World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:17, No:9, 2023

AI-based Radio Resource and Transmission Opportunity Allocation for 5G-V2X HetNets: NR and NR-U networks

Farshad Zeinali, Sajedeh Norouzi, Nader Mokari, Eduard A. Jorswieck

Abstract—The capacity of fifth-generation (5G) vehicle-to-everything (V2X) networks poses significant challenges. To address this challenge, this paper utilizes New Radio (NR) and New Radio Unlicensed (NR-U) networks to develop a vehicular heterogeneous network (HetNet). We propose a framework, named joint BS assignment and resource allocation (JBSRA) for mobile V2X users and also consider coexistence schemes based on flexible duty cycle (DC) mechanism for unlicensed bands. Our objective is to maximize the average throughput of vehicles, while guarantying the WiFi users throughput. In simulations based on deep reinforcement learning (DRL) algorithms such as deep deterministic policy gradient (DDPG) and deep Q network (DQN), our proposed framework outperforms existing solutions that rely on fixed DC or schemes without consideration of unlicensed bands.

Keywords—Vehicle-to-everything, resource allocation, BS assignment, new radio, new radio unlicensed, coexistence NR-U and WiFi, deep deterministic policy gradient, Deep Q-network, Duty cycle mechanism.

I. Introduction

FIFTH-GENERATION (5G) networks are designed to provide high-speed, low-latency, and reliable communication services to a wide range of applications, including the internet of things (IoT), virtual and augmented reality (VR/AR), and vehicle-to-everything (V2X) communications. However, the increasing demand for high-quality services and the exponential growth of connected devices and data traffic pose significant challenges for the capacity and spectral efficiency of 5G networks. Traditional approaches to network design and optimization may not be sufficient to address the capacity challenges of 5G networks, especially in dense urban areas with high user density and traffic volume. The deployment of small cell and heterogeneous network (HetNet) has emerged as a promising solution to address the capacity challenges of 5G networks.

To provide 5G HetNet, we utilize new radio unlicensed band (NR-U) for additional frequency capacity in order to offer a wider spectrum with higher data rates, beside new radio (NR) network. Our main challenge when using unlicensed bands is fair spectrum sharing, due to the existence of other wireless networks such as WiFi. Coexistence mechanisms are required to ensure that different wireless technologies can coexist and share the spectrum without causing interference. Listen before talk (LBT) and carrier sensing adaptive transmission (CSAT) are two mechanisms that have been proposed for coexistence between NR-U and WiFi. LBT involves sensing the channel before transmitting and waiting for a random time if the channel is occupied [1]. Due to the possibility that LBT may not enable NR-U cells to transmit even if necessary, it is not suitable for ultra-reliable low latency communication (URLLC) 5G users. In contrast, CSAT improves coexistence performance and spectral efficiency [2] by sensing the channel state and adapting the duty cycle (DC) of transmission [3]. However, sensing the channel state faces some problems like miss detection and false alarm. As a result, coexistence mechanisms such as LBT and CSAT are highly dependent on many factors including access points, user density, and traffic volume. Therefore, further research is needed to design adequate coexistence mechanisms that can support the increasing demand for high-quality wireless services in dense urban areas [4]. One promising approach to design coexistence mechanisms in 5G networks is the use of machine learning (ML), such as deep reinforcement learning (DRL), to make optimal DC of transmission based on the current traffic and network conditions. DRL has shown great potential in addressing the coexistence problem in 5G networks, especially for V2X applications.

A. Related Works

1) Het-Nets: Several studies have focused on optimizing resource allocation and base station (BS) assignment in HetNets. In [5], the authors presented a two-tier network as well as a BS assignment algorithm to maximize the throughput. Similarly, in [6], the authors presented an optimized joint uplink/downlink resource allocation scheme for orthogonal frequency-division multiple access (OFDMA) networks, which maximizes the sum rate while ensuring fairness. In [7], a sum rate maximizing cell association algorithm was proposed, which assigns users to BSs based on their channel gains. Additionally, a user association algorithm was proposed in [8] for load balancing in HetNets.

2) Unlicensed band: There has been significant research on coexistence mechanisms between NR-U and WiFi in unlicensed bands. In [9], the authors proposed a mechanism for the coexistence of NR-U and WiFi systems in unlicensed bands by allocating bandwidth and transmission opportunities

F. Zeinali, S. Norouzi and N. Mokari are with the Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran, Postal Code: 14115, (e-mail: {zeinali, sajede.noroozi, nader.mokari}@modares.ac.ir).

Eduard A. Jorswieck is with TU Braunschweig, Department of Information Theory and Communication Systems, Braunschweig, Germany (e.jorswieck@tu-braunschweig.de).

TABLE I
COMPARISON OF RELATED WORKS

Paper	*V2X network*	*HetNet*	*Unlicensed band*	*Flexible duty cycle mechanism*	*Resource allocation*	*ML solution*
[5]	×	√	×	×	×	×
[12]	√	×	√	√	×	√
[18]	√	×	√	×	√	×
[10]	×	×	√	√	√	√
[15]	×	×	√	√	×	√
[7]	×	√	×	×	×	×
[17]	×	×	√	×	×	×
Our work	√	√	√	√	√	√

to improve throughput and fairness for both systems. A similar problem was also solved by [10] using DRL algorithms. The authors of [11] presented a method for enabling NR-U to operate in unlicensed spectrum using LBT switching procedures, which can effectively reduce the interference between NR-U and WiFi. In [12], the authors proposed an algorithm based on Q-learning for coexistence between long-term evolution Unlicensed (LTE-U) and WiFi in multi-channel environments, which can optimize the coexistence performance and improve the throughput.

The authors of [13] investigated the impact of WiFi transmissions on Cellular-V2X (C-V2X) performance and demonstrated that coexistence was possible with proper interference management. Furthermore, a coexistence algorithm based on Q-learning for LTE-U and WiFi was proposed in [14], which can optimize transmission power and channel selection to mitigate interference. In [15], the authors proposed a ML-based discontinuous reception (DRX) mechanism for NR-U networks, to improve energy efficiency and reduce interference with WiFi networks. The authors of [16] proposed a coexistence mechanism that assigns multiple bandwidth parts to NR-U and WiFi systems in the unlicensed band by jointly optimizing bandwidth assignment and transmission parameters. A contention resolution algorithm was proposed in [17] for NR-U in shared sub-7 GHz bands, based on gap-based channel access. The algorithm considered channel occupancy information of both NR-U and incumbent WiFi networks to improve coexistence performance.

Overall, these studies demonstrate the importance of coexistence mechanisms to ensure fair sharing of unlicensed spectrum. This will improve the performance of wireless systems in dense urban areas. However, there is still a need for further research to address the challenges of coexistence in the unlicensed band, particularly in the context of 5G systems. In Table I, we compare our work with the most relevant previous research. To the best of our knowledge, no study has investigated joint BS assignment and resource allocation (JBSRA) with considering flexible DC coexistence mechanism in unlicensed bands for V2X networks.

B. Contribution

In this paper, we propose a DRL-based JBSRA framework for NR and NR-U networks in the context of a HetNet with V2X users. Our framework considers the coexistence problem between NR-U and WiFi users, which we address using a flexible DC mechanism. To maximize V2X network throughput, we formulate an optimization problem and utilize deep deterministic policy gradient (DDPG) and deep Q-network (DQN)-DDPG algorithms to solve that. The contributions of this work include the following key items:

- We design a two-step DDPG algorithm, with a first step being BS assignment and a second step being resource allocation, in the context of a HetNet that supports V2X.
- A RL-based flexible DC mechanism is proposed, which takes into account WiFi throughput to mitigate the coexistence issue.
- We propose the DQN-DDPG module, which employs DQN for BS assignment and DDPG for resource allocation, and evaluate its performance against the DDPG module.

II. SYSTEM MODEL AND PROBLEM FORMULATION

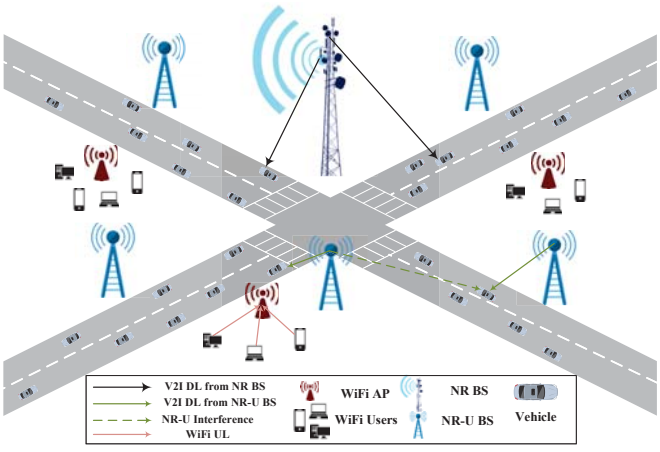

Fig. 1 System model

A. 5G-HetNet

To support the large number of users in 5G, we propose a two-tier HetNet that consists of B_1 macro (NR) BSs and B_2 micro (NR-U) BSs in downlink as shown in Fig. 1. The sets of BSs are denoted as $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2\}$, where $\mathcal{B}_1 = \{1, 2, ..., B_1\}$ and $\mathcal{B}_2 = \{1, 2, ..., B_2\}$ represent the set of macro and micro BSs, respectively. We also define $\mathcal{R} = \{\mathcal{R}_1, \mathcal{R}_2\}$, where $\mathcal{R}_1 = \{1, 2, ..., R_1\}$ and $\mathcal{R}_2 = \{1, 2, ..., R_2\}$ represent the set of

resource blocks (RBs) for macro and micro BSs, respectively. The total number of RBs for macro and micro BSs are denoted by R_1 and R_2. In addition, $\mathcal{V} = \{1, 2, \ldots, v, \ldots, V\}$ demonstrates the set of vehicles, where V represents the total number of vehicles. To indicate the BS and RB assignment, we define the binary variable $\eta^t_{v,b}[r] \in \{0, 1\}$ for vehicle v at time slot t. We have $\eta^t_{v,b}[r] = 1$ if BS $b \in \mathcal{B}$ uses RB $r \in \mathcal{R}$ to transmit data to vehicle v, and $\eta^t_{v,b}[r] = 0$ otherwise. We consider each time slot as 1 ms and contains 12 RBs with a bandwidth of 15 kHz (for a total bandwidth of 180 kHz) according to the 3rd generation partnership project (3GPP) 5G standard [19]. OFDMA is used as multiple access to increase the capacity of the system and enable more users to access the network. Thus, each user can use multiple RBs, but each RB can only be assigned to one user. We calculate the inter-cell interference, as follows:

$$I^t_{v,b}[r] = \begin{cases} \sum_{\substack{\hat{v}\in\mathcal{V}\\ \hat{v}\neq v}} \sum_{\substack{\hat{b}\in\mathcal{B}_1\\ \hat{b}\neq b}} \eta^t_{\hat{v},\hat{b}}[r]\, P_{\hat{v},\hat{b},r}\, h_{v,\hat{b},r}, & \text{if } b \in \mathcal{B}_1, \\ \sum_{\substack{\hat{v}\in\mathcal{V}\\ \hat{v}\neq v}} \sum_{\substack{\hat{b}\in\mathcal{B}_2,\\ \hat{b}\neq b}} \eta^t_{\hat{v},\hat{b}}[r]\, P_{\hat{v},\hat{b},r}\, h_{v,\hat{b},r}, & \text{if } b \in \mathcal{B}_2, \end{cases} \tag{1}$$

where $P_{v,b,r}$ and $h_{v,b,r}$ denote the transmit power and channel gain from BS b to vehicle v on RB r, respectively. The signal-to-interference-plus-noise-ratio (SINR) of vehicle v at time slot t is calculated as follows:

$$\gamma^t_{v,b}[r] = \frac{\eta^t_{v,b}[r]\, P_{v,b,r}\, h_{v,b,r}}{I^t_{v,b}[r] + \sigma^2}, \quad \forall v \in \mathcal{V}, \forall b \in \mathcal{B}, \forall r \in \mathcal{R}, \tag{2}$$

where σ^2 is the noise power. Data rate of vehicle v for BS b on RB r can be expressed using the Shannon capacity formula:

$$R^t_{v,b}[r] = Bw \log_2[1 + \gamma^t_{v,b}[r]], \tag{3}$$

where Bw is the bandwidth for each RB, we can describe the total rate for each vehicle at time slot t as:

$$R^t_v = \sum_{b\in\mathcal{B}} \sum_{r\in\mathcal{R}} \eta^t_{v,b}[r] R^t_{v,b}[r], \tag{4}$$

we can also compute the total and mean data rate as:

$$R^t_{\text{Total}} = \sum_{v\in\mathcal{V}} R^t_v, \tag{5}$$

$$\overline{R}^t = \frac{1}{V} R^t_{\text{Total}}. \tag{6}$$

B. RL-Based Duty Cycle Coexistence Mechanism

As mentioned previously, each vehicle can use licensed (NR) or unlicensed bands (NR-U) to receive packets. However, since WiFi also uses unlicensed bands, if a vehicle uses these bands to receive packets, there will be a collision with WiFi. To address this issue, we propose a DC model based on reinforcement learning (RL) for both WiFi and V2X networks. According to Fig. 2, each time slot is divided into two parts: I) NR-U and WiFi, and II) only WiFi. When part I is enabled, only NR-U users are allowed to access unlicensed bands. The

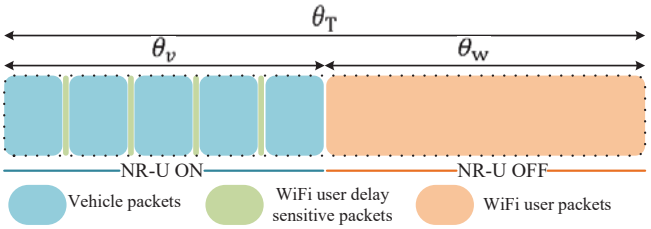

Fig. 2 Duty cycle mechanism

reason is WiFi users that utilize the Carrier-Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol and are not allowed to use the channels during busy periods. However, sub-frame puncturing for delay-sensitive data on the WiFi network is being considered [20]. In Part II, NR-U is disabled, allowing WiFi users to access unlicensed bands easily. The duration of one time slot θ_T, is fixed, and the duration of NR-U θ^t_v and WiFi parts θ^t_w, are flexible at each time slot t.

$$\theta_T = \theta^t_v + \theta^t_w. \tag{7}$$

$R^t_{\text{WiFi}}[v]$ is considered as the data rate of user within proximity of vehicle v at time slot t. In accordance with [21], we can express the throughput of each WiFi user as follows:

$$W^t_v = \theta^t_w \times R^t_{\text{WiFi}}[v], \tag{8}$$

where θ^t_w represents the duration of time that the WiFi user occupies the channel, and $R^t_{\text{WiFi}}[v]$ is a random uniform data rate of the WiFi user v between 600 Mbps and 1200 Mbps, depending on the modulation and coding methods used at time slot t, according to IEEE 802.11ax (WiFi 6) [22]. Therefore, we can calculate the average throughput of WiFi users as:

$$\overline{W}^t = \frac{\sum_v W^t_v}{V}, \tag{9}$$

Similarly, we define the throughput for each vehicle and average throughput of vehicles as follows:

$$T^t_v = \theta^t_v \times R^t_v, \tag{10}$$

$$\overline{T}^t = \frac{\sum_v T^t_v}{V}, \tag{11}$$

where θ^t_v represents the duration time of vehicles' usage, R^t_v is the data rate of vehicle v at time slot t.

C. Optimization Problem

The optimization problem can be expressed as follows:

$$\max_{\boldsymbol{\eta},\boldsymbol{p},\boldsymbol{\theta}} \left\{ \sum_{t=1}^{T} \overline{T}^t \right\} \tag{12a}$$

$$\text{s.t. } C1: \overline{T}^t \geq \underline{T}, \quad \forall t \in T, \tag{12b}$$

$$C2: \overline{W}^t \geq \underline{W}, \quad \forall t \in T, \tag{12c}$$

$$C3: \sum_{v=1}^{V} \sum_{r=1}^{R} \eta^t_{v,b}[r] P^t_{v,b,r} \leq P_{\text{Max}}, \quad , \forall b \in \mathcal{B}, \tag{12d}$$

$$C4: \sum_{v=1}^{V} \eta^t_{v,b}[r] \leq 1, \quad \forall b \in \mathcal{B}, \forall r \in \mathcal{R}, \forall t \in T. \tag{12e}$$

TABLE II
NOTATIONS USED IN THE PAPER

Notation	*Definition*
V / $\mathcal{V}$ / v	Number/set/index of vehicles
B / $\mathcal{B}$ / b	Number/set/index ofs BSs
B_1 / $\mathcal{B}_1$ / b_1	Number/set/index of macro BSs
B_2 / $\mathcal{B}_2$ / b_2	Number/set/index of micro BSs
R / $\mathcal{R}$ / r	Number/set/index of RBs
R_1 / $\mathcal{R}_1$ / r_1	Number/set/index of RBs for macro
R_2 / $\mathcal{R}_2$ / r_2	Number/set/index of RBs for micro
$h_{v,b,r}$	Channel gain from BS b to vehicle v in RB r
$P_{v,b,r}$	Power usage of vehicle v from BS b in RB r
W_v	Throughput of WiFi user within proximity of vehicle v
$\gamma_{v,b}[r]$	SINR between vehicle v and BS b in RB r
$R_{v,b}[r]$	Data rate between vehicle v and BS b in RB r
$\eta_{v,b}[r]$	BS and RB allocation for vehicle v indicator
θ_v/θ_w	Duty cycle of vehicle/WiFi user indicator

The optimization objective is to maximize the average throughput of vehicles. Constraints (12b) and (12c) ensure the throughput of vehicles and WiFi users, respectively. Constraint (12d) is defined to ensure that each BS cannot exceed the maximum transmit power value. Constraint (12e) indicates each RB cannot be assigned to more than one vehicle according to OFDMA.

III. SOLUTION

The purpose of this section is to solve the problem of resource allocation presented in (12a). We have designed DRL algorithms to address this optimization problem.

In this section, we discuss our DRL solutions, DDPG, and combined DQN-DDPG. We also discuss their state space, actions, and reward functions and computational complexity of the utilized algorithms.

A. Deep Deterministic Policy Gradient

DDPG is a type of DRL that learns a Q-function and a policy simultaneously. In this subsection, we explain how it works. DDPG has an actor network and a critic network. The actor decides which action to take, and critics inform the actor about how good the action was and how it needs to be adjusted. The DDPG model is composed of three key elements: state space, action space, and immediate reward.

1) State Space: The observed state of each agent for each vehicle v at time slot t consists of four components: (i) $h^t_{v,b,r}$, which indicates the instant channel information when it is connected to BS b on RB r; (ii) the throughput of the vehicle, T^t_v; (iii) the previous interference from other BSs to vehicles v on the same RB r, $I^{t-1}_{v,b}[r]$; and (iv) the throughput of WiFi users associated with vehicle v used during the previous time slot, W^{t-1}_v. Therefore, the state space of the agent can be described as follows:

$$s^t = \{s^t_1, ..., s^t_v, ..., s^t_V\},$$
$$s^t_v = [h^t_{v,b,r}, T^t_v, I^{t-1}_{v,b}[r], W^{t-1}_v].$$

2) Action space: In each time slot t, based on the observed state, the agent can select actions; BS assignment and RB allocation for vehicle v indicated by η^t_v, the allocated power for transmission p^t_v, DC allocation only for vehicles connected to micro BS θ^t_v. Thus, the agent's action space can be defined as follows:

$$a^t = \{a^t_1, ..., a^t_v, ..., a^t_V\},$$
$$a^t_v = [\eta^t_v, p^t_v, \theta^t_v].$$

3) Immediate reward: In RL algorithms, the agent learns through a reward provided by the environment. In this case, to maximize the mean throughput of the vehicles, we use the immediate reward at each time slot t:

$$r^t = \frac{\overline{T}^t}{\underline{T}} G(W^t_v - \underline{W}) - \kappa F\{\eta^t_v\}, \tag{13}$$

where $\underline{T}$ and $\underline{W}$ are the minimum thresholds for the throughput of vehicles and WiFi users, respectively. The step function G satisfies constraint (12c), and function F satisfies constraint (12e), which are given below by (14) and (15), respectively. Finally, κ is the weight of penalty term used to balance the reward:

$$G(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases} \tag{14}$$

$$F\{\eta^t_v\} = \begin{cases} 1, & \sum_{r \in R} \eta^t_{v,b}[r] \geq 2\,, \\ 0, & \& \text{otherwise}. \end{cases} \tag{15}$$

In DDPG, the actor network is used to generate the action deterministically a^t, and the critic network is used to evaluate the rewards of state-action pair (s^t, a^t). Target networks are also used to enhance the stability of actor and critic networks. The actor network makes its decision based on:

$$a^t = \pi(s^t, \psi). \tag{16}$$

In this case, π represents the policy of the actor network, while ψ represents the weight of the actor network. At each time slot, the agent takes an action a^t, receives an immediate reward r^t, and transitions to a new state $s^{t+1} = \{s^{t+1}_1, s^{t+1}_2, ..., s^{t+1}_V\}$. The experience (s^t, a^t, r^t, s^{t+1}) is then stored in the replay memory buffer D with a size of N. The agent randomly selects a sample i from the buffer, then the actor network can learn from the following loss function:

$$L^{Actor}_{\psi} = -(Q(s^i, \pi(s^i, \psi), \varphi)), \tag{17}$$

where Q is the estimated value of the critic network with weights φ which can evaluate the policy of the actor network. The agent aims to maximize Q or minimize $-Q$ to update the weights of the actor network via the gradient:

$$\psi \leftarrow \psi - \tau_1 \nabla_{\psi} L^{\text{Actor}}_{\psi}. \tag{18}$$

The critic network can learn using the loss function:

$$L^{Critic}_{\varphi} = \sum_i ((r^i + \gamma \tilde{Q}(s^{i+1}, \tilde{\pi}(s^{i+1}, \tilde{\psi})\tilde{\varphi})) - Q(s^i, a^i, \varphi))^2, \tag{19}$$

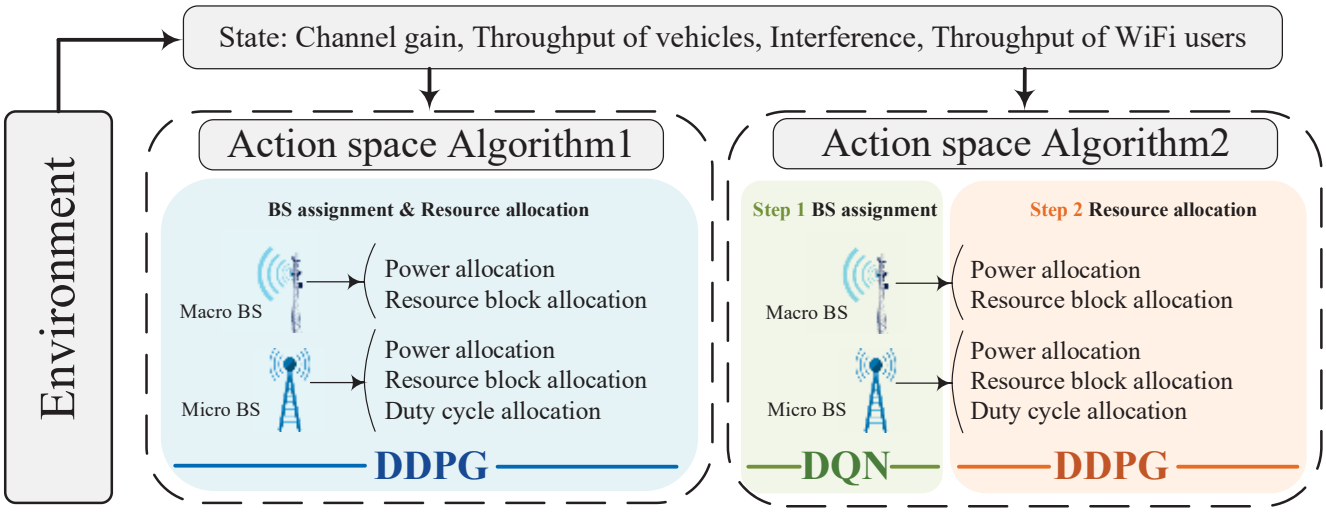

Fig. 3 Frameworks of Algorithms

where $\tilde{Q}$ is the Q-value of the target critic network, and $\tilde{\varphi}$ and $\tilde{\psi}$ are the weights of the target critic network and target actor network, respectively. The weights of the critic network can be updated via the gradient:

$$\varphi \leftarrow \varphi - \tau_2 \nabla_{\varphi} L_{\varphi}^{\text{Critic}}. \tag{20}$$

At the end, the weights of the actor and critic target networks are updated by:

$$\tilde{\psi} \leftarrow \psi + (1-\tau_3)\tilde{\psi}, \tag{21}$$

$$\tilde{\varphi} \leftarrow \varphi + (1-\tau_3)\tilde{\varphi}, \tag{22}$$

where τ_3 controls the fraction of the target network's weights to copy from the main networks. In the next time slot, the algorithm generates new experiences and updates the weights of the networks using new samples from the batch. In summary, all the previous explanations are found in Subsection III-A3.

Initialize the weights ψ of actor network and the weights φ of critic network.
Initialize the weights $\tilde{\psi}$ of target actor network and the weights $\tilde{\varphi}$ of target critic network, $\tilde{\psi} = \psi$, $\tilde{\varphi} = \varphi$.
Initialize the parameters of environment.
for each episode **do**
  Randomly initialize the positions and directions of vehicles.
  for each step t **do**
    Update position of vehicles, according to their directions and speeds.
    Consider $s^t = \{s_1^t, s_2^t, ..., s_V^t\}$ and choose action $a^t = \Pi(s^t \mid \phi)$ according to current policy and exploration noise.
    for each step v **do**
      Evaluate next state s_v^{t+1} using the action a^t.
    end for
    receive $s^{t+1} = \{s_1^{t+1}, s_2^{t+1}, ..., s_V^{t+1}\}$ and
    calculate reward r^t by (13).
    Store transition (s^t, a^t, r^t, s^{t+1}) in replay buffer D with size N.
    Sample minibatch of size B, from D.
    Update weights ψ and φ by minimizing the loss function in 17, (19).
    Update weights $\tilde{\psi}$ and $\tilde{\varphi}$ by (21), (22).
  end for
end for

B. DQN-DDPG

In this section, we explain a combined DQN-DDPG algorithm. The DQN consists of a main network and a target network, both with the same Deep Neural Network (DNN) architecture as shown in Table III. Since the DQN has a discrete output, we use it only for selecting the type of BS assignment action $a_1^t[v]$. However, DDPG produces a continuous output, so it selects the other actions $a_2^t[v]$.

At the beginning of each time slot, DQN selects an action $a_1^t = \{a_1^t[1], a_1^t[2], ..., a_1^t[V]\}$ randomly with probability ϵ, or according to (23) with probability $1-\epsilon$, based on the ϵ-greedy policy. The variable ϵ balances the exploration and experience. Initially, $\epsilon = 1$, and actions are chosen randomly. After each time slot, ϵ is reduced by the epsilon decreasing rate, combining previous experiments with exploration, until it reaches the minimum specified value $\epsilon_{\text{min}} \geq 0$.

$$a_1^t = \arg\max_{a_1}(Q(s^t, a_1^t, \phi)), \tag{23}$$

where $Q(s^t, a^t, \phi)$ is the output Q-value based on the observed state and action of the main Q network with weights ϕ. In our combined algorithm, at each time slot t, $a^t = \{a_1^t, a_2^t\}$ denotes actions, where a_1^t and a_2^t, are selected ones by DQN and DDPG, receptively. Following the same process in the DDPG algorithm, the experience (s^t, a^t, r^t, s^{t+1}) is stored in the replay memory of DQN and DDPG buffers.

In the same manner as the DDPG algorithm, random selected experiences of size B from the replay memory are

utilized to learn the main DQN network such that its loss function can be expressed as follows:

$$L_{\phi}^{\text{Main}} = \sum_{i=1}^{B}((r^i + \tilde{Q}(s^{i+1}, a^{i+1}, \tilde{\phi})) - Q(s^i, a^i, \phi))^2, \quad (24)$$

where $\tilde{Q}(s^i, a^i, \tilde{\phi})$ is the Q-value of the target DQN network with weight $\tilde{\phi}$ based on s^i and a^i. Thus, by minimizing the loss function using the gradient descent method, we can update the weights of the Q network as follows:

$$\phi \leftarrow \phi - \tau_4 \nabla_{\phi} L_{\phi}^{\text{Main}}, \quad (25)$$

where τ_4 represents the learning rate, and $\tilde{\phi}$ is updated periodically by copying ϕ. In summary, all the previous explanations are found in Subsection III-B.

1: Initialize the weights ϕ for DQN and the weights ψ and φ for actor network and critic network of DDPG respectively.
2: Initialize the weights $\tilde{\phi}$ of target network for DQN and $\tilde{\psi}$, $\tilde{\varphi}$ of target networks for DDPG, $\tilde{\phi} = \phi$, $\tilde{\psi} = \psi$, $\tilde{\varphi} = \varphi$.
3: Initialize the parameters of environment.
4: **for** each episode **do**
5: Randomly initialize the positions and directions of vehicles.
6: **for** each step t **do**
7: Update position of vehicles, according to their directions and speeds.
8: Observe $s^t = \{s_1^t, s_2^t, ..., s_V^t\}$ and choose BS action as a_1^t following the ϵ-greedy policy based on s^t
9: choose action $a_2^t = \Pi(s^t \mid \phi)$ according to current policy and exploration noise for power, RB and DC allocation.
10: **for** each vehicle v **do**
11: Evaluate next state s_v^{t+1} using the action $a^t = \{a_1^t, a_2^t\}$
12: **end for**
13: receive $s^{t+1} = \{s_1^{t+1}, s_2^{t+1}, ..., s_V^{t+1}\}$ and calculate reward r^t by()
14: Store transition $(s^t, a_1^t, r^t, s^{t+1})$ in replay buffer D_1 with size N_1.
15: Store transition $(s^t, a_2^t, r^t, s^{t+1})$ in replay buffer D_2 with size N_2.
16: Sample minibatch of size B_1, from D_1.
17: Sample minibatch of size B_2 from D_2.
18: Update weights ϕ by (25).
19: Update weights ψ and φ by minimizing the loss function in (17), (19) respectively.
20: Update the weights $\tilde{\phi}$ of the target network of DQN every 100 steps $\tilde{\phi} \leftarrow \phi$.
21: Update weights $\tilde{\psi}$ and $\tilde{\varphi}$ by (21), (22) respectively.
22: **end for**
23: **end for**

C. Computational Complexity

In this section, we investigate the computational complexity of our proposed algorithms, which is comprised of the size of state and action spaces, and training process. The computational complexity of the DDPG algorithm can be approximated as: $\mathcal{O}(n \times (m/b) \times (s+a+h_1+h_2))$, where n, m, b, s, a, are the number of iterations, the size of the replay buffer, the batch size, the size of the state space and the size of the action space, respectively [23]. Furthermore, h_1 and h_2 indicate the number of neurons in the first and second hidden layer of the critic network. Consequently, we can express the computational complexity of DQN-DDPG algorithm as: $\mathcal{O}(n \times [(\tilde{m}/\tilde{b}) \times (\tilde{s}+\tilde{a}+\sum h_i) + (m/b) \times (s+a+h_1+h_2)])$, where $\tilde{s}$ and $\tilde{a}$ are the size of state and action space in DQN algorithm. Additionally, h_i denotes the number of neurons in the i_{th} hidden layer of the network. Table III summarizes the complexity comparison with or without DC allocation in both algorithms. Hence, DC allocation can increase fairness performance at the expense of a slight increase in complexity.

IV. Simulation Results

This section demonstrates the effectiveness of our DRL models. We implement our algorithm using Python 3.7.9 and the Spyder IDE version 5.2.2. The simulation is conducted on a square area of $(1000 \times 1000\, m^2)$, where vehicles are randomly distributed with random directions and fixed speeds. Each vehicle moves 0.01 meters in each 1 ms time slot, resulting in a vehicle speed of 36 km/h. In order to comply with the 3GPP standard, we use the following parameters as shown in Table IV: the carrier frequency is 5 GHz, the total number of RBs for each BS is 12, and the maximum power of the vehicles is 30 dBm. The antenna heights for the BS and vehicles are 25 and 1.5 meters, respectively.

According to the V2X communication papers [24], our DQN is a five-layer fully connected neural network with three hidden layers, each having 256 neurons. The actor of DDPG has two hidden layers (1024, 512), and the critic network has two hidden layers (512, 256). Both DQN and DDPG use the rectified linear unit (ReLU) as the activation function for the hidden layers. The learning rates of the DQN network, the actor and critic networks of DDPG are 0.003, 0.0001, and 0.001, respectively. We briefly explain the four baselines in the paper:

- **The DQN-DDPG Model:** Based on this model, DQN is used to assign BS and DDPG is used to assign the remaining actions, such as RB assignment, power allocation, and DC allocation.
- **Without Het-Net network:** This method aims to achieve results without using unlicensed spectrum.
- **Fixed DC:** The baseline entails fixing the DC at 0.5 for all unlicensed users at each time slots.
- **Random DC:** In this baseline, DCs are randomly selected for all unlicensed band users at each time slots.

Fig. 4 represents the converge results of the proposed DDPG and DQN-DDPG algorithms, expressed as the average reward over the learning episodes. The results demonstrate that the DDPG algorithm performs better than the DQN-DDPG algorithm in terms of convergence. However, both algorithms achieve similar results in average data rates. The average

TABLE III
COMPLEXITY COMPARISON

Algorithm	*Computational complexity*	**Complexity compared to the same algorithm without DC allocation**	**Fairness compared to the same algorithm without DC allocation**
DDPG with DC allocation	$\mathcal{O}(n \times (m/b) \times (s+a+h_1+h_2))$	3.08 % worse	17.95 % better
DDPG without DC allocation		-	-
DQN-DDPG with DC allocation	$\mathcal{O}(n \times [(\tilde{m}/\tilde{b}) \times (\tilde{s}+\tilde{a}+\sum h_i)$	10.14 % worse	9.16 % better
DQN-DDPG without DC allocation	$+(m/b) \times (s+a+h_1+h_2)])$	-	-

TABLE IV
SIMULATION PARAMETERS

Environment parameters	*Value*
Carrier frequency	5 GHz
Number of RBs for each BS	12
Bandwidth of each RB	15 kHz
Area environment	$1000 \times 1000\ m^2$
Number of vehicles	5, 10, 15, 20
Vehicles speed	36 km/h
BSs and vehicles antenna height	25, 1.5 m
BS and vehicles antenna gains	8, 3 dBi
BS and vehicles receiver noise figure	5, 9 dB
Vehicles maximum power	30 dBm
Vehicles mobility model	Urban case
Noise power σ^2	-114 dBm
Path loss model	128.1 + 37.6log(d)
Shadowing distribution	log-normal
Shadowing standard deviation	8 dB
Decorrelation distance	50 m
Pathloss/shadowing update	Every 100 ms
Fast fading update	Every 1 ms
Fast fading	Rayleigh fading
DNN parameters	***Value***
Experience replay buffer size	10000
Mini batch size	64
Number/size of DQN networks hidden layers	3 / 256, 256, 256
Number/size of actor DDPG networks hidden layers	2 / 1024, 512
Number/size of critic DDPG networks hidden layers	2 / 512, 256
DQN networks learning rate	0.001
Epsilon decreasing rate	0.0005
Minimum epsilon rate	0.01
Critic/Actor networks learning rate	0.001/0.0001
Target networks update parameter	0.0005
Discount factor	0.99
Number of episodes	250
Number of steps per episode	100

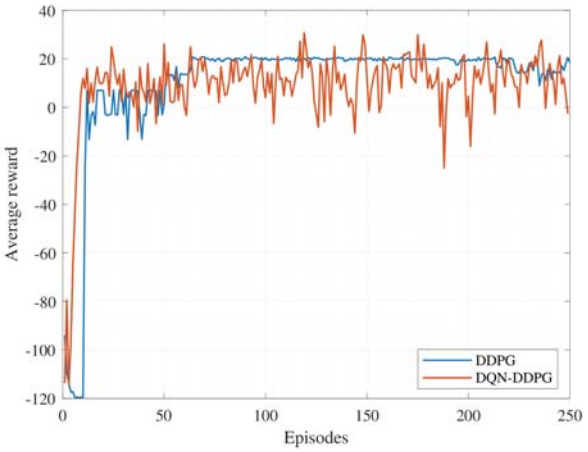

Fig. 4 Mean reward per episodes

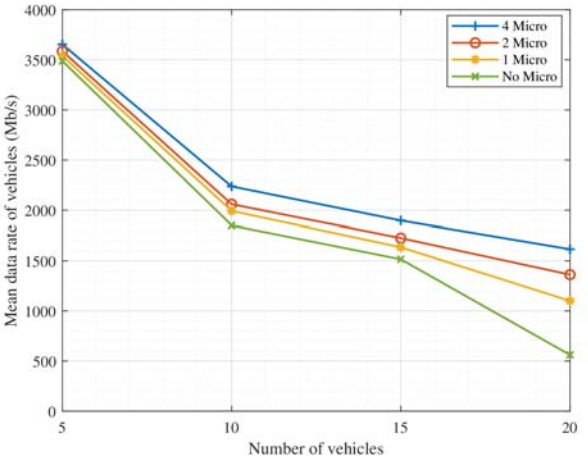

Fig. 5 Mean rate per number of vehicles

data rate of vehicles with varying numbers of micro BSs is presented in Fig. 5. As expected, the data rate decreases as the number of vehicles increases. However, the incorporation of multiple micro BSs can improve data transmission rates by providing multiple RBs for vehicles. The figure clearly shows that the reduction in data rate is less noticeable when four micro BSs are present compared to just one micro BS. Furthermore, when the network is not hybrid (no micro BS), the data rate is at its lowest point. This highlights the benefit of using unlicensed and licensed spectrum simultaneously.

To demonstrate the impact of flexible DC on the fairness factor between vehicles and WiFi users, we utilize Jain's fairness index which is a popular metric used to measure the fairness of multiple users as follows [25]:

$$F(x) = \frac{(\sum_{i=1}^{N} x_i)^2}{N \sum_{i=1}^{N} x_i^2}. \quad (26)$$

The index ranges from 0 to 1, where 0 indicates complete unfairness, and 1 indicates complete fairness. In (26), N is the number of total users (vehicles and WiFi users), and x_i represents the throughput for user i. As illustrated in Fig. 6, our simulation results reveal that the use of a fixed DC or random DC can lead to a lower fairness factor compared to the use of a flexible DC in both algorithms. The flexible DC mechanism in our proposed solution takes into account the throughput of both WiFi and V2X networks. Consequently, the fairness factor increases and network resources are used more efficiently and fairly.

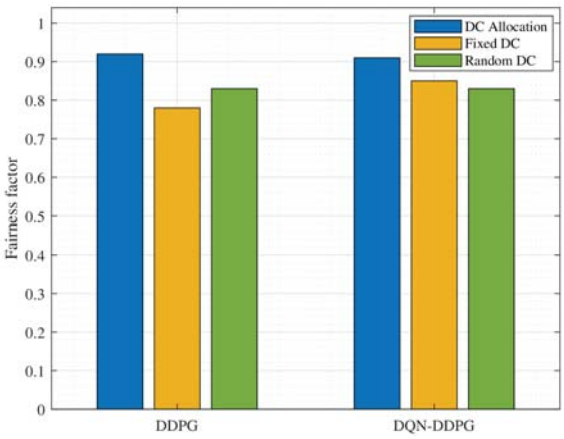

Fig. 6 Impact of duty cycle allocation on fairness factor

V. Conclusion

In this paper, we proposed a DRL based JBSRA scheme for the 5G-V2X network. Our approach addressed the challenge of limited network capacity in 5G by optimizing resource allocation for both NR and NR-U networks. We presented a flexible DC mechanism to mitigate the coexistence problem between NR-U and WiFi users. During the simulation, it was observed that DDPG algorithm demonstrated better convergence performance compared to DQN-DDPG. However, both proposed algorithms exhibited substantial enhancements in the data rates of the V2X network and fairness factor between networks, surpassing the performance of the existing solutions that either do not consider the utilization of unlicensed spectrum or adopt fixed DC.

References

[1] K. S. Kwak, S. Kim, and K. B. Lee, "Coexistence performance of LBT based NR-U with Wi-Fi," in *proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Istanbul , Turkey, September, 2019, pp.1–6.

[2] J. Zhu, Y. Liu, and Y. Jiang, "A novel carrier sensing adaptive transmission mechanism for coexistence between Wi-Fi and NR-U," in *proceedings of IEEE/CIC International Conference on Communications in China (ICCC)*, Changchun , China, 11-13 August, 2019, pp.1–6.

[3] M. Alzenad, S. Ben Jemaa, and S. Alouini, "An overview of coexistence mechanisms for Wi-Fi and NR-U," *IEEE Communications Magazine*, vol. 57, no. 4, pp. 108–114, 2019.

[4] M. Omer, A. Rachedi, and T. Taleb, "Coexistence of Wi-Fi and NR-U: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 22, no. 2, pp. 1036–1070, 2020.

[5] E. Moein, S. K. Taskou, and M. Rasti, "Base Station Assignment Two-tier Dual Connectivity Heterogeneous Networks," in *proceedings of International Symposium on Telecommunications (IST)*, 2018, pp. 474–480.

[6] A. M. El-Hajj and Z. Dawy, "On optimized joint uplink/downlink resource allocation in ofdma networks," in *proceedings of IEEE Symposium on Computers and Communications (ISCC)*, 2011, pp. 248–253.

[7] M. Kim, S. Y. Jung, and S.-L. Kim, "Sum-rate maximizing cell association via dual-connectivity," in *proceedings of International Conference on Computer, Information and Telecommunication Systems (CITS)*, 2015, pp. 1–5.

[8] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.

[9] B. Yin, H. Hu, B. Xi, Q. Liu, Y. Zheng, and Z. Zhang, "Joint Radio Resources Allocation in the Coexisting NR-U and Wi-Fi Networks," in *proceedings of Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, virtual, September, 2021, pp.1532-1538.

[10] L. Wang, M. Zeng, J. Guo, Q. Cui, and Z. Fei, "Joint bandwidth and transmission opportunity allocation for the coexistence between nr-u and wifi systems in the unlicensed band," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2021.

[11] S. Lagen, L. Giupponi, and N. Patriciello, "LBT switching procedures for new radio-based access to unlicensed spectrum," in *proceedings of IEEE Globecom Workshops (GC Wkshps)*, Abu Dhabi, United Arab Emirates, December, 2018, pp.1–6., pp. 1–6.

[12] Su, Yuhan and Du, Xiaojiang and Huang, Lianfen and Gao, Zhibin and Guizani, Mohsen, "LTE-U and Wi-Fi Coexistence Algorithm Based on Q-Learning in Multi-Channel," *IEEE Access*, vol. 6, pp. 13 644–13 652, 2018.

[13] Naik, Gaurang and Jerry Park, Jung-Min, "Impact of Wi-Fi Transmissions on C-V2X Performance," in *proceedings of IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, 2019.

[14] Y. Liu, Y. Jiang, and K. Yang, "Joint cellular-U and WiFi coexistence algorithm based on Q-learning," in *proceedings of IEEE Vehicular Technology Conference (VTC-Fall)*, 2017, pp. 1–5.

[15] E. Rastogi, M. K. Maheshwari, A. Roy, N. Saxena, and D. R. Shin, "Machine Learning-Based DRX Mechanism in NR-Unlicensed," *IEEE Wireless Communications Letters*, vol. 11, no. 5, pp. 1052–1056, 2022.

[16] M. Haghshenas and M. Magarini, "NR-U and Wi-Fi coexistence enhancement exploiting multiple bandwidth parts assignment," in *proceedings of IEEE Annual Consumer Communications Networking Conference (CCNC)*, Las Vegas, NV, USA, January, 2022, pp.1532-1538.

[17] M. Zajac and S. Szott, "Resolving 5G NR-U contention for gap-based channel access in shared sub-7 GHz bands," *IEEE Access*, vol. 10, pp. 4031–4047, 2022.

[18] P. Wang, B. Di, H. Zhang, K. Bian, and L. Song, "Cellular V2X communications in unlicensed spectrum: Harmonious coexistence with vanet in 5G systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 8, pp. 5212–5224, 2018.

[19] 3rd Generation Partnership Project (3GPP), "NR; Physical Channels and Modulation (Release 16)," Technical Specification 38.211, 2020. [Online]. Available: https://www.3gpp.org/ftp//Specs/archive/38$_{s}eries$/38.211/

[20] P. Gawłowicz, A. Zubow, and A. Wolisz, "LtFi: Cross-technology Communication for RRM between LTE-U and IEEE 802.11," *arXiv preprint arXiv:1707.06912*, 2017.

[21] Y. Su, X. Lu, L. Huang, X. Du, and M. Guizani, "TAC-U: A traffic balancing scheme over licensed and unlicensed bands for tactile internet", *Future Generation Computer Systems*, vol. 97, pp. 41–49, 2019.

[22] E. Khorov, A. Kiryanov, A. Lyakhov, and G. Bianchi, "A tutorial on ieee 802.11ax high efficiency wlans," *IEEE Communications Surveys Tutorials,* vol. 21, no. 1, pp. 197–216, 2019.

[23] Gharehgoli, Amir and Nouruzi, Ali and Mokari, Nader and Azmi, Paeiz and Javan, Mohammad Reza and Jorswieck, Eduard A., I-Based Resource Allocation in End-to-End Network Slicing under demand and CSI Uncertainties," *IEEE Transactions on Network and Service Management,* pp. 1–1, 2023.

[24] Parvini, Mohammad and Javan, Mohammad Reza and Mokari, Nader and Abbasi, Bijan and Jorswieck, Eduard A, "AoI-aware resource allocation for platoon-based C-V2X networks via multi-agent multi-task reinforcement learning," *IEEE Transactions on Vehicular Technology*, 2023.

[25] Y. Su, M. LiWang, Z. Gao, L. Huang, S. Liu, and X. Du, "Coexistence of Cellular V2X and Wi-Fi over Unlicensed Spectrum with Reinforcement Learning," in *proceedings of IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.