IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 21 June 2023, accepted 8 July 2023, date of publication 19 July 2023, date of current version 26 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3297099

RESEARCH ARTICLE

AI-Enabled Energy-Aware Carrier Aggregation in 5G New Radio With Dual Connectivity

FAHIME KHORAMNEJAD[1], ROGHAYEH JODA[2], (Senior Member, IEEE), AKRAM BIN SEDIQ[2], GARY BOUDREAU[2], (Senior Member, IEEE), AND MELIKE EROL-KANTARCI[1], (Senior Member, IEEE)

[1]School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada
[2]Ericsson Canada, Ottawa, ON K2K 2V6, Canada

Corresponding author: Melike Erol-Kantarci (melike.erolkantarci@uottawa.ca)

This work was supported in part by NERSC Canada Research Chairs program, in part by MITACS, and in part by Ericsson Canada.

ABSTRACT Aggregating multiple component carriers (CCs) from different frequency bands, also known as Carrier Aggregation (CA), and Dual Connectivity (DC), i.e., concurrently transmitting and receiving from two nodes or cell groups, are employed in 5G and 6G wireless networks to enhance coverage and capacity. In wireless networks with DC and CA, the performance can be boosted by dynamically adjusting the uplink (UL) transmit power level for the user equipments (UEs) and properly activating/deactivating the CCs for the UEs. In this paper, we study the problem of joint dynamic UL power-sharing and CC management. The objective is to simultaneously minimize the delay and power consumption for the UEs. The pertinent problem is a multi-objective optimization problem with both discrete and continuous variables and therefore is hard to solve. We first model it as a multi-agent reinforcement learning (RL) system with compound action to handle the problem. Then, we employ a compound-action actor-critic algorithm to find the optimal policy and propose the Joint Power-Sharing and Carrier Aggregation (JPSCA) algorithm. The performance of the JPSCA algorithm is compared with two baseline algorithms. Our results show that the performance of the JPSCA algorithm in terms of the average rate, delay, and UL transmit power level outperforms the baselines where UL power control and CC management are performed disjointly. For 25 UEs, our proposed JPSCA algorithm decreases the UE power consumption and UE delay by about 28% and 16%, respectively, concerning the all-CC and equal power-sharing schemes.

INDEX TERMS 5G, carrier aggregation, dual connectivity, dynamic uplink power-sharing, reinforcement learning.

I. INTRODUCTION

Supporting a wide range of services, from the enhanced mobile broadband (eMBB) applications to the ultra-reliable low-latency communications (URLLC), 5G resorts to carrier aggregation (CA) and dual connectivity (DC) to boost its capacity and provide enhanced coverage under multiple bands. In CA, the spectrum resources in the form of component carriers (CCs) in different frequency bands are aggregated to increase the capacity of the network and the data rates. Meanwhile, DC or, more generally, multiple connectivity (MC) allow operators to provide LTE and 5G connectivity simultaneously. Combining CA and DC can be an attractive solution for enhancing wireless communication performance such as providing faster data rates, better network efficiency, and improved user experience. In 3GPP Release 15 and thereon, this technology is called E-UTRAN New Radio – Dual Connectivity (EN-DC), where LTE is used as a master node and NR is used as a secondary node. In EN-DC, user equipment (UE) is allowed to connect to an LTE-evolved node B (eNB) simultaneously and a 5G next-generation (NR) node B (gNB), i.e., separately transmitting/receiving LTE and 5G signals and then aggregating the streams. The UEs can enormously benefit from employing joint CA and DC technologies to increase the coverage area, and system throughput, load balancing and mobility robustness [1]. Meanwhile,

The associate editor coordinating the review of this manuscript and approving it for publication was Zaharias D. Zaharis.

these emerging technologies increase complexity and could pose challenges to derive beneficial resource management schemes.

One of the significant challenges arising in uplink (UL) transmission in the EN-DC network is dividing the UL transmit power level between the eNB and gNB, subject to the maximum power budget for the UE. In the literature, [2], [3] two different schemes, Dynamic Power-Sharing (DPS) and Equal Power-Sharing (EPS), have been proposed to adjust the UL transmit power level. The UL transmit power levels to both eNB and gNB are dynamically adjusted in DPS. While, in EPS, the UL transmit power level is equally divided between the eNB and the gNB. Employing EPS may result in poor network performance because the time-varying characteristics for the channels and UE traffic in the 5G networks would not be taken into consideration [3]. This motivates us to focus on the DPS. Specifically, the novelty of this work is tackling the problem of the joint CA and dynamic adjustment of UL transmit power, which has not been addressed before, to the best of our knowledge. This problem contains continuous and discrete variables, UL transmit power level and the indicator variables for activating/deactivating the CCs, respectively. Furthermore, 5G has multi-dimensional and dynamic characteristics. These features motivate us to employ a machine learning (ML) based method, i.e., reinforcement learning (RL), to address the problem of joint UL power control and CA. One may employ some optimization and game theory methods to tackle the problem; however, this can result in a lack of scalability and flexibility. In the following subsection, we first briefly review the most recent related studies and then introduce our contributions and techniques used to address the problem of joint power-sharing and CA.

A. RELATED WORKS

This subsection briefly reviews the most recently proposed resource management schemes in the 5G networks with CA and DC. In the networks with CA technology, some resource management schemes have been derived in [4], [5], [6]. The authors in [4] deploy RL and optimiztion theory to tackle the problem of CC management. The objective is to simultaneously minimize the UE delays and the energy consumed by the UEs to activate/deactivate the secondary component carriers (SCCs). In [5], it is assumed that more than one cell operator can serve the mobile devices, and they can aggregate their CCs. Specifically, a two-layer interacting game is formulated to maximize the system throughput. The upper layer comprises a Stackelberg game and is responsible for adjusting the spectrum price. Meanwhile, the lower game comprises a bargaining game and allocates the spectrum resources to the users. Subject to the constraint of high data rate requirements for the device-to-device (D2D) links, the problem of minimizing the total consumed power has been addressed in [6]. Specifically, the CA technology has been developed in D2D-enabled 5G networks to satisfy the rate requirements for the D2D links and the resource management problem is formulated as a mixed-integer optimization problem. The transformation and variable substitution have been used to address the problem. A two-layer algorithm has been proposed for joint UL power allocation and carrier aggregation.

DC technology was first initiated in standard 3GPP Released 12 for the LTE networks. To employ the non-standalone 5G networks so as to simultaneously connect to the 5G and LTE, the standard 3GPP Released 15 has been developed, which is analyzed in [7]. In the networks with DC technology, some resource management problems have been tackled by the authors in [8], [9], [10], and [11]. To maximize energy efficiency in a heterogeneous network (HetNet) with DC, a joint power control and traffic offloading scheme has been derived in [8]. In [9], in a multi-access edge computing (MEC) enabled network with DC technology, the problem of minimizing the total energy consumption is formulated as a mixed-integer non-linear programming (MINLP) problem. The MINLP problem is first addressed by using some concepts from optimization theory. Then, deep learning is used to derive an intelligent offloading scheme. In a MEC-enabled HetNet with DC technology, to provide the edge devices with adequate resources, sub-6 GHz and mmWave BSs are employed in [10], and the benefits of using the corresponding links for the reliable delivery of the virtual reality (VR) traffic have been studied. In [11], the DC has been used in a UAV-assisted HetNet. The UAVs are responsible for controlling reconfigurable intelligent surfaces (RISs)-providing a strong line-of-sight (LOS) connection with the ground users by operating on microwave channels in the sky. By considering orthogonal multiple access (OMA) over the microwave channel for the macro BSs and non-orthogonal multiple access (NOMA) over the mmWave channel for the small BSs, the authors formally express the problem of minimizing the total DL transmit power level for the macro and small BSs. The problem is decomposed into two subproblems. The former is solved by deriving an intelligent dueling deep Q-Network-based algorithm, and the latter subproblem is tackled by using successive convex approximation in optimization theory.

In an EN-DC network consisting of one eNB and one gNB, to tackle the problem of UL power-sharing, the authors in [3] quantized the continuous UL transmit power levels for the users (i.e., by approximating power levels with amplitudes restricted to a prescribed set of values). Then the Q-learning algorithm [12] was used to address the problem of dynamic power-sharing. In [3], the CA technology has not been considered. Additionally, the Q-learning-based algorithm may not perform efficiently as the action space grows, such as by increasing the quantization levels or expanding the number of users.

Recently, to increase the system throughput and coverage, both DC (or MC) and CA technology have been used in the wireless networks [1], [13]. For instance, in [1], the authors derive a heuristic UE-BS association and CA scheme for the load (i.e., the number of assigned UEs to the BSs) balancing

in the networks. Specifically, a UE selects its serving primary cell based on either received reference signal power or received reference signal quality. Then, a CC management scheme is applied for CA and secondary cell selection. In the LTE HetNets, the problem of jointly maximizing the spectrum efficiency and energy efficiency has been formulated as a bi-objective optimization problem in [13]. The authors derive a resource management scheme for CA, downlink (DL) power control, and user association using some optimization theory concepts.

In the wireless EN-DC networks, the problem of joint UL power control and CC management has not been studied yet. Motivated by this, different than other works, we will formulate and address the problem of joint UL power-sharing, i.e., adjusting the UL transmit power levels for the UEs to eNB and gNBs, and CA for 5G networks.

B. MOTIVATION AND CONTRIBUTIONS

Resource management problems in wireless networks are sequential decision-making problems, and thus, RL has been recently used to address them. Specifically, deep RL (DRL) based methods are employed to tackle the problems in wireless networks with a huge action/state space. In [14] single and multi-agent DRL-based resource management schemes in the AI-enabled networks are surveyed. Additionally, the machine learning-based methods that are used to improve the wireless network performance are studied in [15].

Due to the multi-dimensional and dynamic characteristics of 5G networks, modeling the EN-DC wireless networks with CA presents some challenges. Previously, to deal with the resource management problems in the wireless networks with DC and CA technology, either optimization theory-based [13] or heuristic algorithms [1] were employed. These two methods require complete knowledge about the environment and precisely studying the optimal solution gap, respectively.

In this paper, we consider an EN-DC network with CA technology. In 5G/6G wireless networks, CA and DC are employed to enhance the UE throughput and the network coverage. The more achievable the rate is for the UEs, the less the UE delay, which is obtained at the price of the energy consumed to monitor and activate/deactivate the CCs. Thus, we focus on the problem of jointly minimizing the delay and the energy consumption in EN-DC networks with CA. Furthermore, based on [16], we perform CC management and RB allocation disjointly [4]. Specifically, we focus on dealing with the problem of CA and dynamic power sharing. To the best of our knowledge, this problem has not been studied yet. To allocate the RBs in the CCs, we deploy the round-robin (RR) scheduling. We use multi-agent RL and derive a UL resource management scheme to adjust UE UL transmit power levels to the eNB and gNBs. Specifically, by employing the multi-agent compound action actor-critic (CA2C) method in [17], we propose an intelligent joint power control and CC management scheme to minimize the delay and the total power consumption of UEs. Our main contributions are:

1) PROBLEM FORMULATION

We formally express a resource management problem to tackle the main challenges in 5G wireless networks with CA and DC, such as complex radio resource management, interference management, and compatibility with LTE networks. Subject to the constraint of QoS requirements for the UEs and the constraint on UL transmit power level, we formulate the problem of joint CA and UL power sharing in EN-DC networks as a multi-objective optimization problem where the gNBs concurrently try to optimize their objective function. Specifically, taking care of interference from other gNBs, a gNB aims at jointly minimizing the total delay for the UEs in its coverage area and minimizing their total UL power consumption. The optimized variables correspond to activating a CC for a UE, and adjusting the UL transmit power levels for each UE to the BSs. The first set of optimizing variables are discrete variables, and the second one is continuous.

2) DEVELOPING THE RL SYSTEMS

Having two different sets of discrete and continuous variables, the formulated multi-objective optimization problem has the combinatorial characteristic and thus is hard to solve. For CC management and UL power control, we propose two approaches. In the first approach, we simultaneously adjust the UL transmit power levels and activate/deactivate a CC, and we model the problem as a multi-agent RL problem with compound action space composed of both discrete and continuous actions. The adopted actions correspond to CC activation and the UL transmit power level selection for the UEs to eNB and gNBs. In the second approach, we consider an extreme case that all CCs are activated for the UEs at the NR side. By quantizing the continuous UL transmit power levels for the UEs, we develop a multi-agent RL system with discrete action space. The action space is pertinent to quantized UL transmit power levels for a UE. In both already mentioned RL systems, the gNBs are the learning agents. Further, the state space is given in terms of the UE target (tolerable) delay requirements. The reward (cost) for a given gNB is given in terms of the total delay and the power consumption for the UEs in its coverage area..

3) DERIVING THE INTELLIGENT ALGORITHMS

To address the formulated RL problem with a compound action space, we utilize the CA2C algorithm [17] and introduce a scheme called the Joint Power-Sharing and Carrier Aggregation (JPSCA) algorithm. In addition, we consider a scenario where the UL transmit power levels for the UEs are quantized, leading us to develop a RL system with discrete actions. The RL system with a discrete action space is tackled using the double deep Q-Network (DDQN) algorithm [18], and we propose an algorithm based on DDQN that adjusts both the UL transmit power levels for the UEs and CA. This algorithm is referred to as the Discrete Joint Power-Sharing and Carrier Aggregation (DJPCA) algorithm. To evaluate the performance of the proposed intelligent algorithms,

TABLE 1. Table of notations.

Notation	Description
$\mathcal{B}, \mathcal{M}, \mathcal{M}^k$	Set of gNB, set of CCs, set of CCs for UE k
$\widehat{\mathcal{M}} = \{1\}, \widetilde{\mathcal{M}}$	Set of CC on LTE side, set of CCs on NR side
$\mathcal{K}, \mathcal{K}_b$	Set of UEs, set of UEs in gNB b
$b(k), \mathcal{N}_m$	Serving gNB for UE k, set of RBs in CC m
$\widehat{h}^k_{1,n}$	Path gain between the UE $k \in \mathcal{K}_b$ and eNB over a RB n in PCC (in the LTE side)
$\widetilde{h}^{b(k),k}_{m,n}$	Path gain between that UE and gNB $b(k)$ over a RB n in CC m (PCC or SCC at NR side)
$\widehat{p}^k, \widehat{p}^k_{1,n}$	UL transmit power levels for UE k to the eNB and the one for UE k on RB $n \in \mathcal{N}_1$
$\widetilde{p}^k, \widetilde{p}^k_{m,n}$	UL power levels for UE k to the gNB $b(k)$ and the one for UE k on RB $n \in \mathcal{N}_m$
$\widehat{R}^k, \widehat{R}^k_{1,n}$	UL rate for UE k achieved by transmitting to eNB and the one on RB $n \in \mathcal{N}_1$
$\widetilde{R}^k, \widetilde{R}^k_{m,n}$	UL rate for UE k achieved by transmitting to gNB $b(k)$ on the one on RB $n \in \mathcal{N}_m$
p^k, R^k	Total UL power for UE k, total rate for UE k
$\overline{D}^k, P^k$	Average delay and UL power consumption for UE k
$\mathrm{D}^k_{\mathrm{QoS}}, p^k_{\max}$	Max. tolerable delay and Max. UL power level for UE k
$\widehat{q}^k(t)$	Average number of bits per burst of data for UE k
$\alpha^k_m, \beta^k_{m,n}$	SCC activation indicator for UE k, and RB allocation indicator for UE k in CC $m \in \mathcal{M}$
$\hat{r}^k, \tilde{r}^k$	Continuous action used to adjust the UL power levels for UE k to the eNB and gNB $b(k)$

we compare them with a baseline algorithm where all CCs are activated for a UE, and the UL transmit power levels for a UE to both eNB and gNB are equally split from the total UL power budget of the UE.

In the rest of this paper, we first introduce the system model and problem statement in Section II. The distributed multi-agent RL-based schemes are proposed in Sections III and IV. The simulation results and conclusion are given in Sections V and VI, respectively.

II. SYSTEM MODEL AND PROBLEM STATEMENT

In this work, we consider an EN-DC wireless network with CA technology. In CA, the bandwidth is expanded by aggregating some CCs in the same/different frequency bands. A CC consists of several RBs, and thus the problem of RB allocation and CA are coupled. The details for the network model and problem formulation are given in what follows. The notations are provided in Table 1.

A. NETWORK AND NOTATIONS

In an EN-DC network consisting of an eNB and a set of B NR next generation NodeBs (gNBs) denoted by $\mathcal{B} = \{1, \dots B\}$, we consider the UL transmission where gNBs can use CA. Let us denote the set of non-overlapping and orthogonal CCs adopted by the eNB and the gNBs by $\widehat{\mathcal{M}} = \{1, \dots \widehat{M}\}$ and $\widetilde{\mathcal{M}} = \{\widehat{M}+1, \dots, \widehat{M}+\widetilde{M}\}$, respectively. Specifically, based on [19], we assume that $\widehat{M} = 1$, i.e., the eNB does not use the CA and the CCs in $\widehat{\mathcal{M}}$ are shared among the gNBs. Let $\mathcal{M} = \widehat{\mathcal{M}} \bigcup \widetilde{\mathcal{M}} = \{1, \dots 1+\widetilde{M}\}$. For a given $m \in \mathcal{M}$, we denote $\mathcal{N}_m = \{1, \dots, N_m\}$ as the set of RBs in CC m. Let $\mathcal{M}^k = \{1, \dots M^k\}$ be the set of activated CCs for UE k. $\mathcal{M}^k$ consists of both primary CCs (PCCs) and SCCs for UE k. The PCC is the main carrier assigned and it can be updated either during handover or selecting a cell. Meanwhile, a SCC is an auxiliary CC that can be activated/deactivated at any time to boost the achievable data rate [20], [21]. The functionality of activating/deactivating the SCCs are shown in [4]. Specifically, activating a SCC can enhance the network performance in terms of increasing the achievable rate for each UE and decreasing the UE delays. This enhancement is obtained at the price of consuming more energy to monitor the CCs [4].

In EN-DC networks, a UE is simultaneously connected to the eNB and a gNB. Additionally, there are actually two PCCs, one on LTE side and one on NR side. Therefore, we assume that the CC in the eNB is an always-activated PCC for UE k and the other PCC is in its serving gNB. The PCCs on LTE and NR sides are in low-band and high-band frequencies, respectively. Furthermore, as already mentioned, the eNB does not use the CA, and the RBs in the always-activated PCC on LTE side are exclusively assigned to the UEs, and thus the UEs are not interfering with each other on LTE side. On the NR side, a CC can be shared among the UEs served by a gNB; however, the RBs in the CC are exclusively allocated to the UEs. On the other hand, the UEs in different 5G cells may reuse the RBs in the CCs, and thus they would interfere with each other through the assigned RBs in their PCCs and SCCs.

Let us assume that a set of K UEs, $\mathcal{K} = \{1, \dots, K\}$, is distributed in the network area. We denote the serving gNB for UE k by $b(k)$. Indeed, UE k is in the coverage area for the gNB $b(k)$. Given $\mathcal{K}_b = \{1, \dots, K_b\}$ as the set of UEs in the coverage area for the gNB b, by employing DC technology, a UE $k \in \mathcal{K}_b$ can simultaneously transmit to both eNB and gNB b. Let $\widehat{h}^k_{1,n}$ and $\widetilde{h}^{b(k),k}_{m,n}$ be the path gain between the UE $k \in \mathcal{K}_b$ and eNB over a RB n in PCC (in the LTE side) and the path gain between that UE and its serving gNB $b(k)$ over a RB n in CC m (PCC or SCC at NR side), respectively. For any gNB b and the eNB, we assume that the channel between UE k and gNB b, and the channel between the UE and the eNB have both small and large fading with path loss and shadowing. Also, the channels are time-frequency varying ones.

Our network model and proposed scheme have been briefly illustrated in Fig. 1. Specifically, to activate/deactivate the SCCs, adjust the UL transmit power level to the eNB and gNBs, and allocate the RBs, we develop a scheme with different execution intervals. CA and UL power allocation as the delay insensitive tasks are performed every $\mathrm{T}_{\mathrm{cca}}$, whereas RB allocation, as a delay-sensitive task, is performed every time interval t where $t \leq \mathrm{T}_{\mathrm{cca}}$. Furthermore, RB allocation and CA are performed separately [16]. In this work, we derive a compound-action actor-critic-based algorithm for joint CA and UL power allocation. For RB allocation, we employ round-robin algorithm. The details are given in Sections III.

We first introduce the model for UL transmit power level, UL power consumption, UL throughput and delay. Then, using the models, we conclude this section by formulating

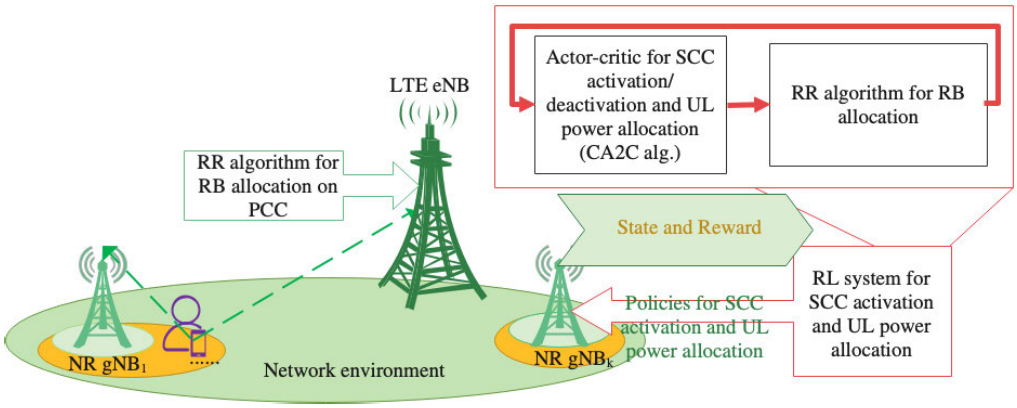

FIGURE 1. Intelligent CA and UL power allocation in the network with DC technology.

the problem of joint CA and UL power allocation in EN-DC networks.

B. MODELS FOR TRANSMIT POWER LEVEL, POWER CONSUMPTION, THROUGHPUT, AND DELAY

1) UL TRANSMIT POWER MODEL FOR UE

In time t, we denote the total transmit power level for UE k by $p^k(t)$:

$$p^k(t) = \widehat{p}^k(t) + \widetilde{p}^k(t). \tag{1}$$

In (1), $\widehat{p}^k(t)$ and $\widetilde{p}^k(t)$ are the UL transmit power level for UE k to the eNB and its serving gNB, respectively. At the time t, let $\widehat{p}^k_{1,n}$ be the UL transmit power level for UE k through the RB $n \in \mathcal{N}_1$ and $\widetilde{p}^k_{m,n}$ be the UL transmit power level for the UE k through RB $n \in \mathcal{N}_m$ where m is an activated SCC for UE k. We have $\widehat{p}^k(t) = \sum_{n\in\mathcal{N}_1} \beta^k_{1,n}\widehat{p}^k_{1,n}(t)$ and $\widetilde{p}^k = \sum_{m\in\widetilde{\mathcal{M}}} \alpha^k_m \sum_{n\in\mathcal{N}_m} \beta^k_{m,n}\widetilde{p}^k_{m,n}(t)$ where α^k_m is the SCC activation indicator for UE k and $\beta^k_{m,n}$ is the RB allocation indicator for UE k in CC $m \in \mathcal{M}$. Additionally, for convenience, from here, we assume that for the always activated PCC m at NR side we have $\alpha^k_m = 1$. The total UL transmit power level for the UE k is bounded, i.e.,

$$p^k(t) \le p^k_{\max}, \tag{2}$$

where $p^k_{\max}$ is the maximum UL transmit power level for UE k.

2) UL POWER CONSUMPTION MODEL FOR UE

For studying the power consumption in UL transmission, the authors in [22] employ the model developed in [23]. As shown in [23], the UL power consumption for UE k depends on its operating frequency band which is characterized by some parameters. Additionally, it depends on UL transmit power level for the UE. In this work, we use the UL power consumption in [23] given by,

$$P^k(t) = \begin{cases} \Theta_L p^k(t) + \Lambda_L & \text{if } p^k(t) \le \Gamma, \\ \Theta_H p^k(t) + \Lambda_H & \text{if } p^k(t) > \Gamma. \end{cases} \tag{3}$$

In (3), the parameters Θ_L, Θ_H, Λ_L, Λ_H, and Γ are the device-based ones and depend on the operating frequency band [23].

3) UL THROUGHPUT MODEL FOR UE

At time t, let $\widehat{\gamma}^k_{1,n}(t)$ denote the SINR for the UE k with eNB over RB n in PCC. Therefore we have $\widehat{\gamma}^k_{1,n}(t) = \frac{\widehat{p}^k_{1,n}(t)\widehat{h}^k_{1,n}(t)}{\sigma^2}$, where σ^2 is the noise. It is noteworthy that the RBs in the low-band PCC at LTE side is exclusively allocated to the UEs, and thus the UEs are not interfering with each other through the RBs in the PCC. Additionally, $\widetilde{\gamma}^k_{m,n}(t)$, the SINR for the UE k with its gNB (i.e., $b(k)$) over RB n in CC m (PCC or SCC), is given by, $\widetilde{\gamma}^k_{m,n}(t) = \frac{\widetilde{p}^k_{m,n}(t)\widetilde{h}^{b(k),k}_{m,n}(t)}{\sum_{k'\in\mathcal{K},k'\neq k} \beta^{k'}_{m,n}\widetilde{p}^{k'}_{m,n}(t)\widetilde{h}^{b(k),k'}_{m,n}(t) + \sigma^2}$. Accordingly, employing the Shannon theorem, we denote the achievable rate for UE k on RB n in PCC by $\widehat{R}^k_{1,n}$ and formally state them as, $\widehat{R}^k_{1,n}(t) = \widehat{B}_{1,n}\log_2\left(1+\widehat{\gamma}^k_{1,n}(t)\right)$ where $\widehat{B}_{1,n}$ is the bandwidth for RB n in PCC at LTE side. Similarly, the achievable rate for UE k on RB n in CC m at NR side is denoted by $\widetilde{R}^k_{m,n}$ and expressed as, $\widetilde{R}^k_{m,n}(t) = \widetilde{B}_{m,n}\log_2\left(1+\widetilde{\gamma}^k_{m,n}(t)\right)$ where $\widetilde{B}_{m,n}$ is the bandwidth for RB n in CC m at NR side. Let $\widehat{R}^k(t)$ and $\widetilde{R}^k(t)$ be the UL rate for UE k achieved by transmitting to eNB and gNB, respectively. By denoting $R^k(t)$ as the total achievable rate for UE k, we have,

$$R^k(t) = \widehat{R}^k(t) + \widetilde{R}^k(t) \tag{4}$$

where $\widehat{R}^k(t) = \sum_{n\in\mathcal{N}_1} \beta^k_{1,n}\widehat{R}^k_{1,n}(t)$ and $\widetilde{R}^k(t) = \sum_{m\in\widetilde{\mathcal{M}}} \alpha^k_m \sum_{n\in\mathcal{N}_m} \beta^k_{m,n}\widetilde{R}^k_{m,n}(t)$.

4) DELAY MODEL FOR UE

As aforementioned, CA is employed at T_{cca} intervals and primarily impacts the UE delays. Hence, the delay for a UE in duration $[t, t+T_{cca}]$ can be given in terms of the time to deliver both remaining bursts of data, and the ones estimated to be in the scheduling queue [4]. Let $\overline{D}^k(t)$ be the average

delay for UE k in time t. In [4], we developed the UE delay model and obtained $\overline{D}^k(t)$ as follows,

$$\overline{D}^k(t) = \frac{\widehat{q}^k(t)}{R^k(t)}, \tag{5}$$

in which, $R^k(t)$ is the UL achievable rate for the user in current time t (which is given in (4)), and

$$\widehat{q}^k(t) = \widehat{q}_c^k(t) \times \left\lfloor \frac{t + T_{cca} - T_f^k}{\widehat{T}^k(t)} \right\rfloor + \overline{q}_r^k(t) \times \widehat{N}_r^k(t). \tag{6}$$

In (6), $\widehat{q}^k(t)$ is the average number of bits for UE k predicted to be transmitted in the current duration, $\overline{q}_r^k(t)$ is the average number of bits in one burst of data from the previous duration, and $\widehat{N}_r^k(t)$ is the number of bursts of data remaining in the scheduling queue from the previous duration. Additionally, $\widehat{q}_c^k(t)$ is the estimated average number of bits in one burst of data arriving at the current duration. $\widehat{T}^k(t)$ is the interval time between the two consecutive bursts and T_f^k is the time that the last burst of payloads of UE k has been arrived. Thus, $\frac{t+T_{cca}-T_f^k}{\widehat{T}^k(t)}$ is the estimation of the average number of bursts arriving at the current duration.

In this work, we state the QoS requirement for a UE k in terms of the maximum delay which can be tolerated by the UE. It can be given by,

$$\frac{\widehat{q}^k(t)}{R^k(t)} \leq \mathrm{D}_{\mathrm{QoS}}^k, \tag{7}$$

where $\mathrm{D}_{\mathrm{QoS}}^k$ is the maximum tolerable delay for UE k. We say UE k meets its QoS requirement if (7) holds for that UE.

C. PROBLEM STATEMENT

In CA, a UE sends and receives through multiple CCs, i.e., PCCs and SCCs, from a single node which can be either an eNB or a gNB. Meanwhile, EN-DC enables a UE to concurrently send and receive its data through the CCs from a master eNB and a secondary gNB.

This work will address the problem of joint UL power control and CA in 5G networks with EN-DC. For the UL transmission, based on [24], we describe the scenario where the eNB selects a CC in a low-frequency band, and a gNB selects three CCs in the mid-band frequency. There are two PCCs for a UE activated from the eNB and its serving gNB, respectively, and we use round-robin (RR) algorithm to allocate the RBs in the PCCs [4]. A UE with a rate-hungry application may not meet its QoS (delay) requirement by only transmitting through the PCCs. Thus, SCC(s) from a gNB may need to be activated. This work assumes that the gNBs use the same spectrum for their CCs. Therefore, the UEs transmitting through their activated CCs from gNBs could interfere. Thus, the UL transmit power levels through their PCCs and SCCs should be adjusted carefully.

As mentioned before, in the networks supporting EN-DC technology, two different schemes, DPS and EPS, can be used to adjust the UL transmit power level. In DPS, the UL transmit power is dynamically calibrated. While in ESP, the UL transmit power level is equally divided between the eNB and the gNB. EPS may degrade the network performance because the time-varying characteristics for the channels and UE traffic in the 5G networks are not taken into consideration [3]. Thus, the network performance (network throughput) would degrade [2]. Based on a study performed in [2], the DPS technique offer 40% more of the average total throughput than ESP. Therefore, we focus on DPS and state the following multi-objective optimization problem for the CA and UL power control,

$$\begin{aligned}
&min. \quad \left[\sum_{k\in\mathcal{K}_b} \overline{D}^k, \sum_{k\in\mathcal{K}_b} P^k \right] \quad \forall b \in \mathcal{B} \\
&s.t. \quad \frac{\widehat{q}^k(t)}{R^k(t)} \leq \mathrm{D}_{\mathrm{QoS}}^k \quad \forall k \in \mathcal{K}, \\
&\qquad p^k(t) \leq p_{\max}^k \quad \forall k \in \mathcal{K}, \\
&\qquad \text{Number of CCs}, \\
&var. \quad \alpha_m^k \in \{0,1\} \quad \forall m \in \mathcal{M}, \forall k \in \mathcal{K} \\
&\qquad \widehat{p}^k, \widetilde{p}^k \quad \forall k \in \mathcal{K}
\end{aligned} \tag{8}$$

In (8), $\mathcal{B}$, $\mathcal{K}_b$, $\mathcal{M}$ are the set of gNBs, set of UEs in gNB b, and the set of CCs, respectively. Further, $\sum_{k\in\mathcal{K}_b} \overline{D}^k$ is the total delay for the UEs in $\mathcal{K}_b$, $\sum_{k\in\mathcal{K}_b} P^k$ is the UL total transmit power level for the UEs in $\mathcal{K}_b$. The above problem is a multi-objective optimization problem. Specifically, each gNB tries to minimize the total delay for the UEs in its coverage area and the total power consumed by the UEs in $\mathcal{K}_b$ to activate/deactivate the SCCs. This multi-objective optimization problem has two sets of integer (binary) and continuous variables. Therefore, it is a mixed-integer programming multi-objective optimization problem and is hard to solve. In this work, we derive an intelligent algorithm to address (8) sub-optimally. Specifically, we develop a multi-agent RL system with a compound-action space (which contains both continuous action and discrete action) and propose **Algorithm 1** to activate/deactivate SCCs for the UEs and adjust their UL transmit power levels to the BSs. At time slot t, given the activated CCs and the UL transmit power levels for the UEs, **Algorithm 4** in Appendix A is employed to allocate the RBs. The details of deriving the intelligent algorithm are given in the next section.

III. INTELLIGENT CA AND UL POWER-SHARING: A COMPOUND-ACTION ACTOR-CRITIC APPROACH

In this section, considering the main objective of the optimization problem (8) for a given gNB b, minimizing the total delay for the UEs in $\mathcal{K}_b$ and at the same time minimizing the power consumption for the UEs in $\mathcal{K}_b$ to activate the SCCs, we state the problem of UL power control and CC management as a multi-agent model-free RL system. The RL system is the one with compound-action space and thus we use the CA2C method to address the RL system, and drive a resource management scheme for joint CA and UL power

control in EN-DC wireless networks. The CA2C method has been proposed in [17] and has been recently used (e.g., by the authors in [25]) to solve a RL system with compound-action space.

A. THE PROPOSED MULTI-AGENT RL SYSTEM WITH COMPOUND-ACTION SPACE

As aforementioned, based on 3GPP, for UL transmission in the 5G networks with EN-DC technology, we consider the scenario wherein only one CC in low-band would be activated by the eNB. We assume that this CC is the always activated PCC in this work. Each UE has another always activated PCC at the NR side; additionally, the SCC would be activated by gNBs for their associated UEs (i.e., the ones in their coverage area). As depicted in Fig. 1, the RR algorithm is used to allocate the RBs on the always-activated PCCs. Furthermore, a multi-agent CA2C-based algorithm is employed to activate the SCCs and adjust the UL transmit power level for both gNBs and eNB. Like the one employed to allocate the RBs in PCCs, the RR algorithm is used by each gNB to allocate the RBs in the activated SCCs to the UEs. To jointly perform CC activation and UL power adjustment, for each gNB b, it is aimed to minimize the total delay for the UEs in $\mathcal{K}_b$ and at the same time minimize the power consumption for the UEs in $\mathcal{K}_b$ to activate the SCCs. This problem can be developed as a multi-agent RL system with a compound action space. The details are given below.

At a given time t, let us denote the immediate reward for a given gNB $b \in \mathcal{B}$ by $\eta_b(t)$. By using (3) and (5), we define $\eta_b(t)$ in terms of the total delay and the total UL power consumption for the UEs in $\mathcal{K}_b$:

$$\begin{aligned}\eta_b(t) &= -\Big(\sum_{k\in\mathcal{K}_b}\frac{\widehat{q}^k(t)}{R^k(t)} + \omega_b \sum_{k\in\mathcal{K}_b} P^k(t)\Big)\\ &= -\Big(\sum_{k\in\mathcal{K}_b}\frac{\widehat{q}^k(t)}{\sum_{n\in\mathcal{N}_1}\beta^k_{1,n}\widehat{R}^k_{1,n}(t) + \sum_{m\in\widetilde{\mathcal{M}}}\alpha^k_m\sum_{n\in\mathcal{N}_m}\beta^k_{m,n}\widetilde{R}^k_{m,n}(t)}\\ &\quad + \omega_b\sum_{k\in\mathcal{K}_b}P^k(t)\Big). \end{aligned} \tag{9}$$

In (9), ω_b is the unit price for the total amount of power consumed by the UEs assigned to the gNB b in order to activate the SCCs. Let Φ_b be the long-term reward for gNB b. Φ_b can be expressed as the weighted sum of the short-term reward $\eta_b(t)$ as follows,

$$\Phi_b = \sum_{t=0}^{T-1}\lambda^t\eta_b(t). \tag{10}$$

In (10), we have $0 \le \lambda \le 1$. Specifically, if $\lambda = 0$ then $\Phi_b = \eta_b(0)$; otherwise, by passing the time, the weighted immediate reward (i.e., $\lambda^t\eta_b(t)$) becomes smaller and has negligible impact on long-term reward Φ_b.

In this work, we are going to address the problem where each gNB b tries to maximize its long-term reward (minimizing its cost), which is given in (10). This problem can be restated as a stochastic game, and the Markov Decision Process (MDP) can be used to address the game, i.e., finding the Nash equilibrium at each state. However, complete knowledge about the environment would not be available to us in the 5G network (e.g., the transition probability between the states in the MDP). We formally express the MDP as a multi-agent model-free RL system to tackle this issue. The authors have used this method in [26] and [27] as well.[1] To address the RL system, we employ the CA2C algorithm proposed [17].

1) SET OF AGENTS AND REWARD FUNCTION

The set of gNBs, $\mathcal{B} = \{1, \ldots, B\}$, is the set of agents. The immediate and long-term reward functions for each agent b are given in (9) and (10), respectively.

2) STATE SPACE

In our proposed multi-agent RL system, all agents have the same state space which is given in terms of the delay requirement for the UEs. Let us denote the state space by $\mathcal{S}$ and define it as,

$$\mathcal{S} = \Big\{\mathbf{s}(t) \mid \mathbf{s}(t) = [s^k(t)]_{k\in\mathcal{K}} \text{ and } s^k(t) \in \{0, 1\}\Big\}. \tag{11}$$

Specifically, $s^k(t) = 1$ if the target delay requirement in (7) is satisfied for UE k. Otherwise, i.e., if (7) is not satisfied for UE k, we have $s^k(t) = 0$. The state space comprises a vector of zeros and ones, with each element representing whether the delay requirement for a user is met or not. Each NR BS can collect this information about its covered UEs and share it with the LTE BS, which can subsequently broadcast it among all NR BSs. The overhead of broadcasting the vector of zeros and ones among the NR BSs would be low. Based on (11), we have $|\ \mathcal{S}\ | = 2^K$ and therefore the dimension of state space, $|\ \mathcal{S}\ |$, grows exponentially when the number of the UEs, K, increases. In the following subsection, to handle the exponential growth of the state space due to the increasing number of UEs, we use a DRL-based algorithm to address the RL system.

3) ACTION SPACE

For a given gNB b (agent b) in the multi-agent RL system, the action corresponds to activating/deactivating the SCCs for the UEs covered by the gNB and adjusting its assigned UEs' UL transmit power level to gNB b and eNB. Let $\mathcal{A}_b$ be the action space for gNB b. So, we have

$$\begin{aligned}\mathcal{A}_b = \Big\{\mathbf{a}_b = \Big(\boldsymbol{\alpha}_b(t), \mathbf{r}_b(t)\Big) \mid \boldsymbol{\alpha}_b(t) = [\alpha^k_m(t)]_{k\in\mathcal{K}_b, m\in\widetilde{\mathcal{M}}} \text{ and}\\ \mathbf{r}_b(t) = \Big[\Big(\widehat{r}^k(t), \widetilde{r}^k(t)\Big)\Big]_{\forall k\in\mathcal{K}_b}\Big\}\end{aligned} \tag{12}$$

In (12), $\boldsymbol{\alpha}_b(t)$ is composed of zero and one. Specifically, we have $\alpha^k_m = 1$ if the SCC m is activated for UE $k \in \mathcal{K}_b$ and $\alpha^k_m = 0$, otherwise. Therefore, $\boldsymbol{\alpha}_b(t)$ is a discrete action.

[1] Due to the space limitation, we have dropped the details of formulating the stochastic game and MDP and refer the reader to the works above in [26] and [27].

Additionally, $\mathbf{r}_b(t)$ is the action used to adjust the UL transmit power level for the UEs in $\mathcal{K}_b$ to gNB b and eNB. Hence, $\mathbf{r}_b(t)$ is a continuous action. Specifically, given $\Big(\boldsymbol{\alpha}_b(t), \mathbf{r}_b(t)\Big)$, the UL transmit power levels for that UE to the eNB and to the gNB b for that UEs are given by,

$$\begin{cases} \widehat{p}^k(t) = \widehat{r}^k(t), \quad \widetilde{p}^k(t) = \widetilde{r}^k(t) \quad \text{if } \widehat{r}^k(t) + \widetilde{r}^k(t) \leq p^k_{\max}, \\ \widehat{p}^k(t) = \min(\widehat{r}^k(t), \dfrac{p^k_{\max}}{2}), \text{ and} \\ \widetilde{p}^k(t) = \min(\widetilde{r}^k(t), \dfrac{p^k_{\max}}{2}) \qquad \text{if } \widehat{r}^k(t) + \widetilde{r}^k(t) > p^k_{\max}. \end{cases} \tag{13}$$

Based on the above discussion, for the gNB b, the action space $\mathcal{A}_b$ consists of both discrete and continuous actions and thus is a compound-action space with a high dimension. Thus, the compound-action and state spaces will be huge for our proposed multi-agent RL system. To address it, finding the optimal policy corresponding to the best action $\mathbf{a}^*_b$, we employ the framework developed in [17].

B. THE PROPOSED COMPOUND-ACTION ACTOR-CRITIC ALGORITHM

Taking advantage of the deep deterministic policy gradient (DDPG) algorithm [28] and deep Q-network (DQN) algorithm [29], the authors in [17] have derived a CA2C method to learn the optimal policy. In this subsection, by employing the CA2C algorithm in [17], we propose a learning approach to derive the optimal policy to handle both continuous and discrete action, i.e., adjusting the UL transmit power levels of each UE to gNB and eNB, respectively, and activating/deactivating the SCCs for the UEs.

Let $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \ldots \boldsymbol{\pi}_B)$ be the policy profile for the gNBs. Specifically, for a given gNB b, $\boldsymbol{\pi}_b$ is a function from a given state $\mathbf{s}$ to action $\mathbf{a}_b$, i.e., $\boldsymbol{\pi}_b : \mathcal{S} \mapsto \mathcal{A}_b$. So, the policy profile $\boldsymbol{\pi}$ corresponds to patterns of behaviour for the gNBs at the different states of the environment and thus the policy for them should be optimized to maximize the long-term reward for the gNBs (given in (10)). Let $\boldsymbol{\pi}^* = (\boldsymbol{\pi}^*_1, \ldots \boldsymbol{\pi}^*_B)$ be the optimal policy profile for the gNBs. To find the optimal policy for a given gNB b, both current and future reward for the agent should be taken into consideration. For a given state $\mathbf{s}$ and action $\mathbf{a}_b$, let $Q_b(\boldsymbol{\pi}_b, \boldsymbol{\pi}_{-b}, \mathbf{s}, \mathbf{a}_b)$ be the Q-function for gNB b where $\boldsymbol{\pi}_b$ is the policy for agent (gNB) b and $\boldsymbol{\pi}_{-b} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{b-1}, \boldsymbol{\pi}_{b+1} \ldots \boldsymbol{\pi}_B)$ is the policy for the others. For ease of reference, from here on, we use the notation $Q_b(\mathbf{s}, \mathbf{a}_b)$ for gNB b's Q-function. Specifically, $Q_b(\mathbf{s}, \mathbf{a}_b)$ is defined in terms of the expectation of the weighted sum of the short-term reward for the agent [30]. Based on what is discussed in [17] and [26], the optimal policy for an agent gNB b is the policy under which the Q-function for that agent is maximized, i.e.,

$$\mathbf{a}^*_b = \pi^*_b(\mathbf{s}) = \arg \max_{\mathbf{a}_b \in \mathcal{A}_b} Q^*_b(\mathbf{s}, \mathbf{a}_b), \quad \forall \mathbf{s} \in \mathcal{S}. \tag{14}$$

The action space is composed of continuous and discrete actions in the above problem.[2] To find the optimal policy, we use the CA2C algorithm in [17]. The details are given in what follows.

For a given gNB (agent) b, let us decompose the optimal policy $\boldsymbol{\pi}^*_b$ into two optimal policies to adjust the UL transmit power levels to eNB and gNB b, and to activate/deactivate SCCs, respectively. For a gNB b, given state $\mathbf{s}$ and discrete action $\boldsymbol{\alpha}_b$ (which corresponds to activating the SCCs for the UEs in $\mathcal{K}_b$), let $\boldsymbol{\nu}^*_b(\mathbf{s}, \boldsymbol{\alpha}_b)$ denote the optimal policy to adjust the UL transmit power levels to the LTE and NR BSs. The best action for activating the SCCs for the UEs in $\mathcal{K}_b$ is obtained as follows [17],

$$\alpha^*_b = \arg \max_{\boldsymbol{\alpha}_b \in \mathcal{C}_b} Q^*_b\Big(\mathbf{s}, \big(\boldsymbol{\alpha}_b, \boldsymbol{\nu}^*_b(\mathbf{s}, \boldsymbol{\alpha}_b)\big)\Big), \tag{15}$$

where $\mathcal{C}_b = \{\boldsymbol{\alpha}_b = [\alpha^k_m(t)]_{k \in \mathcal{K}_b, m \in \widetilde{\mathcal{M}}} \mid \alpha^k_m \in \{0, 1\}\}$. In (15), finding the exact value for Q^*_b and $\boldsymbol{\nu}^*_b$ is challenging because the action and state space are high dimensional. To address this issue, we use the deep neural network (DNN) to approximate them. Specifically, we employ the CA2C algorithm in [17] wherein the DQN and DDPG algorithms are employed to train the corresponding DNNs separately. The details are given below.

As illustrated in Fig. 2, for a given agent b, two separated DNNs- actor DNN with parameter $\boldsymbol{\theta}_b$ and critic DNN with parameter $\mathbf{w}_b$- are used to approximate the Q^*_b and $\boldsymbol{\nu}^*_b$, respectively. Let us denote the parametrized Q-function and the parametrized policy for adjusting the UL transmit power level to the eNB and gNB b by $Q_b(\mathbf{s}, \mathbf{a}_b, \mathbf{w}_b)$ and $\boldsymbol{\nu}_b(\mathbf{s}, \boldsymbol{\alpha}_b, \boldsymbol{\theta}_b)$, respectively. For gNB b (agent b), the state $\mathbf{s}$ and $\boldsymbol{\alpha}_b$ (which are pertinent to the activated SCCs for the UEs in $\mathcal{K}_b$) are first given as the input to the actor DNN. The actor DNN approximates $\boldsymbol{\nu}_b(\mathbf{s}, \boldsymbol{\alpha}_b, \boldsymbol{\theta}_b)$ to obtain the continuous action $\mathbf{r}_b(t)$. Then, given state $\mathbf{s}$ and $\mathbf{r}_b(t)$ to the critic network, the Q-function for agent b, $Q_b(\mathbf{s}, \mathbf{a}_b, \mathbf{w}_b)$, is approximated. Based on (15), $Q_b(\mathbf{s}, \mathbf{a}_b, \mathbf{w}_b)$, is used to activate the SCCs for the UEs in $\mathcal{K}_b$.

Training procedure: To train the actor and critic DNNs, the DDPG and DQN algorithms in [28] and [29] are used, respectively [17]. In both aforementioned algorithms, to prevent overoptimism and instability, the concept of target network and online network are employed. Specifically, the actor and critic target networks are with parameters $\boldsymbol{\theta}^-_b$ and $\mathbf{w}^-_b$, respectively. Additionally, the parameters for the actor and critic online networks are $\boldsymbol{\theta}_b$ and $\mathbf{w}_b$, respectively. At each step, we use the soft update method for updating the parameters for the target networks, i.e., $\boldsymbol{\theta}^-_b = \tau\boldsymbol{\theta}^-_b + (1-\tau)\boldsymbol{\theta}_b$ and $\mathbf{w}^-_b = \tau\mathbf{w}^-_b + (1-\tau)\mathbf{w}_b$ where τ is the fixed parameter. The replay buffer is used to train the actor and critic online DNNs (i.e., setting the parameters $\boldsymbol{\theta}_b$ and $\mathbf{w}_b$, respectively).

[2] While discretizing the action space can simplify the problem, it can also pose significant challenges in learning the policy for activating/deactivating a CC and adjusting the UL transmit power level of the UE to the BSs. This is an important consideration because a less accurate policy can lead to degradation in system performance.

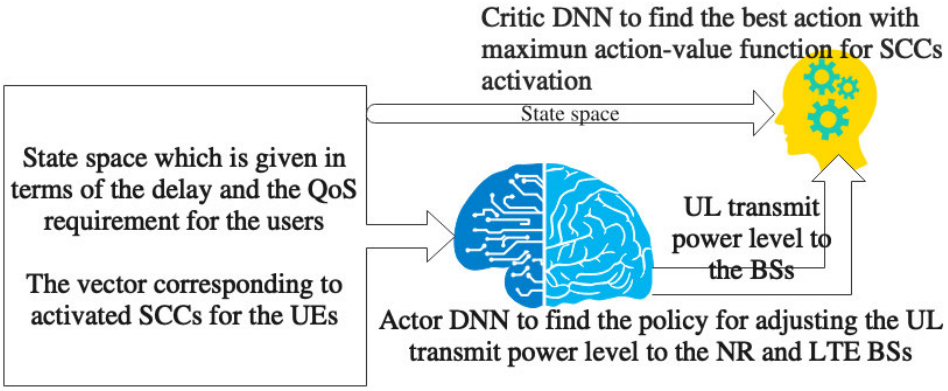

FIGURE 2. Actor-critic-based method of training the DNNs used to approximate Q_b^* and v_b^* for an agent b.

Specifically, for a given agent b, at a given time t, let us denote the tuple of current space, current action, next state and reward as the experience $\mathbf{e}_b^t = [\mathbf{s}(t), \mathbf{a}_b(t), \mathbf{s}', \eta_b(t)]$. The replay buffer for the agent b is denoted by $\mathcal{D}_b = \{\mathbf{e}_b^t\}$. Additionally, we denote the functions evaluated by the actor and critic networks by $J_b(\boldsymbol{\theta}_b)$ and $L_b(\mathbf{w}_b)$, respectively. The function $J_b(\boldsymbol{\theta}_b)$ is given in terms of the average Q-function for agent b and $L_b(\mathbf{w}_b)$ is given in terms of the average difference between the Q-function and target value for agent b. The details for functions $J_b(\boldsymbol{\theta}_b)$ and $L_b(\mathbf{w}_b)$ and the corresponding target value are given in [17]. For more details, we refer the readers to this reference. We mention the updating function for $\boldsymbol{\theta}_b$ and $\mathbf{w}_b$ as follows [17],

$$\boldsymbol{\theta}_b = \boldsymbol{\theta}_b + \zeta \nabla_{\boldsymbol{\theta}_b} J_b(\boldsymbol{\theta}_b), \tag{16a}$$

$$\mathbf{w}_b = \mathbf{w}_b - \zeta \nabla_{\mathbf{w}} L_b(\mathbf{w}_b), \tag{16b}$$

where ζ is the learning rate. By following the procedure illustrated in Fig. 2 and using (16a) and (16b), we propose the algorithm to train the actor and critic networks in **Algorithm 5** and **Algorithm 6**, respectively [17]. See Appendix B.

The proposed CA2C-based algorithm: Now, by using the training algorithm for the actor and critic algorithms, **Algorithm 5** and **Algorithm 6**, we derive a CA2C-based algorithm to activate the SCCs and adjusting the UL transmit power level for each UEs to its serving gNB and eNB. We call it the Intelligent *Joint power-sharing and CA (JPSCA)* algorithm. It is noteworthy that, to make a trade-off between exploration and exploitation, given the continuous action $\mathbf{r}_b$, we use the $\epsilon-$greedy algorithm to adopt the discrete action in the critic side which is given by,

$$\boldsymbol{\alpha}_b = \begin{cases} \arg\max\limits_{\boldsymbol{\alpha}_b \in \mathcal{C}_b} Q_b\big(\mathbf{s}, (\boldsymbol{\alpha}_b, \mathbf{r}_b), \mathbf{w}_b\big) & \text{with probability of } 1-\epsilon, \\ \text{randomly adopted from } \mathcal{C}_b & \text{with probability of } \epsilon. \end{cases} \tag{17}$$

Accordingly, the details of the Intelligent JPSCA algorithm are given in **Algorithm 1**.

In the **JPSCA** Algorithm, for a gNB b, at each state $\mathbf{s}$, given the activated SCC for the UEs in $\mathcal{K}_b$, the policy for UL transmit power level for the UEs to the eNB and that to gNB b are updated by using the actor networks. By using the output of the actor online network, i.e., $\mathbf{r}_b$, the UL transmit power levels to the BSs are obtained through the power control scheme derived in (13). As mentioned before, the RR algorithm assigns the RBs to the UEs sharing a CC. Specifically, we modify the RR algorithm and allocate the number of RBs to the UEs concerning their QoS requirements. Then, the output of the actor DNN and the state are given as input to the critic DNNs to activate/deactivate the CCs for the UEs in $\mathcal{K}_b$. At each step, the transition of the current state, current action, reward and the next state is captured in the experience memory for the agent (gNB) b. Specifically, at each step, a mini-batch of the transmissions is uniformly chosen and used to train the actor and critic DNNs. **Algorithm 5** and **Algorithm 6** are used to train the DNNs of the critic and the DNNs of the actor, respectively.

IV. THE DDQN-BASED UL POWER-SHARING AND CA

In this section, we develop the two algorithms called **ACPS** (All-activated CCs and Power Sharing) and **DJPCA** (Discrete Joint Power sharing and CA), respectively. The details are given in what follows.

1) ACPS ALGORITHM

To derive this algorithm, we separate the CA scheme from the UL power-sharing scheme. For CA, we consider the extreme scenario where all CCs are activated for the UEs. For the UL power-sharing, we quantize the continuous UL transmit power levels for the UE to the eNB and the gNBs, i.e., approximate them by ones whose amplitudes are restricted to a prescribed set of values. So, given the activated CCs for the UEs, we develop the multi-agent RL system for UL power-sharing. In the RL system, the set of agents are the

Algorithm 1 The Proposed **JPSCA** Algorithm

input : The action set for each gNB b, i.e., $\mathcal{A}_b,\ \forall b \in \mathcal{B}$, target delay requirement for the UEs in $\mathcal{K}_b$

output: Optimal policy to activate/deactivate the SCCs for the UEs in $\mathcal{K}_b$, and adjust the UE UL transmit power levels to the serving gNB b and eNB

Initialization: Experience memory $\mathcal{D}_b, b \in \mathcal{B}$, actor online network parameters $\boldsymbol{\theta}_b, b \in \mathcal{B}$, actor target network parameters $\boldsymbol{\theta}_b^- = \boldsymbol{\theta}_b, b \in \mathcal{B}$, critic online network parameters $\mathbf{w}_b, b \in \mathcal{B}$, critic target network parameters $\mathbf{w}_b^- = \mathbf{w}_b, b \in \mathcal{B}$

for *episode* $= 1 : \mathrm{T_{cca}} : \mathrm{T_{sim}}$ **do :**

  Initialize the network state $\mathbf{s}$;

  for *step* $= 1 : T$ **do :**

    Given state $\mathbf{s}$ and the continuous action $\mathbf{r}_b$, each gNB $b \in \mathcal{B}$ uses its critic online network and approximates action-value $Q_b(\mathbf{s}, (\boldsymbol{\alpha}_b, \mathbf{r}_b), \boldsymbol{\theta}_b), \forall \boldsymbol{\alpha}_b \in \mathcal{C}_b$ and adopts $\boldsymbol{\alpha}_b$ by using (17);

    Given state $\mathbf{s}$ and the discrete action $\boldsymbol{\alpha}_b$ (corresponds to activating the SCCs for the UEs in $\mathcal{K}_b$), $\mathbf{r}_b$ is obtained, which is used to adjust the UL transmit power levels for the UEs in $\mathcal{K}_b$ by using (13);

    Use **Algorithm 4** to allocate the RBs to the UEs sharing the SCCs and PCCs and adjust the UL transmit power levels through them;

    The next state $\mathbf{s}'$ is obtained via the message passing and $\eta_b,\ \forall b \in \mathcal{B}$ is obtained by each gNB b. Set $\mathbf{s} \leftarrow \mathbf{s}'$;

    Store experience $\mathbf{e}_b^t = [\mathbf{s}, \mathbf{a}_b, \eta_b, \mathbf{s}']$ in memory $\mathcal{D}_b$ for $b \in \mathcal{B}$;

    Each gNB $b \in \mathcal{B}$ samples random mini-batch of transitions $[\mathbf{s}, \mathbf{a}_b, \eta_b, \mathbf{s}']$ from $\mathcal{D}_b$;

    The parameters of the DNNs for each gNB b are updated by using **Algorithm 5** and **Algorithm 6**;

  end

end

Algorithm 2 The **ACPS** Algorithm

input : The action set for each gNB b, i.e., $\mathcal{A}_b^{\mathrm{DDQN}},\ \forall b \in \mathcal{B}$, target delay requirement for the UEs in $\mathcal{K}_b$, and the set of activated CCs for the UEs in $\mathcal{K}_b$

output: Optimal policy to adjust UE UL transmit power levels to the serving gNB b and eNB

Initialization: Experience memory $\mathcal{D}_b^{\mathrm{DDQN}}, b \in \mathcal{B}$, online network parameters $\mathbf{O}_b, b \in \mathcal{B}$, target network parameters $\mathbf{O}_b^- = \mathbf{O}_b, b \in \mathcal{B}$, $\mathcal{M}^k = \mathcal{M}\ \forall k \in \mathcal{K}$;

for *episode* $= 1 : T$ **do :**

  Initialize the network state $\mathbf{s}$;

  for *step* $= 1 : T$ **do :**

    Given state s, each gNB $b \in \mathcal{B}$ uses its online network to approximates its action-value $Q^k(\mathbf{s}, \mathbf{a}_b, \mathbf{O}_b), \forall \mathbf{a}_b \in \mathcal{A}_b^{\mathrm{DDQN}}$;

    Given state $\mathbf{s}$, each gNB $b \in \mathcal{B}$ employ the ϵ-greedy policy in (17) to adopt $\widetilde{q}_q^k$ for all $k \in \mathcal{K}_b$ (with $Q_b^{\mathrm{DDQN}}(\mathbf{s}, \mathbf{a}_b, \mathbf{O}_b)$ and $\mathcal{A}_b^{\mathrm{DDQN}}$);

    Use **Algorithm 4** to allocate the RBs to the UEs sharing the CCs and adjust the UL transmit power levels through them;

    The message passing is employed to get the next state $\mathbf{s}'$ and $\mathbf{s} \leftarrow \mathbf{s}'$;

    The transition $[\mathbf{s}, \mathbf{a}_b, \eta_b, \mathbf{s}']$ in memory $\mathcal{D}_b^{\mathrm{DDQN}}$;

    Each gNB $b \in \mathcal{B}$ samples random mini-batch of transitions from $\mathcal{D}_b^{\mathrm{DDQN}}$ and update the parameters for online network, $\mathbf{O}_b$, by using equation (29) in [26];

    $C_U \leftarrow C_U + 1$;

    if $C_U == N$ **then** $\mathbf{O}_b^- = \mathbf{O}_b, b \in \mathcal{B}$; $C_U \leftarrow 0$;

  end

end

set of gNBs, i.e., $\mathcal{B} = \{1, \ldots, B\}$, the reward function for each agent and the state space are the ones in (9) and (11), respectively. For the action space, let us denote the quantized UL transmit power levels for a given UE k to its serving gNB b by $\widetilde{q}_q^k$ where $q \in \mathcal{Q} = \{1, \ldots Q\}$. So, the action space for agent b is denoted by $\mathcal{A}_b^{\mathrm{DDQN}}$ and given by $\mathcal{A}_b^{\mathrm{DDQN}} = \{\mathbf{a}_b^{\mathrm{DDQN}} \mid \mathbf{a}_b^{\mathrm{DDQN}} = [\widetilde{q}_q^k]_{k \in \mathcal{K}_b, q \in \mathcal{Q}}\}$. Therefore, we have $\mid \mathcal{A}_b^{\mathrm{DDQN}} \mid = \mid \mathcal{K}_b \mid^Q$, and thus, an increase in either the number of UEs in $\mathcal{K}_b$ or the number of quantization levels results in an explosion in the dimension of the action space for agent b. Accordingly, the Q-learning algorithm [12] used by the authors in [3] does not provide sufficient performance to address the computational needs of the RL system. To tackle this issue, we employ the DDQN algorithm [18] in DRL. The DDQN algorithm uses the target network, online network, and experience memory to prevent overoptimism. The details are given in [18], and we refer the readers to this reference. The details for the **ACPS** algorithm, which is based on the DDQN algorithm, are given in **Algorithm 2**.

2) **DJPCA** ALGORITHM

In this algorithm, we quantize the continuous UL transmit power levels for the UE to the eNB and the gNBs and develop the multi-agent RL system for joint UL power-sharing and CA. In the RL system, the set of gNBs, i.e., $\mathcal{B} = \{1, \ldots, B\}$ is the set of agents, the reward function for each agent is given by (9) and (11) is used for state space. For a given agent $b \in \mathcal{B}$, the action space corresponds to activating/deactivating

Algorithm 3 The **DJPCA** Algorithm

input : The action set for each gNB b, i.e., $\mathcal{A}'_b$, $\forall b \in \mathcal{B}$, target delay requirement for the UEs in $\mathcal{K}_b$, and the set of activated CCs for the UEs in $\mathcal{K}_b$

output: Optimal policy to adjust UE UL transmit power levels to the serving gNB b and eNB and activate/deactivate a SCC

Initialization: Experience memory $\mathcal{D}'_b$, $b \in \mathcal{B}$, online network parameters $\mathbf{O}'_b$, $b \in \mathcal{B}$, target network parameters $\mathbf{O}'^{-}_b = \mathbf{O}'_b$, $b \in \mathcal{B}$, $\mathcal{M}^k = \mathcal{M}\ \forall k \in \mathcal{K}$;
for *episode* $= 1 : T$ **do :**
  Initialize the network state $\mathbf{s}$;
  for *step* $= 1 : T$ **do :**
    Given state s, each gNB $b \in \mathcal{B}$ uses its online network to approximates its action-value $Q'_b(\mathbf{s}, \mathbf{a}'_b, \mathbf{O}'_b), \forall \mathbf{a}'_b \in \mathcal{A}'_b$;
    Given state $\mathbf{s}$, each gNB $b \in \mathcal{B}$ employ the ϵ-greedy policy in (17) to adopt $\widetilde{q}^k_q$ and activate a SCC, i.e., α^k_m, for all $k \in \mathcal{K}_b$;
    Use **Algorithm 4** to allocate the RBs to the UEs sharing the CCs and adjust the UL transmit power levels through them;
    The message passing is employed to get the next state $\mathbf{s}'$ and $\mathbf{s} \leftarrow \mathbf{s}'$;
    The transition $[\mathbf{s}, \mathbf{a}'_b, \eta_b, \mathbf{s}']$ in memory $\mathcal{D}'_b$;
    Each gNB $b \in \mathcal{B}$ samples random mini-batch of transitions from $\mathcal{D}'_b$ and update the parameters for online network, $\mathbf{O}'_b$, by using equation (29) in [26];
    $C_U \leftarrow C_U + 1$;
    if $C_U == N$ **then** $\mathbf{O}'^{-}_b = \mathbf{O}'_b$, $b \in \mathcal{B}$;
    $C_U \leftarrow 0$;
  end
end

a CC for a UE $k \in \mathcal{K}_b$ and adjusting its UL transmit power to the gNB b. We denote the action space for agent b by $\mathcal{A}'_b$ and define it as $\mathcal{A}'_b = \left\{\mathbf{a}'_b = \left(\boldsymbol{\alpha}_b, \mathbf{a}^{\text{DDQN}}_b\right) \mid \boldsymbol{\alpha}_b = [\alpha^k_m]_{k\in\mathcal{K}_b, m\in\widetilde{\mathcal{M}}}$ and $\mathbf{a}^{\text{DDQN}}_b = [\widetilde{q}^k_q]_{k\in\mathcal{K}_b, q\in\mathcal{Q}}\right\}$. As aforementioned, for a given gNB (agent) b and its assigned UE $k \in \mathcal{K}_b$, α^k_m corresponds to activating/deactivating the CC m for the UE and $\widetilde{q}^k_q$ is the quantized UL transmit power level for the UE k to its serving gNB b. To solve the RL system, we employ the DDQN algorithm and derive the **DPCA** in **Algorithm 3**.

V. SIMULATION RESULTS

As depicted in Fig. 1, we consider an EN-DC network consisting of one eNB and K number of UEs. Additionally, $B = 5$ gNBs are randomly inserted in the coverage area of the eNB. The eNB has a circular coverage area with a radius of 500 meters, and the circular coverage area for each gNB has a radius of 50 meters. A given gNB b can serve a number of K_b UEs uniformly distributed over its coverage area where $2 \leq K_b \leq 5$. For our setup, based on [19], only one CC in low-band, 800 MHz, can be activated by the eNB for the UEs. We consider this CC the always activated PCC for the UEs at the LTE side. Also, for each gNB, three CCs are operating in mid-band frequency (3.5 GHz). Two of these CCs are considered SCCs that can be activated/deactivated at any time, and one of them is the always activated PCC for the UEs at the NR side. For the DNNs, an input layer, three hidden layers and an output layer are considered. The number of neurons in the hidden layers is 128, 512 and 1024, respectively. The optimal gradient descent algorithm, ReLU activation function and Adam optimization are used to update the weights for the DNNs. Through the simulations, we set the hyper-parameters, e.g., ϵ and learning rate ζ. The simulation parameters and the structure of the DNN are given in Table 2.

In this work, FTP traffic, i.e., FTP model 3 is employed [31] to model the bursty traffic, and we assume the file arrives at the scheduler in one burst. In FTP model 3, the number of UEs is fixed, and each UE's files arrive with the Poisson distribution. Thus, the inter-arrivals have exponential distributions, and therefore, the file arrivals have Poisson distributions with λ_{T} inter-arrival rate. Additionally, for the file size, a log-normal distribution [32] with parameters μ as mean and σ as standard deviation is selected. That is, given $\overline{q}^k$ as the average file size for a UE k, $\ln(\overline{q}^k)$ has a normal distribution $N(\mu, \sigma)$ and we have $\overline{q}^k = \exp(\mu + 0.5 \times \sigma^2)$.

In this section, we evaluate the performance of the **JPSCA** algorithm in comparison with the following baseline algorithms: **ACPS** algorithm, **DJPCA** algorithm and **AEPS** (All-activated CCs with Equal power-sharing) algorithm. All CCs are activated for the UEs in both baseline **ACPS** and **AEPS** algorithms. In **ACPS**, the DDQN algorithm [18] is used to adjust the UL transmit power for each UE to the BSs. While in **DJPCA** algorithm the DDQN algorithm is used for joint CA and UL power sharing. Note that for the **AEPS** algorithm, the UL transmit power level for the UE $k \in \mathcal{K}_b$ to its serving gNB b and that to the eNB are the same. The details for the **ACPS** algorithm are given in **Appendix IV**.

A. CONVERGENCE BEHAVIOUR OF JPSCA ALGORITHM

To study the convergence behaviour for the reward function of a given gNB by using the proposed **JPSCA** algorithm, we consider a scenario with a eNB and two gNBs. There are six UEs in the network, uniformly distributed in the coverage area. **JPSCA** algorithm is a CA2C-based algorithm. The convergence of the CA2C algorithm has been proven in [17]. Hence, using the **JPSCA** algorithm, the average weighted reward function for each gNB converges to a stationary point. This is shown in **Fig. 3**, as well. It is noteworthy that the DNNs for different agents (gNBs) do not necessarily have the same weights. Accordingly, the stationary point for the average weighted reward function for each gNB would not be the same because each gNB uses its DNNs to adjust the UL

TABLE 2. Simulation parameter and hyper-parameters for the learning algorithms.

Parameter	Value
Max. total UL transmit power level for a UE	26 dBm
Average target delays for the UEs	150 msec
T_{cca}	200 msec
Number of CCs in $\widehat{\mathcal{M}}$ and $\widetilde{\mathcal{M}}$, (i.e., $\widehat{M}$ and $\widetilde{M}$), respectively	1, 3
Carrier Setting	CC1 is the CC in low-band frequency (800 MHz) with 5 MHz bandwidth CC2 and CC3 are the CCs in mid-band frequency (3.5 GHz) with 10 MHz bandwidth
PHY numerology	15 KHz sub-carrier spacing, PRB size of 12 sub-carrier (180 kHz), 1ms slot length
Traffic Model and parameters	3GPP FTP model 3, $[\mu, \sigma] = [12.5, 0.35]$, $\lambda_T \in \{10, 15, 20\}$
Number of episodes and steps	100, 100
Discount factor(λ), ϵ, learning rate (ζ), τ	0.9, 0.1, 0.01, 0.01
Replay memory size for both CA2C and DDQN algorithm	500
Optimizer and activation function	Adam and ReLU
Device parameters- $\Theta_L, \Lambda_L, \Theta_H, \Lambda_H, \Gamma_D$ for mid-band frequency dbm	14.25 mW/dbm, 2.1625 W, 117.5 mW/dbm, 1.22 W, 16 dbm
Device parameters- $\Theta_L, \Lambda_L, \Theta_H, \Lambda_H, \Gamma_D$ for low-band frequency dbm	7.5 mW/dbm, 1.325 W, 10^{-6} mW/dbm, 2.3 W, 16 dbm

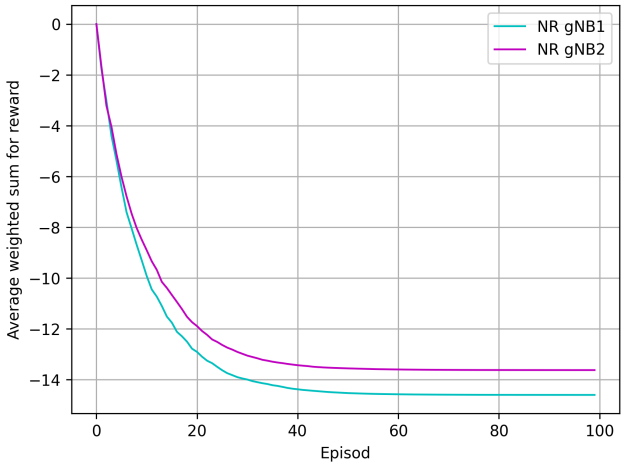

FIGURE 3. Convergence behaviour for CA2C-based (JPSCA) algorithm.

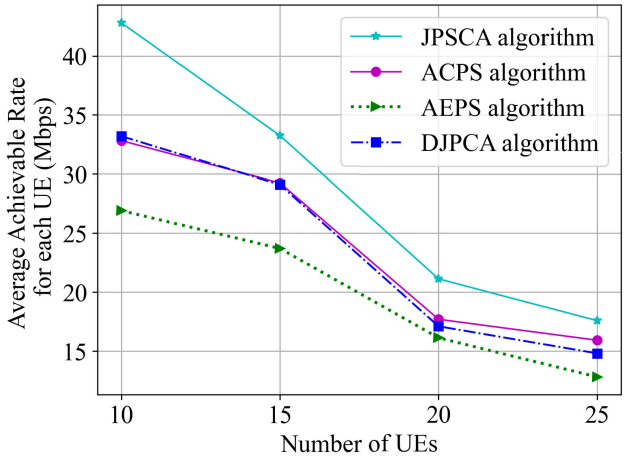

FIGURE 4. Average rate per UE for CA2C-based (JPSCA) algorithm vs. DDQN-based discrete joint power sharing and CA (DJPCA) algorithm.

transmit power level, and activate/deactivate the SCCs for the UEs in its coverage area.

B. PERFORMANCE EVALUATION

To evaluate the performance of **JPSCA** algorithm, we compare it with the performance of **DJPCA**, **ACPS** and **AEPS** algorithms in terms of the average rate per UE, average delay per UE, average number of activated CCs per UE, average UL consumed power per UE, and average UL transmit power levels for each UE to its assigned gNB and eNB which are illustrated in **Figs. 4, 5, 6, 7, 8a, 8b**, respectively. In the simulations, we have considered $B = 5$, and different number of UEs, i.e., $K \in \{10, 15, 20, 25\}$.

Based on **Fig. 4**, by employing **DJPCA**, **ACPS**, **AEPS** and **JPSCA** algorithms, the average achievable rate per UE decreases as the number of UEs increases. The more the number of UEs, the less the number of allocated RBs to each UE. The achievable rate per UE by employing **JPSCA** is higher than those obtained by using **DJPCA**, **AEPS** and

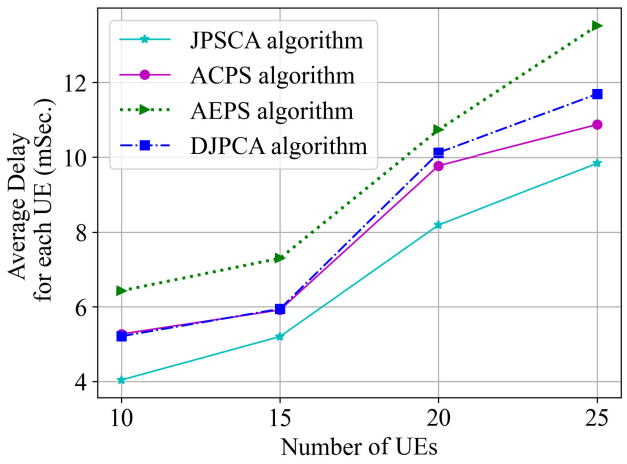

FIGURE 5. Average delay per UE for CA2C-based (JPSCA) algorithm vs. DDQN-based discrete joint power sharing and CA (DJPCA) algorithm.

ACPS algorithms. In **DJPCA** and **JPSCA**, the joint control of CA and UL power is performed. However, **DJPCA** employs quantization for the UL transmit power level of UEs.

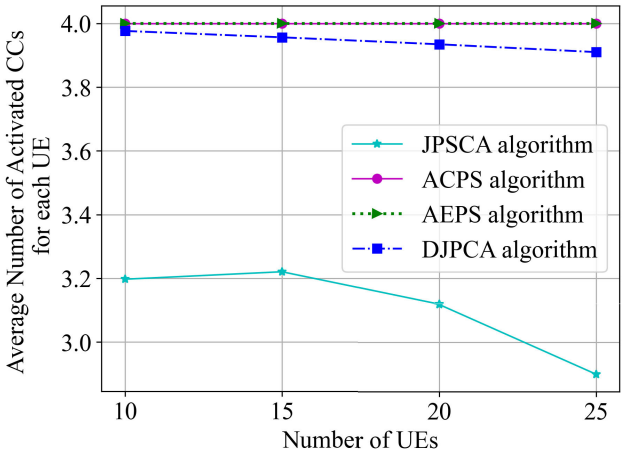

FIGURE 6. Average number of activated CCs per UE for CA2C-based (JPSCA) algorithm vs. DDQN-based discrete joint power sharing and CA (DJPCA) algorithm.

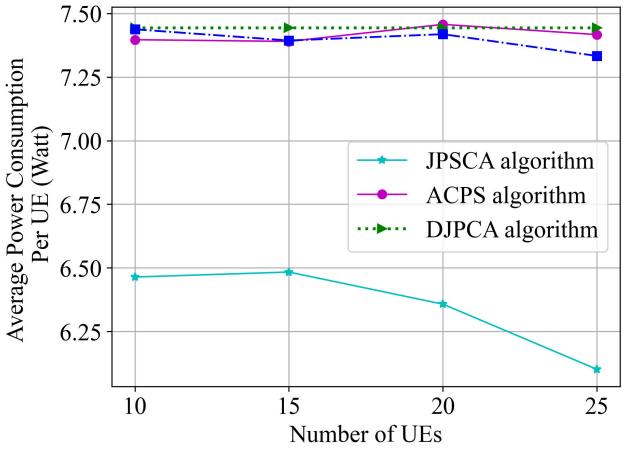

FIGURE 7. Average UL power consumption per UE for CA2C-based (JPSCA) algorithm vs. DDQN-based discrete joint power sharing and CA (DJPCA) algorithm.

This quantization imposes limitations on the granularity of power control, potentially exacerbating interference issues. As a consequence, the achievable data rate per UE obtained from **DJPCA** is lower compared to **JPSCA**, where the UL transfer power level remains continuous and unquantized. The continuous power control in **JPSCA** would enable finer adjustments, thereby improving the achievable rate per UE and reducing interference levels. In the baseline AEPS and ACPS, all CCs are activated for each UE. On one hand, in **JPSCA** algorithm, CC management and UL power control are performed jointly. Additionally, by activating all CCs for all UEs in **AEPS** and **ACPS** algorithms, the interference through the RBs in CCs increases which could result in more degrading the rate for the UEs. It is noteworthy that since the UL power control for each UE in **ACPS** algorithm is intelligently performed by using the DDQN algorithm, the performance of **ACPS** algorithm in terms of achievable rate for each UE is higher than that for **AEPS** algorithm where the UL transmit power levels for each UE to its serving gNB and the eNB are the same. Similarly, we can evaluate the performance of **DJPCA**, **AEPS**, **ACPS** and **JPSCA** algorithms in terms of the delay for each UE. Specifically, based on **Fig. 5**, as the number of UEs grows, which declines the rate for each UE, the delay per UE increases. However, the performance of the **JPSCA** algorithm in terms of the delay for each UE is better than **DJPCA**, **AEPS** and **ACPS** algorithms. For instance, for up to almost 25 UEs, the delay obtained by each of UEs using the **JPSCA** is around the target delay for the URLLC applications in 5G. This can be because of intelligently joint power and CC management in **JPSCA** algorithm by using the CA2C algorithm. In **DJPCA**, as mentioned earlier, the transmit power level is quantized, which leads to increased interference. Consequently, this interference causes a decrease in data rate and an increase in delay for each UE. Meanwhile, in the **ACPS** algorithm, where the UL transmit power level per UE is intelligently adjusted, the delay per UE is lower than that for **AEPS** algorithm.

For the small cell NRs with limited resources, using the **JPSCA** algorithm for CA and UL power control results in better performance in terms of the average number of activated CCs, UL power consumption for each UE. According to **Figs. 6 and 7**, the average number of activated SCCs per UE and the average UL power consumption are lower than those obtained by employing **DJPCA**, **ACPS** and **AEPS** algorithms. This is because of intelligent joint CC management and UL power-sharing in **JPSCA** algorithm. Compared to **JPSCA**, **DJPCA** activates a greater number of component CCs. In **DJPCA**, the quantization of the uplink (UL) transmit power level can have an impact on the selection and activation of CCs, resulting in suboptimal utilization of these CCs. Also, since the UL power-sharing in **ACPS** is performed by using the DDQN algorithm, it has better performance in comparison with **AEPS** algorithm.

As observed in **Fig. 8**, by using **JPSCA** algorithm, the UL transmit power level per UE to its serving gNB is higher than those of the **DJPCA**, **ACPS** and **AEPS** algorithms. Based on (3), the UL power consumption for a UE is a linear function over both UL transmit power level and the operating frequency. By using the **JPSCA** algorithm, as illustrated in Fig. 6, the number of activated SCCs per UE is reduced. This prevents the dramatic increase in the total UL power consumption for each UE. However, as the number of activated SCCs decreases, to intelligently make a trade-off between minimizing the delay and minimizing the UL power consumption, the UL transmit power level per UE to its serving gNB increases (which results in a more achievable rate for each UE).

Illustrated by **Fig. 8b**, as the UL transmit power levels for the UEs to gNBs increase, the UL transmit power level for the UEs to the eNB would decrease (based on constraint (2)). This results in a lower total UL power consumption for each UE. For the number of 25 UEs, there is an increase in the UL transmit power level to the eNB per UE. On the one

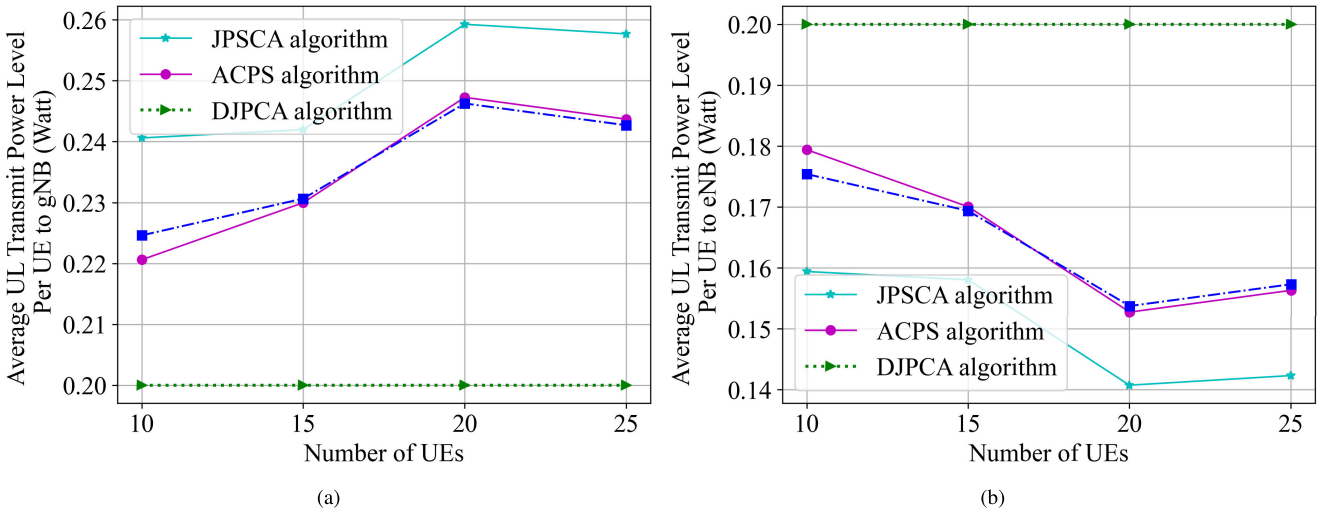

FIGURE 8. Average UL transmit power level per UE to gNB (a) , and average UL transmit power level per UE to eNB (b) for CA2C-based (JPSCA) algorithm vs. DDQN-based discrete joint power sharing and CA (DJPCA) algorithm.

hand, to serve 25 UEs, compared to 20 UEs, using the RR algorithm results in fewer RBs allocated to each UE in the PCC. On the other hand, setting the delay degradation rate for 25 UEs to be the same as that for 20 UEs, it is necessary to prioritize minimizing the total UEs' delay concerning the UE power consumption. Thus, a trade-off between the UL power consumption and the delay for the UEs using the **JPSCA** algorithm occurs when the average number of activated SCCs at the NR side and the allocated RB at the LTE side decreases. This is obtained by increasing the UL transmit power level per UE to the eNB.

VI. CONCLUSION

In this paper, to the best of our knowledge, for the first time, we proposed an intelligent joint dynamic power-sharing and CA scheme in EN-DC networks and showed it outperforms all-CC activated based algorithms and the algorithm employing equal power-sharing. Specifically, the problem of minimizing the delay and UE power consumption with both continuous and discrete variables for UL power adjustment and activating/deactivating the SCCs has been formally expressed. To address this problem, a compound action actor-critic algorithm (CA2C-based algorithm) has been proposed to jointly adjust the UL transmit power level for the UEs to the BSs and activate the SCCs. As shown in the simulation results, such an intelligent, joint resource management scheme performs better in power consumption and the number of activated CCs. Additionally, the proposed CA2C-based algorithm has a better achievable rate and delay performance than all CCs activated algorithms where the UL power control and CA are performed separately. Additionally, the CA2C-based algorithm has better performance that the DDQN-based joint power sharing and CA where the transmit power level for the UEs are quantized. In dynamic network environments where channel conditions change rapidly, quantization may not allow users to adapt their transmit power levels quickly enough which may result in limited options for power allocation across CCs. This lack of adaptability can cause inefficient power usage and greater number of activated UEs and thus reduced overall system performance. From the theoretical perspective, the CA2C-based algorithm performs better than DDQN-based algorithm because it is better suited to handle the joint optimization of discrete and continuous action spaces, and it is able to capture the interdependencies between these two types of actions more effectively.

.

Algorithm 4 RB Allocation and Power Adjustment on the RBs

input : $t, \widehat{p}^k(t), \widetilde{p}^k(t), \alpha_m^k \ \forall k \in \mathcal{K}, \ \forall m \in \mathcal{M}$
output: $\beta_{m,n}^k, \widehat{p}_{1,n}^k(t), \widetilde{p}_{m,n}^k(t)$
for *each gNB* $b \in \mathcal{B}$ **:**
  for *each CC* $m \in \mathcal{M}$ **:**
    For each UE $k \in \mathcal{K}_b^m$, use RR algorithm to allocate a RB $n \in \mathcal{N}_m$ to the UE and set $\beta_{m,n}^k = 1$
  for *each* $k \in \mathcal{K}_b$ **:**
    $$\widehat{p}_{1,n}^k(t) = \frac{\widehat{p}^k(t)}{\sum_{n \in \mathcal{N}_1} \beta_{1,n}^k}; \widetilde{p}_{m,n}^k(t) = \frac{\widetilde{p}^k / \mid \mathcal{M}^k \mid}{\sum_{n \in \mathcal{N}_m} \beta_{m,n}^k}$$

APPENDIX A RB ALLOCATION AND POWER ADJUSTMENT ON THE RBS

As mentioned in [16], the CC management and RB allocation can be performed either jointly or dis-jointly. Similar to [4], we perform CA and RB allocation separately. Given the set of activated CCs, to allocate the RBs in the CCs, we deploy RR algorithm which considerably reduced the complexity of the

Algorithm 5 Critic Network Training Algorithm

input : A sample experience $\mathbf{e}_b^t = [\mathbf{s}, \mathbf{a}_b, \eta_b, \mathbf{s}']$, online network parameter $\boldsymbol{\theta}_b$, target network parameter $\boldsymbol{\theta}_b^-$, parameter τ;
output: Updated parameters for critic online networks and critic target networks;

1 Given $\mathbf{s}'$ and $\boldsymbol{\alpha}_b(t)$ (which corresponds to activating SCCs), obtain the output for the actor online network, i.e., $\mathbf{r}_b(t) = \left[\left(\widehat{r}^k(t), \widetilde{r}^k(t)\right)\right]_{\forall k \in \mathcal{K}_b}$ and calculate $\mathbf{p}_b(t) = \left[\left(\widehat{p}^k(t), \widetilde{p}^k(t)\right)\right]_{\forall k \in \mathcal{K}_b}$ using (13);
2 Given $\mathbf{r}_b(t)$ and $\mathbf{s}'$ as input to critic target network, obtain the Q-function and update $\boldsymbol{\alpha}_b(t)$ based on (15);
3 Given $\boldsymbol{\alpha}_b(t)$ and $\mathbf{s}'$ as input to actor target network, update $\mathbf{r}_b(t)$ and $\mathbf{p}_b(t)$;
4 Obtain the target value for the critic as the summation of $\eta_b(t)$ and the Q-function calculated by the critic's target network, which is multiplied by the discount factor;
5 Employing critic optimizer (e.g., Adam algorithm), update the parameters for the critic online network using (16b);
6 Update the parameter for critic target network as $\mathbf{w}_b^- = \tau \mathbf{w}_b^- + (1-\tau)\mathbf{w}_b$.

Algorithm 6 Actor Network Training Algorithm

input : A sample experience $\mathbf{e}_b^t$, online network parameter $\mathbf{w}_b$, target network parameter $\mathbf{w}_b^-$, parameter τ;
output: Updated parameters for actor online networks and actor target networks;

1 Calculate the gradient of function $J_b(\boldsymbol{\theta}_b)$ with respect to continuous action in all sampled experience form replay buffer $\mathcal{D}_b$ (by using equation (28) in [17]);
2 Using the actor optimizer (e.g., Adam algorithm), update actor online parameter based on (16a);
3 Update the parameter for actor target network as $\boldsymbol{\theta}_b^- = \tau \boldsymbol{\theta}_b^- + (1-\tau)\boldsymbol{\theta}_b$.

CA and RB allocation. Additionally, the UL transmit power level for each UE is equally divided among its activated CCs and its allocated RBs. For a given gNB b, let us denote $\mathcal{K}_b^m$ as the set of UEs sharing a CC m and define it as $\mathcal{K}_b^m = \{k \in \mathcal{K}_b \mid \alpha_m^k = 1\}$. The details of RB allocation and the power adjustment through them are given in **Algorithm 4**.

APPENDIX B ACTOR AND CRITIC TRAINING ALGORITHMS

The algorithms to train the actor and critic networks are given in **Algorithms 6 and 5, respectively.**

REFERENCES

[1] I. de la Bandera, D. Palacios, and R. Barco, "Multinode component carrier management: Multiconnectivity in 5G," *IEEE Veh. Technol. Mag.*, vol. 16, no. 2, pp. 40–47, Jun. 2021.
[2] "5G NR uplink enhancements," MediaTek, Hsinchu, Taiwan, White Paper PDFULEWPA4 0119, 2018. [Online]. Available: https://newsletter.mediatek.com/hubfs/mwc/download/ul-enhancements.pdf
[3] A. Chaudhari, N. Kumar, and P. Rao, "A novel Q- learning assisted dynamic power sharing for dual connectivity scenario," in *Proc. IEEE 17th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2020, pp. 1–6.
[4] F. Khoramnejad, R. Joda, A. B. Sediq, H. Abou-Zeid, R. Atawia, G. Boudreau, and M. Erol-Kantarci, "Delay-aware and energy-efficient carrier aggregation in 5G using double deep Q-networks," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6615–6629, Oct. 2022.
[5] S. Kim, "Two-level game based spectrum allocation scheme for multi-flow carrier aggregation technique," *IEEE Access*, vol. 8, pp. 89291–89299, 2020.
[6] R. Li, P. Hong, K. Xue, M. Zhang, and T. Yang, "Energy-efficient resource allocation for high-rate underlay D2D communications with statistical CSI: A one-to-many strategy," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4006–4018, Apr. 2020.
[7] J. F. Monserrat, F. Bouchmal, D. Martin-Sacristan, and O. Carrasco, "Multi-radio dual connectivity for 5G small cells interworking," *IEEE Commun. Standards Mag.*, vol. 4, no. 3, pp. 30–36, Sep. 2020.
[8] H. Cui and F. You, "User-centric resource scheduling for dual-connectivity communications," *IEEE Commun. Lett.*, vol. 25, no. 11, pp. 3659–3663, Nov. 2021.
[9] C. Li, H. Wang, and R. Song, "Intelligent offloading for NOMA-assisted MEC via dual connectivity," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2802–2813, Feb. 2021.
[10] Z. Gu, H. Lu, P. Hong, and Y. Zhang, "Reliability enhancement for VR delivery in mobile-edge empowered dual-connectivity sub-6 GHz and mmWave HetNets," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2210–2226, Apr. 2022.
[11] A. Khalili, E. M. Monfared, S. Zargari, M. R. Javan, N. M. Yamchi, and E. A. Jorswieck, "Resource management for transmit power minimization in UAV-assisted RIS HetNets supported by dual connectivity," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1806–1822, Mar. 2022.
[12] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992.
[13] J.-S. Liu, C. R. Lin, and Y.-C. Hu, "Joint resource allocation, user association, and power control for 5G LTE-based heterogeneous networks," *IEEE Access*, vol. 8, pp. 122654–122672, 2020.
[14] A. Feriani and E. Hossain, "Single and multi-agent deep reinforcement learning for AI-enabled wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 1226–1252, 2nd Quart., 2021.
[15] M. Elsayed and M. Erol-Kantarci, "AI-enabled future wireless networks: Challenges, opportunities, and open issues," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 70–77, Sep. 2019.
[16] H. Lee, S. Vahid, and K. Moessner, "A survey of radio resource management for spectrum aggregation in LTE-advanced," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 2, pp. 745–760, 2nd Quart., 2014.
[17] J. Hu, H. Zhang, L. Song, R. Schober, and H. V. Poor, "Cooperative Internet of UAVs: Distributed trajectory design by multi-agent deep reinforcement learning," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 6807–6821, Nov. 2020.
[18] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. 30th Conf. Artif. Intell. (AAAI)*, Mar. 2016, pp. 2094–2100.
[19] *Rel-17 Dual Connectivity (DC) of X Bands (X=1,2) LTE Inter-Band CA (xDL/xUL) and Y Bands (y=3-x) NR Inter-Band CA (yDL/yUL)*, document 3GPP TR 37.717-33, Tech. Rep., Jul. 2020.
[20] M. Elsayed, R. Joda, H. Abou-Zeid, R. Atawia, A. B. Sediq, G. Boudreau, and M. Erol-Kantarci, "Reinforcement learning based energy-efficient component carrier activation-deactivation in 5G," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 1–6.
[21] R. Joda, M. Elsayed, H. Abou-Zeid, R. Atawia, A. B. Sediq, G. Boudreau, and M. Erol-Kantarci, "QoS-aware joint component carrier selection and resource allocation for carrier aggregation in 5G," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2021, pp. 1–6.
[22] B. Dusza, C. Ide, L. Cheng, and C. Wietfeld, "CoPoMo: A context-aware power consumption model for LTE user equipment," *Trans. Emerg. Telecommun. Technol.*, vol. 24, no. 6, pp. 615–632, Oct. 2013.

[23] B. Dusza, C. Ide, L. Cheng, and C. Wietfeld, "An accurate measurement-based power consumption model for LTE uplink transmissions," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2013, pp. 49–50.
[24] *Rel-17 Dual Connectivity (DC) of X Bands (x=2,3,4) LTE Inter-Band CA (xDL/1UL) and 1 NR FR1 Band (1DL/1UL) and 1 NR FR2 Band (1DL/1UL)*, document 3GPP TR 37.717-21-22, Tech. Rep., Dec. 2019.
[25] M. Akbari, M. R. Abedi, R. Joda, M. Pourghasemian, N. Mokari, and M. Erol-Kantarci, "Age of information aware VNF scheduling in industrial IoT using deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2487–2500, Aug. 2021.
[26] F. Khoramnejad and M. Erol-Kantarci, "On joint offloading and resource allocation: A double deep Q-network approach," *IEEE Trans. Cognit. Commun. Netw.*, vol. 7, no. 4, pp. 1126–1141, Dec. 2021.
[27] N. Zhao, Y.-C. Liang, D. Niyato, Y. Pei, M. Wu, and Y. Jiang, "Deep reinforcement learning for user association and resource allocation in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5141–5152, Nov. 2019.
[28] 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," 2015, *arXiv:1509.02971*.
[29] V. Mnih, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
[30] R. S. Sutton and A. G. Barto, *Reinforcement Learning, An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
[31] *Study on User Equipment (UE) Power Saving in NR*, document 3GPP TR 38.840, v16.0.0 (Release 16), Jun. 2019.
[32] 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.

FAHIME KHORAMNEJAD received the B.Sc., M.Sc., and Ph.D. degrees in computer engineering from the Amirkabir University of Technology, Tehran, Iran, in 2000, 2009, and 2018, respectively, and the Ph.D. degree in electrical engineering and computer science from the University of Ottawa, ON, Canada, in 2023. From January 2017 to July 2017, she was a Visiting Researcher with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. Her current research interests include resource allocation in next-generation wireless networks, AI-enabled networks, optimization theory, machine learning, deep learning, and reinforcement learning.

ROGHAYEH JODA (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1998, and the M.Sc. and the Ph.D. degrees in electrical engineering from the University of Tehran, Tehran, in 2001 and 2012, respectively. She was a Visiting Scholar with the Polytechnic Institute, NYU, Brooklyn, NY, USA. From September 2013 to August 2014, she was a Postdoctoral Fellow with the University of Padua, Padua, Italy. In November 2014, she joined the ICT Research Institute, Tehran, as a Research Assistant Professor. Then, she was a Visiting Researcher with the University of Ottawa, Ottawa, Canada, and in June 2022, she joined Ericsson, Ottawa, as a System Developer. Her current research interests include communication theory, information theory, resource allocation, optimization and machine learning with application to wireless networks, and 5G and 6G networks.

AKRAM BIN SEDIQ is currently a Senior Specialist in radio access network (RAN) edge analytics with Ericsson Canada, where he is securing embedded machine intelligence in RAN to facilitate automated and adaptive RAN solutions, driving early-phase system designs of 4G/5G wireless systems, generating patents, and managing university collaborations. His research interests include radio resource management (RRM), machine-learning-based RRM, data analytics, optimization, dual connectivity, control channel design, energy-saving, and constellation design. His work resulted in more than 30 granted and filed patents and more than 40 peer-reviewed publications.

GARY BOUDREAU (Senior Member, IEEE) received the B.A.Sc. degree in electrical engineering from the University of Ottawa, in 1983, the M.A.Sc. degree in electrical engineering from Queens University, in 1984, and the Ph.D. degree in electrical engineering from Carleton University, in 1989. From 1984 to 1989, he was a Communications Systems Engineer with Canadian Astronautics Ltd., and from 1990 to 1993, he was a Satellite Systems Engineer with MPR Teltech Ltd. From 1993 to 2009, he was with Nortel Networks in a variety of wireless systems and management roles within the CDMA and LTE basestation product groups. In 2010, he joined Ericsson Canada, where he is currently the Director of RAN Architecture and Performance with the North American CTO Office. His research interests include digital and wireless communications, signal processing, and machine learning.

MELIKE EROL-KANTARCI (Senior Member, IEEE) is currently the Canada Research Chair of AI-Enabled Next-Generation Wireless Networks and an Associate Professor with the School of Electrical Engineering and Computer Science, University of Ottawa. She is also the Founding Director of the Networked Systems and Communications Research (NETCORE) Laboratory. She has received numerous awards and recognitions. She is the co-editor of three books on smart grids, smart cities, and intelligent transportation. She has over 180 peer-reviewed publications. She has delivered more than 70 keynotes, plenary talks, and tutorials around the globe. She is on the editorial board of the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, IEEE INTERNET OF THINGS JOURNAL, IEEE COMMUNICATIONS LETTERS, IEEE NETWORKING LETTERS, *IEEE Vehicular Technology Magazine*, and IEEE ACCESS. She has acted as the general chair and the technical program chair for many international conferences and workshops. Her research interests include AI-enabled wireless networks, 5G and 6G wireless communications, smart grid, and the Internet of Things. She is an IEEE ComSoc Distinguished Lecturer and an ACM Senior Member.