in the system, C_i^m is high for the proposed solution and the other recently proposed solutions [25]–[27] except greedy algorithm which provides a constant C_i^m but below, ξ_m. However, with the increase of c, in the system, C_i^m decays for the proposed solution and also for the other solutions. However, the decay of the C_i^m for the proposed algorithm is slow enough to not fall below the ξ_m as compared to the other solutions [25]–[27] which decay quickly. Therefore, the proposed CQL algorithm successfully meets the minimum QoS requirements of UE_i^m and provides a C_i^m of 2 b/s/Hz which is twice the ξ_m in a cluster of sixteen SCs. However, the C_i^m provided by the Q-DPA [25] and FAQ [27], decay very rapidly with an increase in density of c, and therefore, fail to provide C_i^m to meet ξ_m. Both Q-DPA [25] and FAQ [27] could provide QoS to only six and ten SCs, respectively as

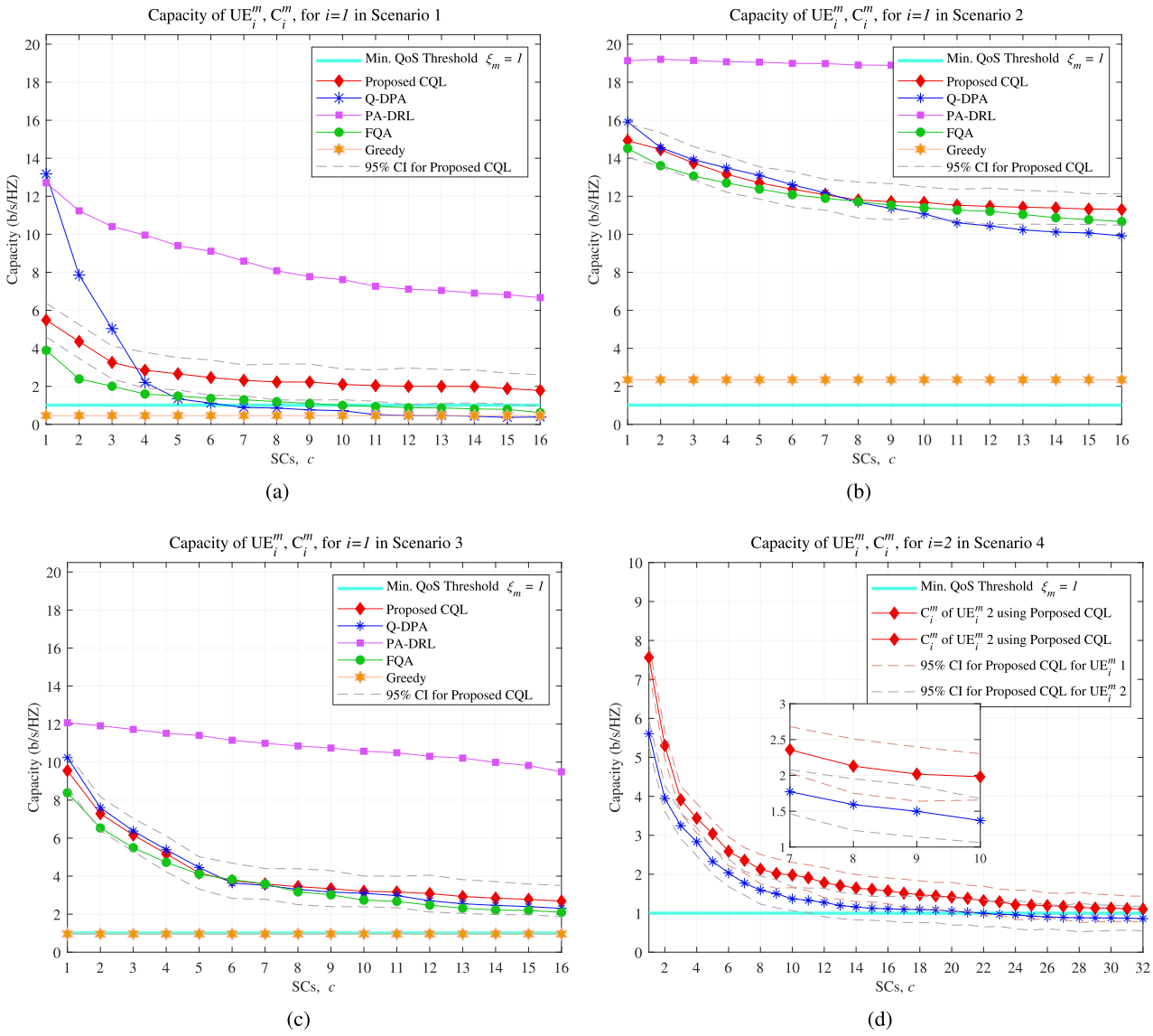

FIGURE 6. Simulation results for capacity of UE$_i^m$,C_i^m, in scenario 1-4, Fig.5, (a) C_i^m in scenario 1, (b) C_i^m in scenario 2, (c) C_i^m in simulation 3, and (d) C_i^m in scenario 4.

compared to the proposed solution which provided QoS up to sixteen SCs. Therefore, proposed solution can support QoS to 62.5% and 37.5% higher number of SCs as compared to Q-DPA [25] and FAQ [27], respectively. However, PA-DRL [26] which has been proven biased to UE$_i^m$, supported QoS for sixteen SCs by maintaining the C_i^m above the ξ_m.

In scenario 2, which is a case of low CrI and high CoI as UE$_i^m$, where $i = 1$, is present close to BS^m and away from the cluster of SCs as shown in Fig.5b. The proposed solution, Q-DPA [25], and FAQ [27] provided a mean capacity of 12 b/s/Hz as shown in Fig. 6b. However, FAQ [26] and the non-adaptive greedy power allocation performed in a similar way as for the simulation scenario 1. The behavior of the C_i^m in simulation scenario 3, remains similar to scenario 2, except a decrease in C_i^m is observed due to increased distance of BS^m and UE$_i^m$.

Despite the increase in the number of UE$_i^m$ and UE$_{c,k}^s$s in scenario 4, Fig.5d, where the $i = k = 1, 2$, the proposed CQL performed in a similar way as for simulation scenario 1-3 as shown in Fig.6d and provided QoS to UE$_i^m$. Initially, C_i^m was high but decays with the increase in density for both UE$_i^m$, however, at the density of sixteen SCs in the system, the C_i^m became nearly constant for both of the UE$_i^m$ in the system. The simulations results for scenario 4 prove the capability of the CQL to meet the minimum QoS requirements for UE$_i^m$ even in the ultra-high density of SCs.

2) MINIMUM CAPACITY OF UE^s

Providing QoS to all UE$_{c,k}^s$ in a cluster of a large number of SCs is a difficult task due to ultra-densification and dynamic conditions in HetNets. To ensure QoS in ultra-dense SC HetNets, minimum UE$_{c,k}^s$ capacity, $C_{c,k}^c$, in a cluster of c SCs,

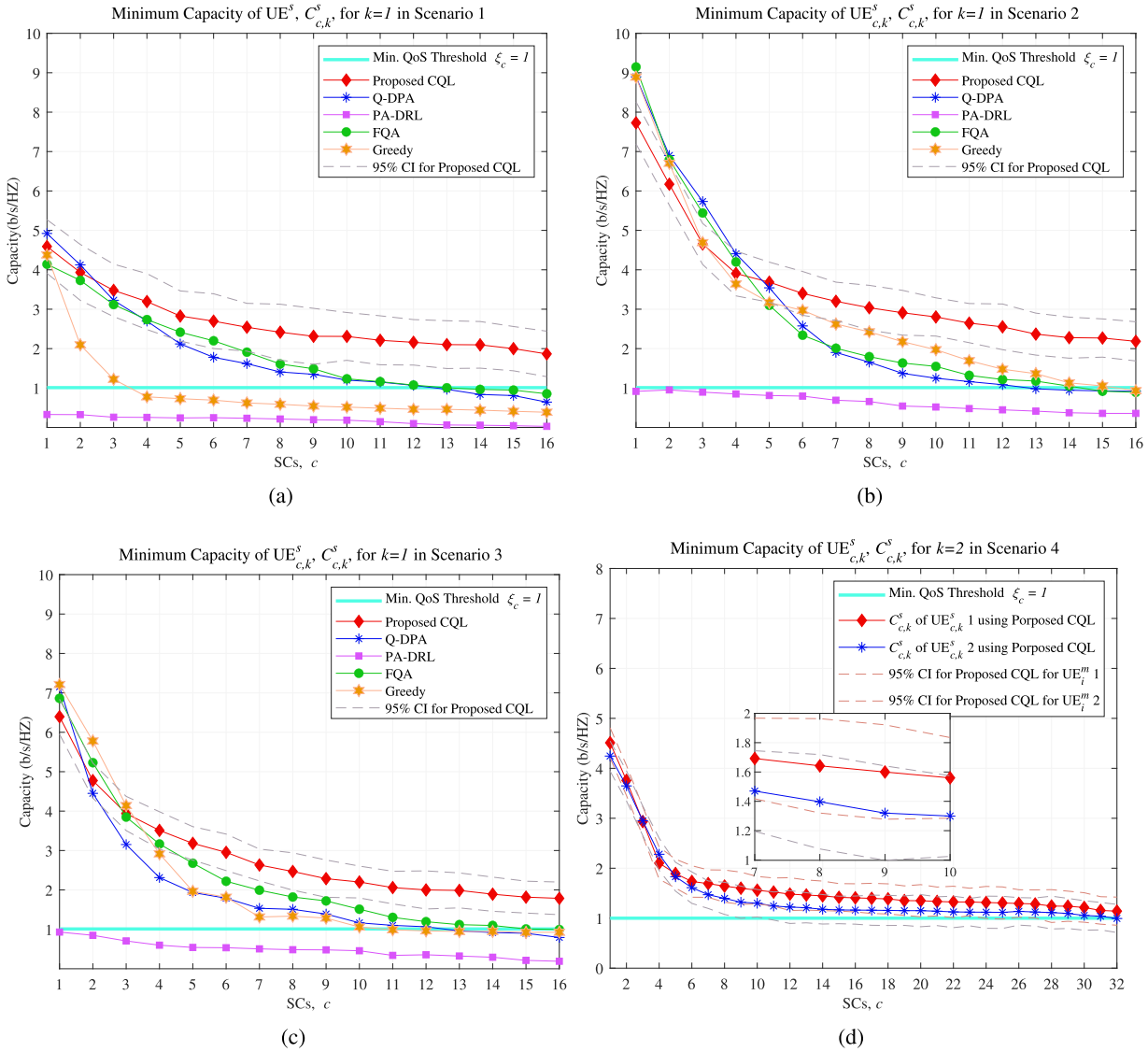

FIGURE 7. Simulation results for Minimum capacity of UE$^s_{c,k}$, $C^s_{c,k}$, in scenario 1-4, Fig.5 (a) $C^s_{c,k}$ in scenario 1, (b) $C^s_{c,k}$ in scenario 2, (c) $C^s_{c,k}$ in scenario 3, and (d) $C^s_{c,k}$ in simulation scenario 4.

should always be greater than or equal to ξ_c. The $C^s_{c,k}$ is also a decaying function of c due to an increase in CoI and CrI. $C^s_{c,k}$ was measured in all four simulation scenarios of Fig.5 using the proposed solution, recently proposed solutions in literature [25]–[27], and non-adaptive greedy power allocation. The simulation results are presented in Fig.7. In Fig.7, ξ_c is represented using a turquoise color line.

In the simulation scenario 1-3, Fig.5, the minimum value of $C^s_{c,k}$ provided by the proposed CQL was 2b/s/Hz which is twice the ξ_c. Therefore, the proposed CQL provided QoS to all sixteen SCs in simulation scenario 1-3, Fig.5, whereas other recently proposed solutions, [25]–[27], could provide $C^s_{c,k}$ above ξ_c to few SCs only. Q-DPA [25] provided $C^s_{c,k}$ above the ξ_c to 12, 13, and 12 SCs in scenarios 1, 2 and 3 respectively whereas the FQA [27] provided $C^s_{c,k}$ in a similar pattern to Q-DPA [25] which remain above the ξ_c for 13, 14 and 14 SCs in the scenario 1, 2 and 3 respectively. In comparison to the proposed CQL, Q-DPA [25] and FQA [27], PA-DRL [26] failed to provide $C^s_{c,k}$ above the ξ_c for any UE^s in all three scenarios due to biasness of its RF to UE^m capacity as discussed in VII-1. The non-adaptive greedy power allocation which results in high CoI and CrI due to maximum transmit power of the BS^s could provide $C^s_{c,k}$ above the ξ_m for only 3,15, and 10 SCs in simulation scenario 1, 2, and 3 receptively.

In ultra-dense simulation scenario 4, Fig.5d, where the number of UE^m_i and UE$^s_{c,k}$ are 2/SC, the proposed CQL provided QoS to all UE^s in a similar way as for simulation scenarios 1-3 as shown in Fig.7d. Despite, the number of the SCs, c, and number of UE^s are twice in simulation scenario 4 as compared to simulation scenario 1-3, proposed CQL provided C^c greater than or equal to ξ_c to all UE^s where

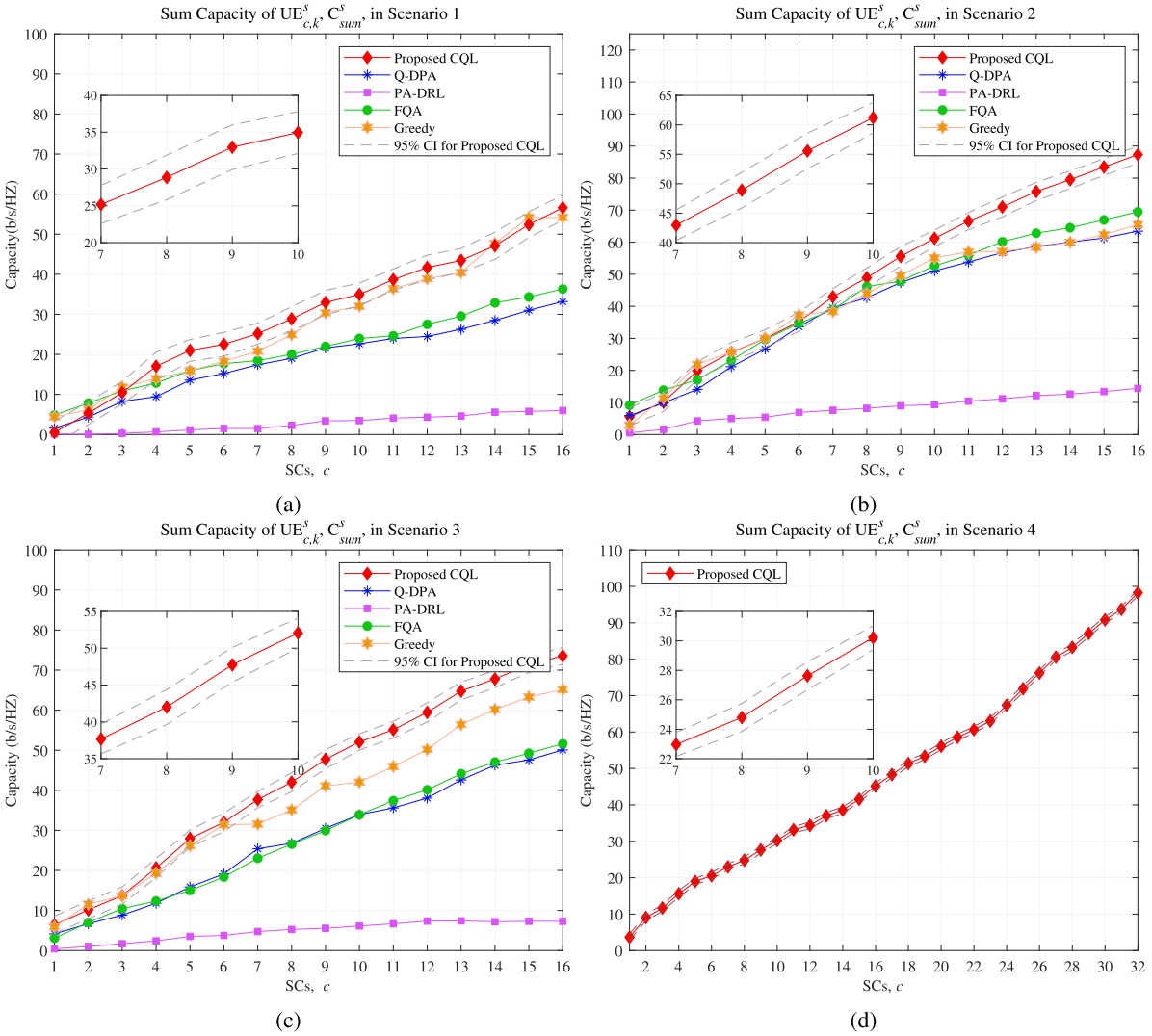

FIGURE 8. Simulation results for sum capacity of UE$^s_{c,k}$, C^s_{sum}, in scenario 1-4, Fig.5 (a) C^s_{sum} in scenario 1, (b) C^s_{sum} in scenario 2, (c) C^s_{sum} in scenario 3, and (d) C^s_{sum} in scenario 4.

all other recently proposed solutions in literature could not meet the minimum QoS requirements.

3) SUM CAPACITY OF UE^s

The sum capacity of the UE^s, C^s_{sum}, which represents the throughput of the system, is an important KPI of the resource allocation algorithms in the ultra-dense SC HetNets. In contrast to the, C^m_i and $C^s_{c,k}$, the C^s_{sum} is an increasing function of c. Like the C^m_i, and $C^s_{c,k}$, C^s_{sum} was measured in all four simulation scenarios of Fig.5 using the proposed CQL, recently proposed solutions in literature [25]–[27], and non-adaptive greedy power allocation algorithm. The results for C^s_{sum} are presented in Fig.8. C^s_{sum} is not a QoS related parameter, therefore, there is no minimum value of C^s_{sum}. However, a higher value of C^s_{sum} shows the capability of a solution to efficiently handle the interferences, resulting in high throughput of the system.

The proposed CQL outperformed the other solutions [25]–[27] and provided a higher C^s_{sum}, in all of the interference scenarios whereas the performance of greedy power allocation remained close to the performance of the proposed solution as shown in Fig.8a-c. In simulation scenario 4, the proposed CQL provided C^s_{sum}, nearly twice the C^s_{sum} provided in simulation scenarios 1-3 which shows the capability of the proposed algorithm to provide higher throughput even in ultra-dense and high interference scenarios.

4) SUM POWER OF UE^s

The sum power of the BS^s, P_{sum}, is the sum of the power transmitted by all BS^s in the system. A high P_{sum} value indicates that BS^s are transmitting at high powers and therefore will cause CoI and CrI to neighboring UE^s of other SCs and UE^m whereas the low value of P_{sum} indicates the effectiveness

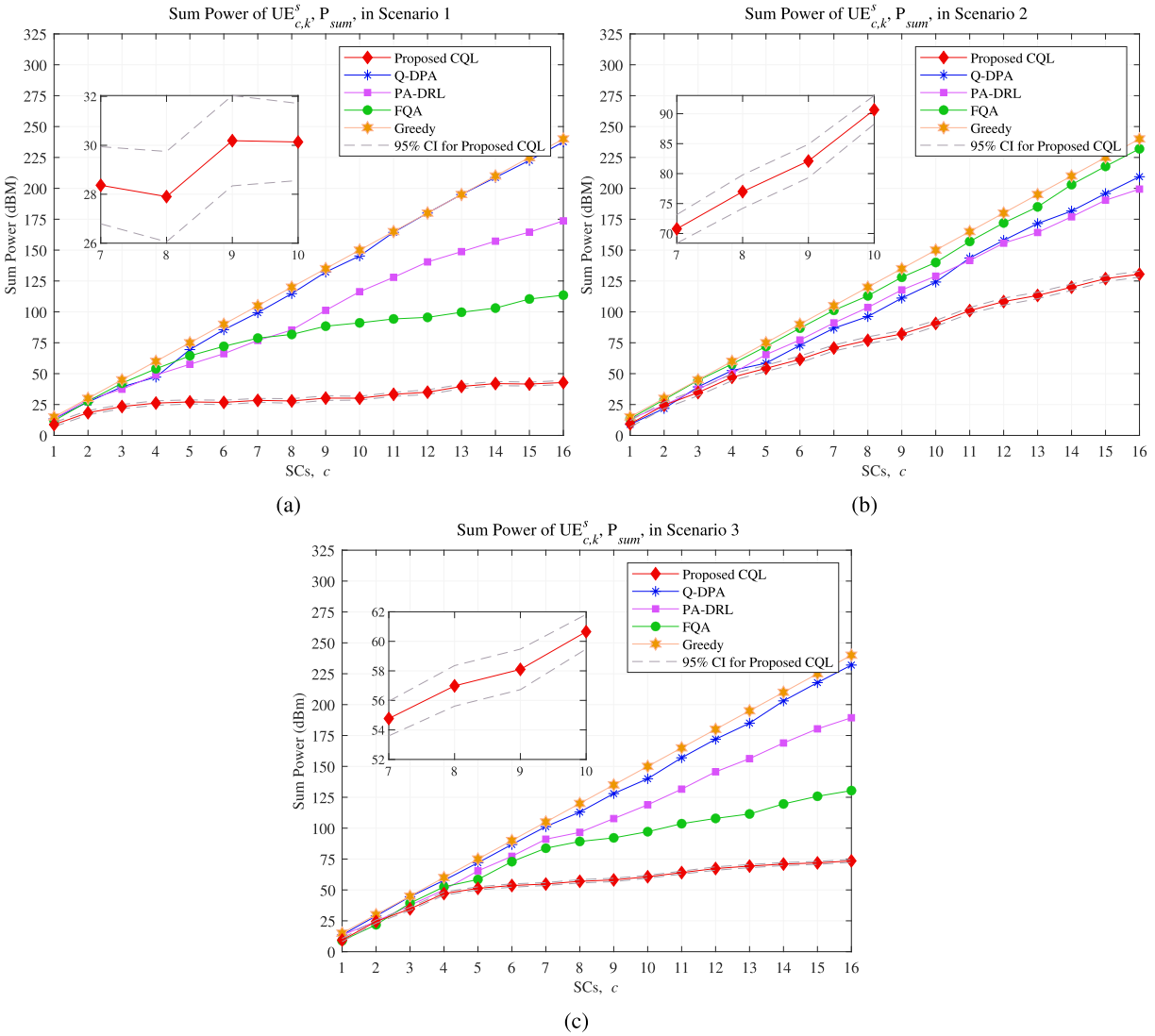

FIGURE 9. Simulation results for sum power of UE^s, P_{sum}, in scenario 1-3, Fig.5 (a) P_{sum} in scenario 1, (b) P_{sum} in scenario 2, and (c) P_{sum} in simulation scenario 3.

of adaptive control of the transmit power of the BS^s which will result in effective mitigation of CoI and CrI. Transmitting at high powers will also significantly reduce the EE of the individual BS^s and as well as the overall SC HetNets. The P_{sum} which is an increasing function of c, is measured in simulation scenario 1-3, presented in Fig.5 using the proposed solution, recently proposed solutions in literature [25]–[27], and non-adaptive greedy power allocation for BS^s and results are presented in Fig.9. In all scenarios 1-3, the proposed solution successfully controlled the transmit power and P_{sum} remain significantly less than the greedy power allocation and other solutions [25]–[27]. The P_{sum} using Q-DPA [25], remained close to the greedy power allocation which is maximum non-adaptive power allocation. The PA-DRL [26] and FAQ [27] performed comparatively better than Q-DPA [25], however, their performance lag in other QoS-related KPIs.

In simulation scenario 1, which is a case of high CoI and CrI, proposed solution optimally controls the transmit power according to interference scenario. Therefore, using the proposed solution, the P_{sum} remain limited at 47dBm, as in Fig.9a with a cluster of 16 SCs which is 81% less than the Q-DPA [25] and greedy power allocation for the cluster of the same size. The PA-DRL [26] and FAQ [27] performed comparatively better than [25] but still 59% and 27% higher than the proposed solution.

A similar behavior of P_{sum} using the proposed solution can be observed in Fig.9b and Fig.9c for simulation scenarios 2 and 3, where P_{sum} remain limited to 125dBm and 75dBm, respectively. Therefore, the proposed CQL algorithm successfully controls the transmit power of the BS^s in the system to mitigate CoI and CrI simultaneously in all three simulation scenarios.

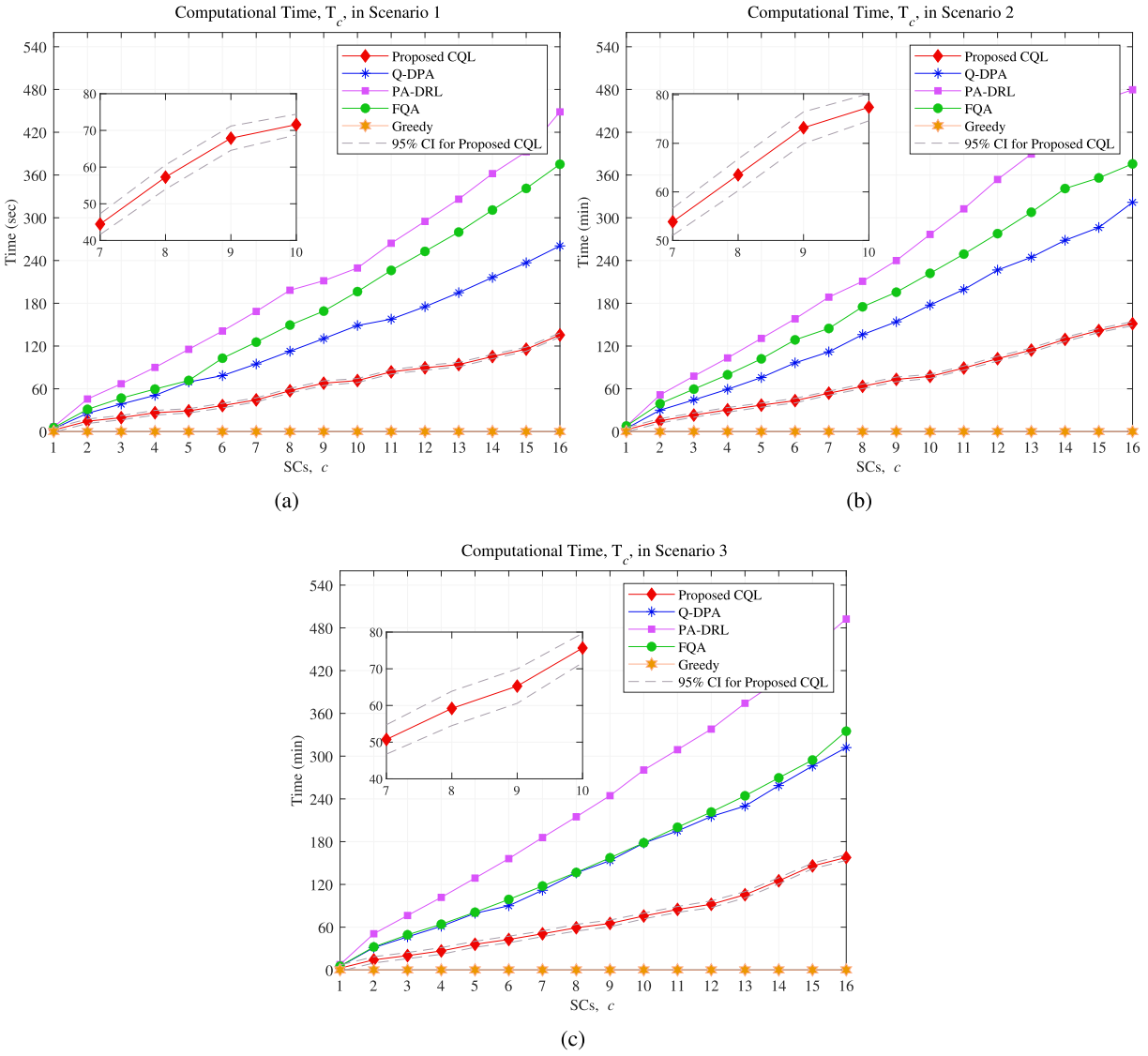

FIGURE 10. Simulation results for computational time, T_c, in scenario 1-3, Fig.5 (a) T_c in scenario 1, Fig.5a, (b) T_c in scenario 2, Fig.5b, and (c) T_c in scenario 3, Fig.5c.

5) COMPUTATIONAL TIME

In the ultradense SC HetNets, conditions are dynamic and therefore the robustness of the system is an important parameter to address the dynamic conditions. Computational time is the measure of the total time required by a BS^s entering in the system for self-organization and self-optimization. A less computational time shows the robustness of the convergence of the RF of a QL algorithm in the self-optimization process. The computational time, T_c in CQL becomes more important as compared to the IQL due to cooperating signaling among the BS^s. The T_c of CQL remains slightly higher than the IQL when the number of SCs, c, are greater than or equal to 2.

To analyze the T_c of the proposed CQL algorithm, it has been measured in simulation scenario 1-3 of Fig.5 using the proposed solution, recently proposed solutions in literature [25]–[27], and non-adaptive greedy power allocation for BS^s. As shown in Fig. 10a-c, T_c remain highest for the PA-DRL [26] and zero for non-adaptive greedy power allocation due to being non-adaptive. However, proposed solution performed significantly better than all of the three adaptive power allocation algorithms, [25]–[27], in all three simulation scenario of Fig.5. The proposed solution requires only 2.25 minutes for convergence as compared to 5, 6.5, and 7.5 minutes by Q-DPA [25], PA-DRL [26], and FQA [27], respectively, with negligible change in all three simulation scenarios. Therefore, the proposed solution is more robust as compared to the other recently proposed solutions and require significantly less computational time in cluster of 16 SCs.

6) CONVERGENCE ANALYSIS

In the simulation parameters, the maximum number of QL iterations are set 75×10^3. Although the QL iterations is a user-defined parameter but it has a great impact on the accuracy of the QL and computational time. The QL is

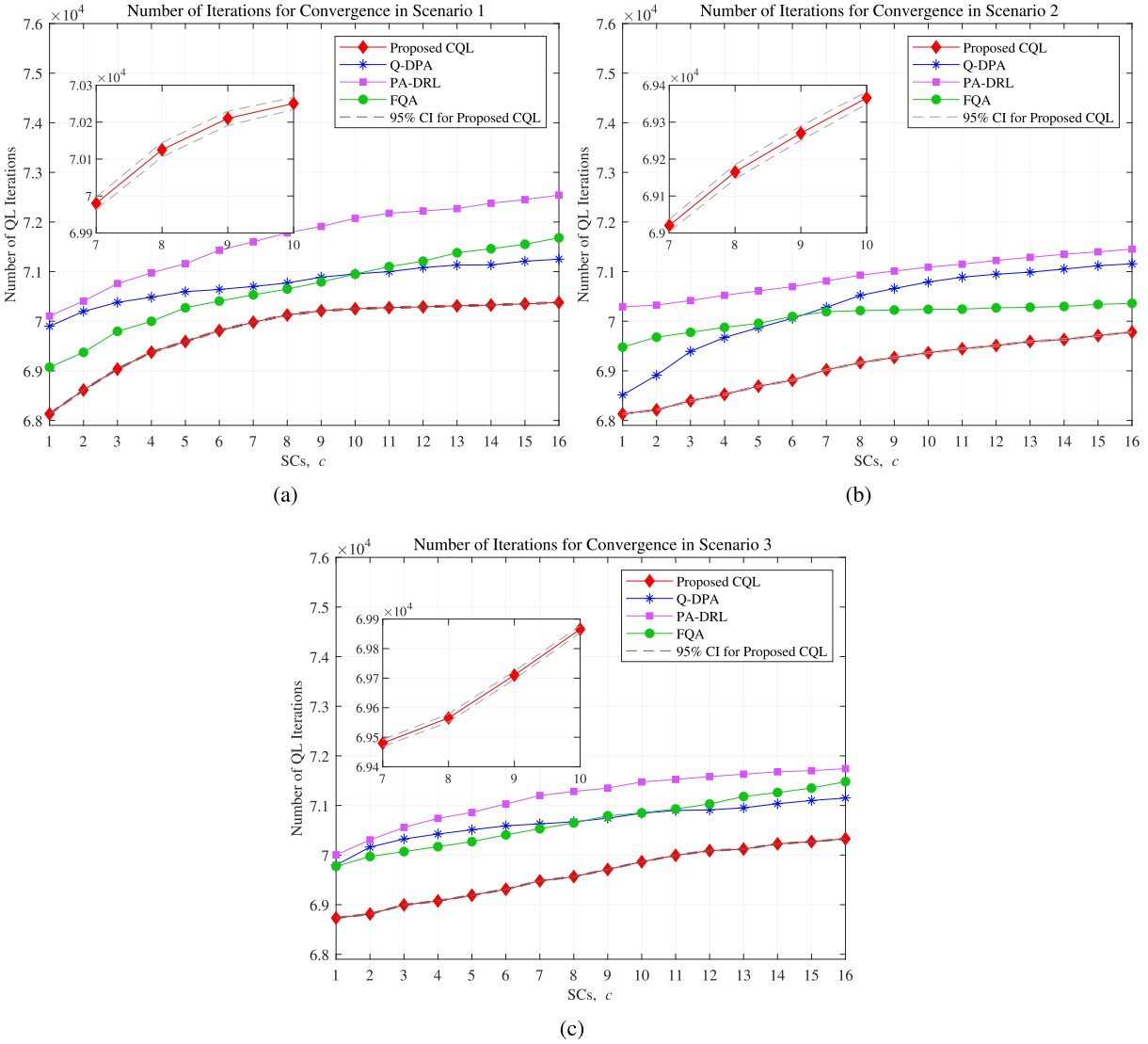

FIGURE 11. Convergence Analysis for scenario 1-3, Fig.5 (a) number of QL iterations in scenario 1, Fig.5a, (b) number of QL iterations in scenario 2, Fig.5b, and (c) number of QL iterations in scenario 3, Fig.5c.

converged if error magnitude is less than 0.001 for 1000 consecutive iterations. The proposed CQL converged in the less number of QL iterations in all three simulation scenarios as compared to the other recently proposed solutions Q-DPA [25], PA-DRL [26], and FQA [27] as shown in Fig. 11. Although QL iterations are an increasing function of SCs but due to CL and sharing of QT rows, the increase in QL iterations remained below the maximum iterations threshold with increase in number of SCs. The QL iterations for Q-DPA [25] and FQA [27] remain close to each other whereas PA-DRL [26] remained the most computational extensive.

7) JAIN's FAIRNESS INDEX

In the dynamic SC HetNets where the conditions are changing continuously and RRM is adaptive to the conditions, it is strongly desired that radio resources are distributed evenly among the SC in the HetNets so that an even throughput can be achieved. Otherwise, an unfair resource allocation will result in an uneven throughput distribution where some of the SCs will strive for resources. Therefore, measuring the fairness of radio resource allocation among the SCs is a widely used metric in SC HetNets. To evaluate the fairness of the proposed solution, we utilized the Jain's Fairness Index [45]. The Jain's Fairness Index is defined as follows:

$$JFI = \frac{\left(\sum\limits_{k\in c, c\in\mathcal{C}} C^s_{c,k}\right)^2}{c\sum\limits_{k\in c, c\in\mathcal{C}} (C^s_{c,k})^2} \tag{16}$$

The value of the JFI lies between 0 and 1 where 1 represents the maximum fairness. The JFI has been measured in

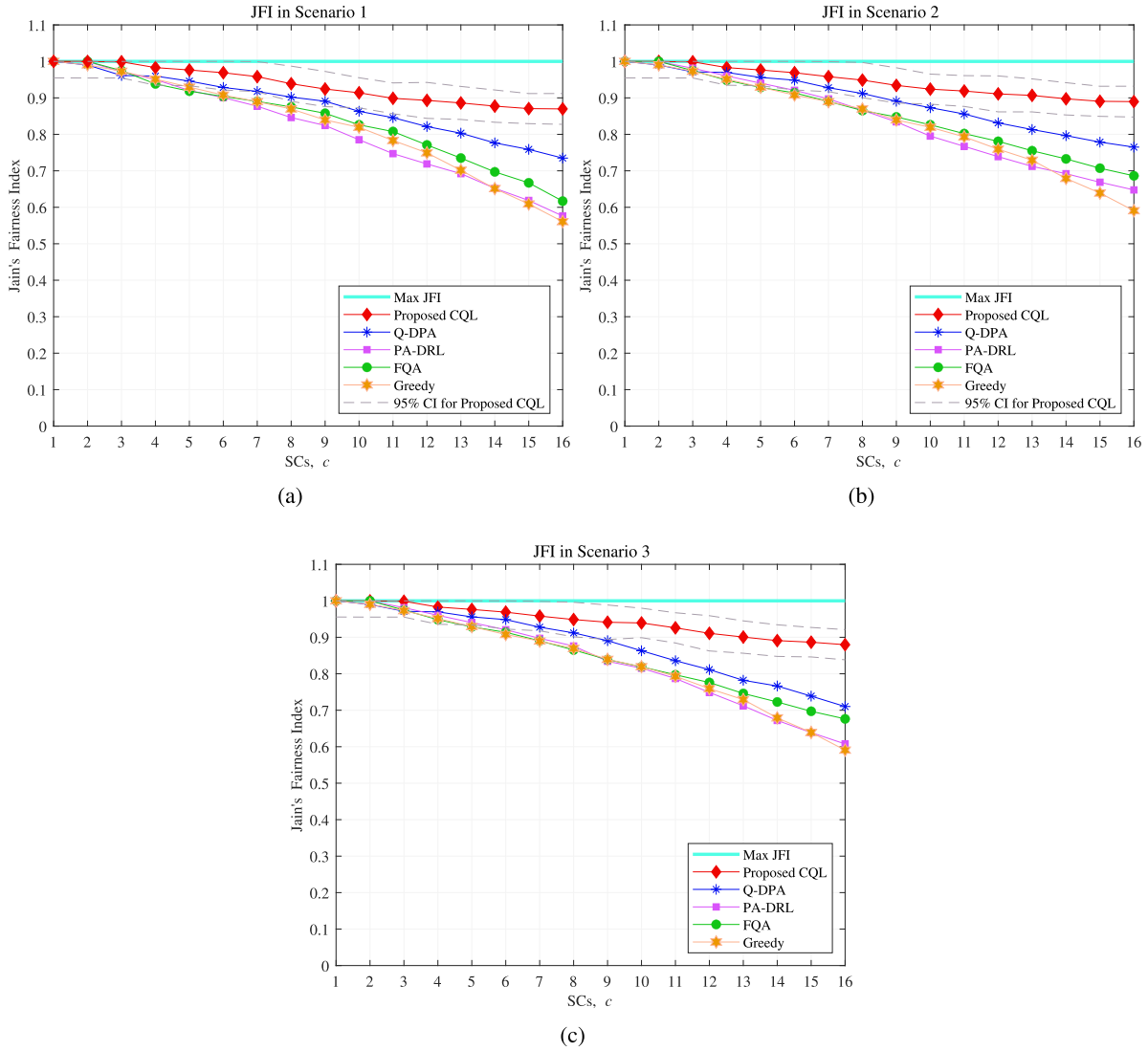

FIGURE 12. Simulation results for JFI in scenario 1-3, Fig.5 (a) JFI in scenario 1, Fig.5a, (b) JFI in scenario 2, Fig.5b, and (c) JFI in scenario 3, Fig.5c.

simulation scenarios 1-3 of Fig.5 using the CL-based proposed solution, other recently proposed solutions in literature [25]–[27], and non-adaptive greedy power allocation for BS^s. The JFI is a decreasing function of the density of SCs. Therefore, the JFI decreases as the number of SCs increase in the system for all of the simulated solutions. However, the rate of decrease of JFI for the proposed solution is much less as compared to the other solutions, [25]–[27], and non-adaptive greedy power allocation. PA-DRL [26] and FAQ [27] and greedy power allocation performed worst and JFI values fall to 0.6 as compared to Q-DPA [25] and the proposed solution which maintained 0.75 and 0.9 in all three simulation scenarios with a cluster size of 16 SCs. Simulation results show that the proposed solution can fairly allocate radio resources among the SCs in a large cluster of SC for even distribution of throughput.

8) PERFORMANCE COMPARISON OF CQL AND IQL

Despite the proposed CQL algorithm has performed better than the recently proposed solutions in the literature, [25]–[27], in terms of various QoS KPIs as discussed in the previous subsections, but its comparison with the IQL algorithm [10] is important to find an optimal learning strategy i.e. either CL or IL. We have compared the proposed CQL algorithm with our previously proposed IQL algorithm for interference mitigation through adaptive power allocation [10] in the simulation scenarios of Fig.5 for the four KPIs, C_i^m, $C_{c,k}^s$, C_{sum}^s and T_c. The comparison is presented in Fig. 13.

Comparison of C_i^m using CL and IL [10] based QL is presented in Fig.13a. It can be observed that the CQL algorithm performed very close to the IQL algorithm. However, its performance is better than the IQL algorithm in simulation

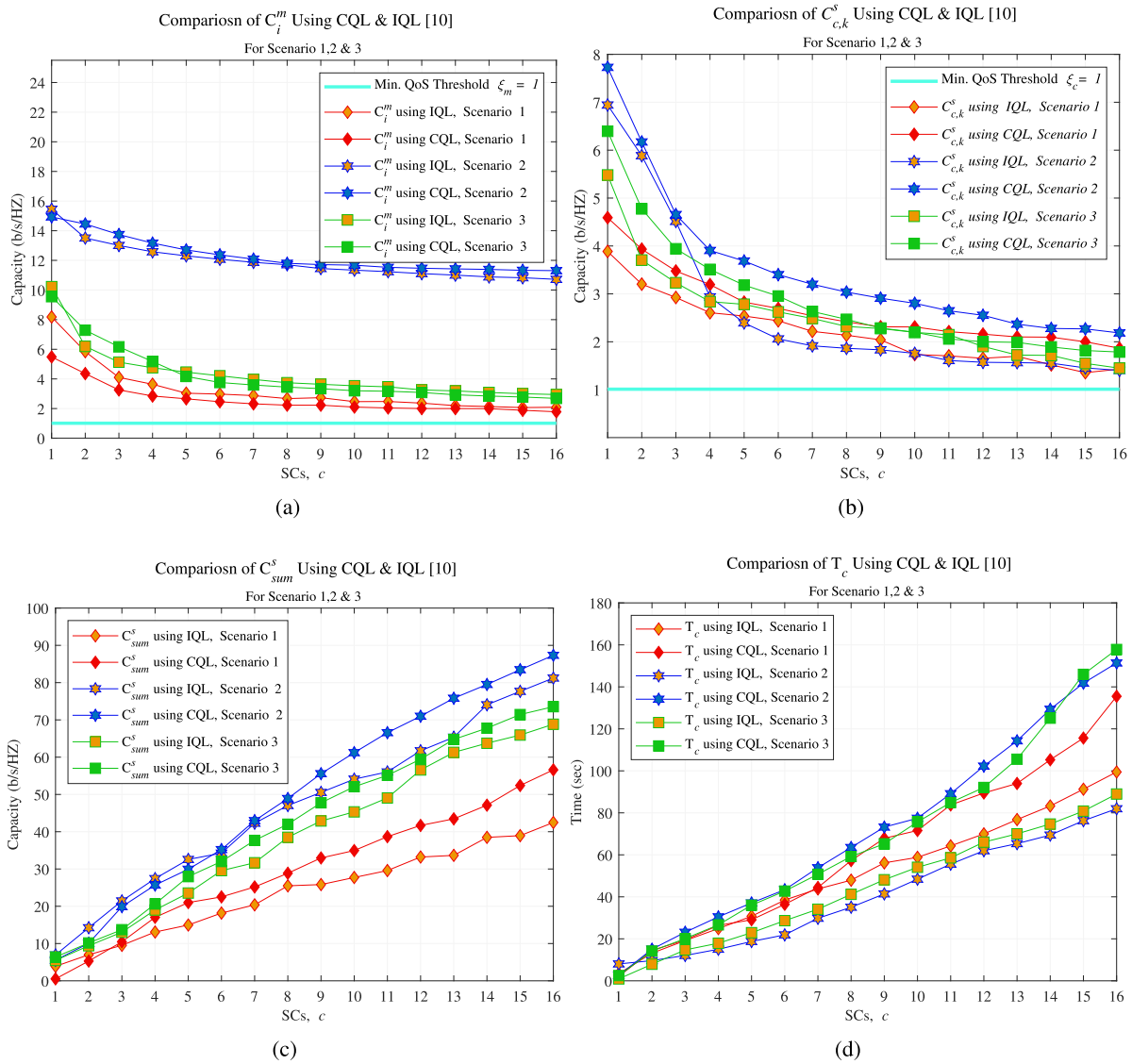

FIGURE 13. Comparison of simulation results for C_i^m, $C_{c,k}^c$, C_{sum}^s and T_c using the proposed CQL and IQL [10], (a) comparison of C_i^m using CQL and IQL [10] in scenario 1-3, (b) comparison of $C_{c,k}^s$ using CQL and IQL [10] in scenario 1-3, (c) comparison of C_{sum}^s using CQL and IQL [10] in simulation scenario 1-3, and (d) comparison of T_c using CQL and IQL [10] in scenario 1-3.

scenarios 2 and 3 but equal to IL-based algorithm in scenario 1. Therefore, cooperation among the SCs do not significantly impact the capacity of the UE^m which is also in line with results presented in the [25].

Fig.13b presents the performance comparison for the $C_{c,k}^s$ using CL and IL [10] based QL. In contrast to the C_i^m, there is a significant positive impact of CL on the $C_{c,k}^s$. In all three simulation scenarios, the CQL algorithm performed significantly better than the IQL algorithm in the same scenario as shown in Fig.13b. There is an improvement of 48%, i.e. 1.35 b/Hz/s to 2.0 b/Hz/S, using the CL as compared to IL.

The CL-based algorithm has improved the $C_{c,k}^s$, therefore, C_{sum}^s also improved significantly in all three simulation scenarios as shown in Fig. 13c. The minimum improvement in C_{sum}^s is for scenario 3 which is 7.4% and maximum improvement is in highest interference scenario 1 which is 38%.

The improvements in $C_{c,k}^s$ and C_{sum}^s using the CQL algorithm are at the cost of communication overhead and computational time, T_c. In the CQL, all the cooperating BS^s transmit and receive the entries of QT, therefore, computational time increases as compared to the IL paradigm. Despite the T_c of the proposed CL-based proposed QL algorithm is significantly less than the other recently proposed CL-based solutions in literature, Q-DPA [25], PA-DRL [26], and FQA [27], but IQL has slightly less T_c as compared to CQL as shown in the Fig. 13d. A similar trend is observed for T_c in all simulation scenarios of Fig.5 using CL and IL.

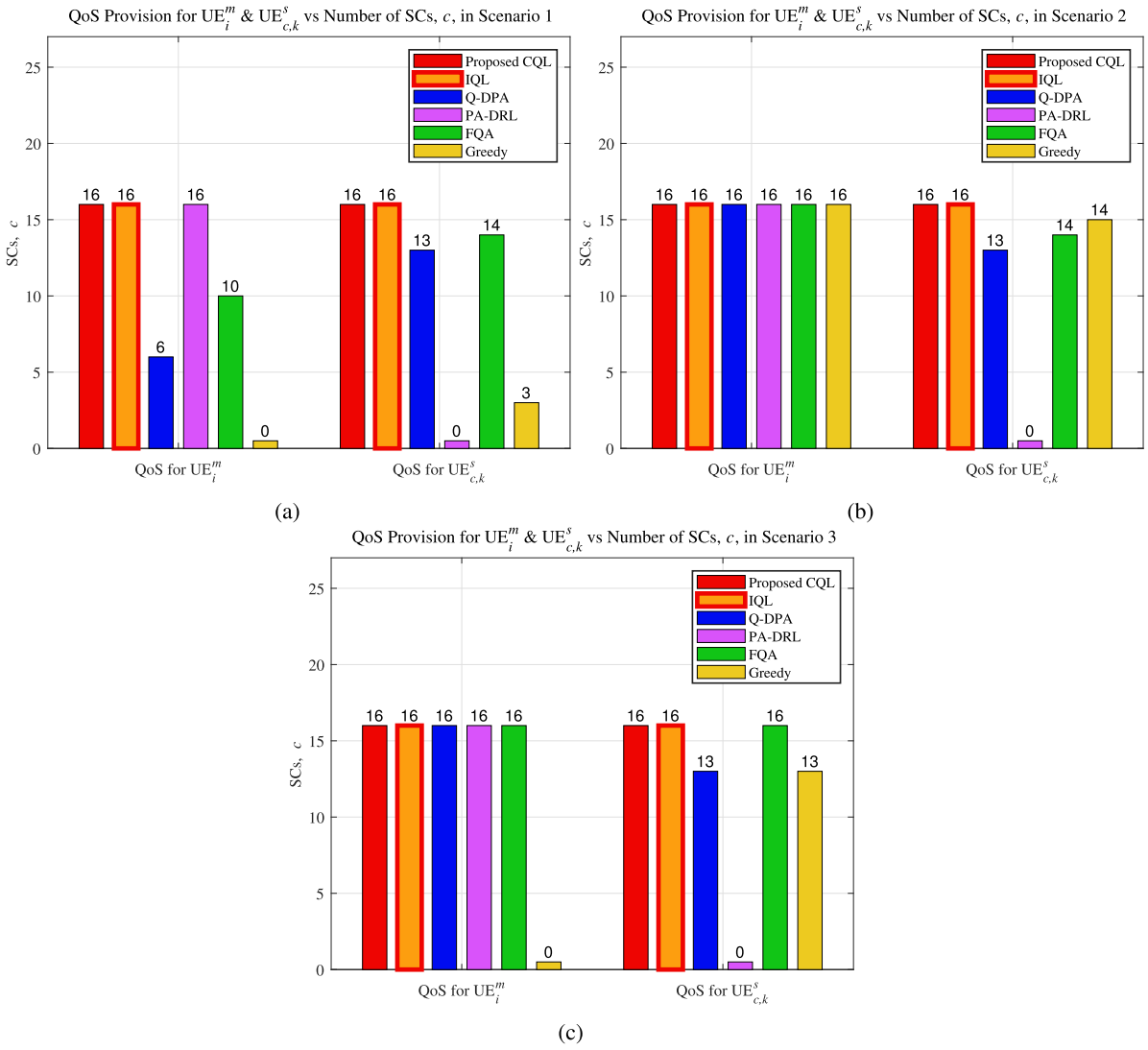

FIGURE 14. Comparison of QoS provision for UE$_i^m$ and UE$_{c,k}^s$ as function of c by proposed CQL, IQL [10], Q-DPA [25], PA-DRL [26], FQA [27], and non-adaptive greedy power allocation (a) simulation scenario 1, Fig.5a, (b) simulation scenario 2, Fig.5b, and (c) simulation scenario 3, Fig.5c.

9) QoS ANALYSIS

The results for QoS KPIs, C_i^m, and $C_{c,k}^s$ are presented in Fig. 6-Fig. 8 for simulation scenario 1-3, Fig.5a-Fig. 5c, are summarized in Fig. 14 for the proposed CQL algorithm, Q-DPA [25], PA-DRL [26], FQA [27], greedy power allocation and our previously proposed IL-based solution [10].

The results presented in Fig. 6-Fig. 8 shows that proposed CQL and previously proposed IQL [10] successfully provide QoS in terms of C_i^m, and C^c to all the 16 SCs in the cluster. However, the CQL algorithm outperforms IQL with a significant increase in $C_{c,k}^s$ and hence C_{sum}^s at the cost of slightly increased computational time, as discussed previously. On the other hand Q-DPA [25], PA-DRL [26], and FQA [27] and greedy power allocation could not meet the minimum QoS requirements for UE^m and UE^s simultaneously for the cluster of 16 SCs. In a very low interference scenario 3, [27] provided QoS requirement to both UE^m and UE^s but failed in scenarios 1and 2.

VIII. CONCLUSION

In this research article, we have explored the CQL algorithm for JRRM to provide QoS in ultra-dense HetNets for 5G and future CN by mitigating CoI and CrI simultaneously through adaptive power allocation in various interference scenarios based on 3GPP specifications. In the CQL algorithm, BS^s share their information of QT obtained through IL with the BS^s of neighboring SCs in the cluster and utilize each other's experience to learn an optimal policy. However, joint RF (JRF) is applied for optimal power allocation by all the cooperating BS^s in the cluster. The proposed CQL algorithm successfully mitigated the CoI and CrI and provided QoS to UE^m and UE^{s}s in the cluster of 16 SCs where other recently

proposed solutions in literature and greedy power allocation fail to meet the QoS requirements for both UE^m and UE^{s}s simultaneously. The proposed CQL provides C_i^m and $C_{c,k}^s$ nearly 2 b/s/Hz which is twice the minimum QoS threshold for UE^m and UE^s capacities, ξ_m and ξ_c respectively. In comparison to the IL paradigm, CL has no impact on UE^m's capacity in the case of ultra-dense SC HetNets. However, there is a significant improvement in UE^s's capacity, $C_{c,k}^s$, and sum capacity of the cooperating SCs in the cluster, C_{sum}^s. An increase of 48% and 34% is observed in $C_{c,k}^s$, and C_{sum}^s, respectively, using the CL as compared to IL. The increase in the $C_{c,k}^s$ and C_{sum}^s is at the cost of slightly increased computational time, T_c which is a function of the number of SCs, c, in the cluster. In this research, we simulated a cluster size of 16 SCs, 37.5% more SCs according to 3GPP TR36.872 by adding SCs in the cluster one by one. However, in the future, an optimal size of the cluster may be found to minimize the computational time in CL. Simulation results show that the proposed CQL algorithm not only outperformed other recently proposed algorithms and non-adaptive greedy power allocation but it proves its significance over the IL paradigm.

REFERENCES

[1] R. N. Mitra and D. P. Agrawal, "5G mobile technology: A survey," *ICT Exp.*, vol. 1, no. 3, pp. 132–137, Dec. 2015.

[2] N. Panwar, S. Sharma, and A. K. Singh, "A survey on 5G: The next generation of mobile communication," *Phys. Commun.*, vol. 18, pp. 64–84, Mar. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1874490715000531

[3] I. F. Akyildiz, S. Nie, S.-C. Lin, and M. Chandrasekaran, "5G roadmap: 10 key enabling technologies," *Comput. Netw.*, vol. 106, pp. 17–48, Sep. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128616301918

[4] A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.

[5] S. Manap, K. Dimyati, M. N. Hindia, M. S. Abu Talip, and R. Tafazolli, "Survey of radio resource management in 5G heterogeneous networks," *IEEE Access*, vol. 8, pp. 131202–131223, 2020.

[6] C. Niu, Y. Li, R. Q. Hu, and F. Ye, "Fast and efficient radio resource allocation in dynamic ultra-dense heterogeneous networks," *IEEE Access*, vol. 5, pp. 1911–1924, 2017.

[7] M. A. Adedoyin and O. E. Falowo, "Combination of ultra-dense networks and other 5G enabling technologies: A survey," *IEEE Access*, vol. 8, pp. 22893–22932, 2020.

[8] K.-L. A. Yau, J. Qadir, C. Wu, M. A. Imran, and M. H. Ling, "Cognition-inspired 5G cellular networks: A review and the road ahead," *IEEE Access*, vol. 6, pp. 35072–35090, 2018.

[9] T. O. Olwal, K. Djouani, and A. M. Kurien, "A survey of resource management toward 5G radio access networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1656–1686, 3rd Quart., 2016.

[10] M. U. Iqbal, E. A. Ansari, and S. Akhtar, "Interference mitigation in HetNets to improve the QoS using Q-learning," *IEEE Access*, vol. 9, pp. 32405–32424, 2021.

[11] P. Mach and Z. Becvar, "Energy-aware dynamic selection of overlay and underlay spectrum sharing for cognitive small cells," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4120–4132, May 2017.

[12] P. Zhang, X. Yang, J. Chen, and Y. Huang, "A survey of testing for 5G: Solutions, opportunities, and challenges," *China Commun.*, vol. 16, no. 1, pp. 69–85, Jan. 2019.

[13] A. Morgado, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, "A survey of 5G technologies: Regulatory, standardization and industrial perspectives," *Digit. Commun. Netw.*, vol. 4, no. 2, pp. 87–97, Apr. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352864817302584

[14] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A survey on 5G usage scenarios and traffic models," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 905–929, 2nd Quart., 2020.

[15] M. E. Morocho Cayamcela and W. Lim, "Artificial intelligence in 5G technology: A survey," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2018, pp. 860–865.

[16] *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN), Overall Description*, 3rd Generation Partnership Project (3GPP), Technical Report (Release8), document TS 36.300, Oct. 2020, version 16.3.0. [Online]. Available: https://portal.3gpp.org/

[17] A. Galindo-Serrano and L. Giupponi, "Distributed Q-learning for interference control in OFDMA-based femtocell networks," in *Proc. IEEE 71st Veh. Technol. Conf.*, May 2010, pp. 1–5.

[18] J. R. Tefft and N. J. Kirsch, "A proximity-based Q-learning reward function for femtocell networks," in *Proc. IEEE 78th Veh. Technol. Conf. (VTC Fall)*, Sep. 2013, pp. 1–5.

[19] B. Wen, Z. Gao, L. Huang, Y. Tang, and H. Cai, "A Q-learning-based downlink resource scheduling method for capacity optimization in LTE femtocells," in *Proc. 9th Int. Conf. Comput. Sci. Educ.*, Aug. 2014, pp. 625–628.

[20] H. Saad, A. Mohamed, and T. ElBatt, "Distributed cooperative Q-learning for power allocation in cognitive femtocell networks," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Sep. 2012, pp. 1–5.

[21] H. Saad, A. Mohamed, and T. ElBatt, "A cooperative Q-learning approach for online power allocation in femtocell networks," in *Proc. IEEE 78th Veh. Technol. Conf. (VTC Fall)*, Sep. 2013, pp. 1–6.

[22] J. R. Tefft and N. J. Kirsch, "Accelerated learning in machine learning-based resource allocation methods for heterogenous networks," in *Proc. IEEE 7th Int. Conf. Intell. Data Acquisition Adv. Comput. Syst. (IDAACS)*, Sep. 2013, pp. 468–473.

[23] R. Amiri and H. Mehrpouyan, "Self-organizing mm wave networks: A power allocation scheme based on machine learning," in *Proc. 11th Global Symp. Millim. Waves (GSMM)*, May 2018, pp. 1–4.

[24] R. Amiri, H. Mehrpouyan, L. Fridman, R. K. Mallik, A. Nallanathan, and D. Matolak, "A machine learning approach for power allocation in HetNets considering QoS," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–7.

[25] R. Amiri, M. A. Almasi, J. G. Andrews, and H. Mehrpouyan, "Reinforcement learning for self organization and power control of two-tier heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 3933–3947, Aug. 2019.

[26] Q. Su, B. Li, C. Wang, C. Qin, and W. Wang, "A power allocation scheme based on deep reinforcement learning in HetNets," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2020, pp. 245–250.

[27] W. AlSobhi and A. H. Aghvami, "QoS-aware resource allocation of two-tier HetNet: A Q-learning approach," in *Proc. 26th Int. Conf. Telecommun. (ICT)*, Apr. 2019, pp. 330–334.

[28] Z. Tang, W. Ji, and Q. Hu, "Optimal power allocation for multi-user linear network coded cooperation system," *IEEE Access*, vol. 7, pp. 7093–7103, 2018.

[29] G. Yu, R. Liu, Q. Chen, and Z. Tang, "A hierarchical SDN architecture for ultra-dense millimeter-wave cellular networks," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 79–85, Jun. 2018.

[30] M. Dirani, Z. Altman, and M. Salaun, "Autonomics in radio access networks," in *Autonomic Network Management Principles*, N. Agoulmine, Ed. New York, NY, USA: Academic, 2011, ch. 7, pp. 141–166. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123821904000073, doi: 10.1016/B978-0-12-382190-4.00007-3.

[31] H. Nouira. (Mar. 2015). *SON in LTE: The What, the Where, and the Why*. [Online]. Available: https://www.nokia.com/blog/son-lte-what-where-and-why/

[32] *Telecommunication Management; Self-Organizing Networks (SON); Concepts and requirements*, 3rd Generation Partnership Project (3GPP), (Release 8), document TR 32.500, Jul. 2020, version 12.1.0. [Online]. Available: https://portal.3gpp.org/

[33] *Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self Configuration and Self-Optimization Network Use Cases and Solutions*, 3rd Generation Partnership Project (3GPP), (Release8), document TR 36.902, Apr. 2011, version 9.3.1. [Online]. Available: https://portal.3gpp.org/

[34] M. Nohrborg. (Oct. 2020). *Self-Organizing Networks*. [Online]. Available: https://www.3gpp.org/technologies/keywords-acronyms/105-son
[35] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, and G. Ostrovski, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
[36] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2019, *arXiv:1509.02971*.
[37] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018. [Online]. Available: http://incompleteideas.net/book/the-book-2nd.html
[38] M. Tan, *Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents*. San Francisco, CA, USA: Morgan Kaufmann, 1997, pp. 487–494.
[39] *Small Cell Enhancements for E-UTRA and E-UTRAN–Physical Layer Aspects*, 3rd Generation Partnership Project (3GPP), (Release12), document TR 36.872, Dec. 2013, version 12.1.0. [Online]. Available: https://portal.3gpp.org/
[40] B. Abuhaija, "Performance analysis of LTE multiuser flat downlink power spectrum and radio resources scheduling," *J. High Speed Netw.*, vol. 18, no. 3, pp. 173–184, 2012.
[41] *Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects*, 3rd Generation Partnership Project (3GPP), (Release 9), document TR 36.814, Mar. 2017, version 9.2.0. [Online]. Available: https://portal.3gpp.org/
[42] C. Jin, Z. Allen-Zhu, S. Bubeck, and M. I. Jordan, "Is Q-learning provably efficient?" in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2018, pp. 4868–4878.
[43] K. Rastogi, J. Lee, F. Harel-Canada, and A. Joglekar, "Is Q-learning provably efficient? An extended analysis," 2020, *arXiv:2009.10396*.
[44] M. Liu, "Optimal number of trials for Monte Carlo simulation," *Valuation Res. Rep.*, 2017. [Online]. Available: https://www.valuationresearch.com/wp-content/uploads/2019/11/SpecialReport_MonteCarloSimulation Trials-11-19.pdf
[45] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," 1998, *arXiv:cs/9809099*.

MUHAMMAD USMAN IQBAL received the B.Sc. degree from the University College of Engineering and Technology (UCE&T), Bahauddin Zakariya University, Multan, Pakistan, in 2009, and the M.S. degree from the School of Electrical Engineering and Computer Science (SEECS), National University of Science and Technology (NUST), Islamabad, Pakistan, in 2013, both in electrical engineering. He is currently pursuing the Ph.D. degree in electrical engineering with the Department of Electrical and Computer Engineering (ECE), COMSATS University Islamabad, Lahore Campus, Pakistan. He served at the University College of Textile Engineering (UCTE), Bahauddin Zakariya University, as a Lecturer in electrical engineering, from June 2009 to June 2014. He has been working as a Lecturer with the Department of Electrical and Computer Engineering (ECE), COMSATS University Islamabad, Lahore Campus, since July 2014. His research interests include digital signal processing, digital image processing, wireless communication, and deep learning.

EJAZ AHMAD ANSARI received the B.Sc. degree (Hons.) from the University of Engineering and Technology (UET), Lahore, in 1990, and the M.Sc. degree from the Georgia Institute of Technology (G Tech), Atlanta, USA, in 1995, both in electrical engineering, the M.B.A. degree in marketing from the University of the Punjab (PU), Lahore, in 2000, and the D.Eng. degree in telecommunications engineering from the School of Engineering and Technology (SET), Asian Institute of Technology (AIT), Bangkok, Thailand, in 2009. Earlier, he served at the Water and Power Development Authority (WAPDA), Pakistan, from April 1991 to August 2000. He served at the Lahore University of Management and Sciences (LUMS) as a Lecturer, from September 2000 to December 2002. He served at the Department of Electrical Engineering, COMSATS Institute of Information Technology (CIIT), Lahore, as an Assistant Professor, from January 2003 to June 2014. Since July 2014, he has been serving with the Department of Electrical and Computer Engineering, CUI, as an Associate Professor. He has been working as the Head of the Department of Electrical and Computer Engineering (ECE), COMSATS University Islamabad (CUI), Lahore Campus, since June 2020. His research interests include multirate signal and image processing and their modeling, performance analysis of wireless networks, and communication theory. He is a Lifetime Member of Pakistan Engineering Council (PEC), Pakistan, and an IEEE Reviewer of *Wireless Sensor Networks*.

SALEEM AKHTAR received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 1991, the D.E.A. degree in digital telecommunication systems from the École Nationale Supérieure des Télécommunications, Paris, France, in 1997, and the Ph.D. degree in mobile CN from the École Nationale Supérieure des Télécommunications, Paris, in 2001. From December 1997 to July 2001, he was a Research Associate with the Network and Services Department, Institut National des Télécommunications, Paris. From October 2001 to September 2002, he was a Research Fellow with the Network and Services Department, Institut National des Télécommunications. He is currently working as a Principal Engineer with the Department of Electrical and Computer Engineering, COMSATS University Islamabad, Lahore. His primary research interests include quality of service (QoS) provisioning and radio resource management in heterogeneous wireless networks.

ALI NAWAZ KHAN (Member, IEEE) received the B.S. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2003, and the Ph.D. degree in information and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2008. He is currently an Assistant Professor with the Electrical and Computer Engineering Department, COMSATS University Islamabad, Lahore. He is the Head of the Wireless Sensor Networks Research Group and supervises graduate and postgraduate research in the areas of mobile networks, mobile healthcare applications, wireless sensor networks, and energy efficient MAC protocols. He is an active reviewer of several internationally abstracted journals and a Program Committee Member of Frontiers of IT Conference (2009 onwards).