IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 May 2024, accepted 16 June 2024, date of publication 12 July 2024, date of current version 22 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3427124

RESEARCH ARTICLE

Power Control of 5G-Connected Vehicular Network Using PPO-Based Deep Reinforcement Learning Algorithm

MOSTAFA RAEISI AND ABU B. SESAY, (Life Senior Member, IEEE)
Department of Electrical and Software Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada

Corresponding author: Mostafa Raeisi (mostafa.raeisiziaran@ucalgary.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada; and in part by the Schulich School of Engineering, University of Calgary.

ABSTRACT In this paper, we propose a novel power control in vehicular 5G-connected network using Deep Reinforcement Learning (DRL) algorithm. We investigate power allocation for Connected Autonomous Vehicles (CAVs) on uplink connections in mm-wave bands between the CAVs and Roadside Units (RSUs). Our objective is to achieve the desired uplink transmission capacity using the minimum required power and minimize co-channel interference for neighboring cells. To achieve this goal, we use the Proximal Policy Optimization (PPO) algorithm implemented by modified actor-critic architecture to solve the problem. In the proposed architecture, a Deep Neural Network (DNN) model is used to gain the desired outputs of the problem. The suggested approach is fully compatible with the existing 3GPP-based 5G architecture and uses the available quantized information in cellular users' measurement reports which provides seamless integration within existing RAN architectures. The performance of the proposed algorithm is compared with multiple power control algorithms in various road conditions. Simulation results show that the proposed algorithm outperforms the 3GPP-based power control algorithm in the dynamic road environment.

INDEX TERMS 5G, autonomous vehicle, deep neural network, power allocation, power control, proximal policy optimization, reinforcement learning, vehicular network.

I. INTRODUCTION

The cellular network is continuously evolving to respond to the growing demands of wireless users for more extensive connections with higher quality of service across various fields and applications. Intelligent Transportation Systems (ITS) represent one emerging field that can benefit from cellular connections. Autonomous vehicles leverage diverse types of sensors for environmental perception to navigate on the road in disparate road/weather conditions. By adding the communication capability to autonomous vehicles, Connected Autonomous Vehicles (CAVs) can take advantage of the added information and processing resources for safer driving. CAVs can access edge/cloud processing resources, high-speed Internet connection, and other safety features through cellular connections. Also, CAVs can use the added data through wireless communication to enhance their decision-making confidence, improve drive safety and control, and increase commutation ease and pleasure. Vehicles' fast-moving ability, high data rate requirement for safety and infotainment systems, and their battery-powered energy source call for efficient resource allocation in their cellular connection links [1], [2], [3], [4].

One of the promising technologies to connect CAVs to the Internet, cloud infrastructure, and edge computing resources is the Fifth-Generation (5G) networks. CAVs communicate with the Road Side Units (RSUs) installed alongside the road to connect to the 5G network. Power allocation is

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

a crucial task in wireless communication, especially in Vehicle-to-Infrastructure (V2I) connection between a CAV and a 5G RSU, due to the variety of different types of battery-powered users [5]. Efficient power allocation can significantly improve connection quality, energy efficiency, and spectral efficiency [6]. The high mobility of CAVs and the small diameter of mm-wave cells in 5G make efficient resource allocation more challenging [7]. Intelligent uplink power control can reduce co-channel interference at RSUs, enhance link reliability, guarantee users' quality of service, save on users' energy consumption, improve network capacity, and boost communication performance.

The 5G's uplink power control, based on the 3rd Generation Partnership Project (3GPP) standard, is a sophisticated mechanism. It relies on measurement reports from users that contain various variables measured at the user equipment. This mechanism aims to establish a balance between providing the desired quality of service and minimizing interference for other users. Like previous generations, 5G's power control includes both open-loop and closed-loop control. The closed-loop power control uses techniques like Transmit Power Control (TPC) to dynamically adjust the users' power level at each time step in an iterative manner. Hence, for a new user or a significant change in the channel condition, it might take several time steps for TPC to get to the right power level [8].

The integration of Artificial Intelligence (AI) and Machine Learning (ML) techniques, particularly Reinforcement Learning (RL), has emerged as a promising approach for optimizing resource allocation in 5G vehicular networks. RL algorithms, such as Deep Q-Learning and Proximal Policy Optimization, offer the capability to adaptively adjust power control strategies based on real-time feedback from the network environment. By leveraging historical data and learning from interactions with the network, RL-based power control methods can effectively optimize system performance while mitigating interference and improving overall network efficiency. This paradigm shift towards AI and ML-powered solutions underscores the importance of exploring novel techniques to meet the evolving demands of 5G vehicular networks.

Many researchers are working on power control algorithms for 5G networks with different objectives. Most of the recent publications are focused on downlink channels, Vehicle-to-Vehicle (V2V) and Device-to-Device (D2D) links, Non-Orthogonal Multiple Access (NOMA) systems, and Unmanned Aerial Vehicle (UAV) aided networks [9], [10], [11], [12]. Resource allocation methods in uplink connection can be divided into two general groups of centralized and distributed methods. In distributed systems, each user makes a decision for their own power based on its algorithm and assumptions on other users' decisions [12], [13]. The centralized power control algorithms are mainly gNodeB-based, where a gNodeB (5G's base-station - Next Generation Node B) collects the channel conditions from all users, makes decisions for all users, and sends the power commands to users [14], [15]. In this work, we adhere to the existing centralized uplink power control structure of the 5G network and utilize a centralized method. This approach offers

TABLE 1. Alphabetized abbreviations Table.

Abbr.	Definition
3GPP	3rd Generation Partnership Project
5G	Fifth-Generation wireless network
BPRE	Bits Per Resource Element
BS	Base Station
BWP	Bandwidth Part
CAV	Connected Autonomous Vehicle
CID	Cell Identification
CSI	Channel State Information
CU	Central Unit
D2D	Device-to-Device
DCI	Downlink Control Information
DDPG	Deep Deterministic Policy Gradient
DNN	Deep Neural Network
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
DU	Distributed Unit
EIRP	Effective Isotropic Radiated Power
FR2	Frequency Range 2
FSPL	Free Space Path Loss
IFR	Interference
ITS	Intelligent Transportation Systems
LoS	Line-of-Sight
MEC	Mobile Edge Computing
mm-wave	Millimeter-wave
NOMA	Non-Orthogonal Multiple Access
PDCCH	Physical Downlink Control Channel
PL	Path-Loss
PPO	Proximal Policy Optimization
PRACH	Physical Random Access Channel
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
QoS	Quality of Service
RL	Reinforcement Learning
RRC	Radio Resource Control
RS	Reference Signal
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSSI	Received Signal Strength Indicator
RSU	Roadside Unit
SCS	Sub-Carrier Spacing
SDN	Software-Defined Network
SINR	Signal to Interference plus Noise Ratio
SNR	Signal to Noise Ratio
SRS	Sounding Reference Signal
TPC	Transmit Power Control
TS	Technical Specification
UAV	Unmanned Aerial Vehicle
UE	User Equipment
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
WLAN	Wireless Local Area Network

the advantage of effective interference management and a comprehensive understanding of neighboring user behavior, leveraging the instantaneous network limits inherent to centralized methodologies. To our knowledge, our study represents the first application of the Proximal Policy Optimization (PPO) algorithm for centralized power control in the uplink channel of 5G's V2I connections. Our emphasis lies in integrating quantized user-reported values of link conditions in accordance with 3GPP standards, setting our work apart from existing DRL-based approaches.

In this paper we propose a novel power control algorithm for vehicular users in 5G network at millimeter-wave (mm-wave) bands, leveraging a Deep Reinforcement Learning (DRL) algorithm. Our objective is to achieve the required capacity by minimizing power consumption at the user end, using only quantized reported values available at the gNodeB. The PPO algorithm is used to train a Deep Neural Network and the trained model is used in each gNodeB for power control for all users under its coverage. The proposed scheme relies solely on the available information at the gNodeB by users' reports, and it is compatible with the current structure of the 5G network standardized by 3GPP. Similar to the conventional 5G network, the uplink power level and resource allocation decisions are made at the gNodeB and then sent to users over downlink control channels. The key contributions of this paper are as follows:

- A new power control algorithm for uplink channels in a 5G vehicular network with small cells over mm-wave bands. The objective is to minimize the achieved capacity error from the desired capacity while satisfying the required Quality of Service (QoS). This is achieved by considering the individual user power level and minimizing user power consumption.
- A DNN with a novel discrete state space and action space in the Radio Resource Control (RRC) function of a gNodeB's central units. This model makes centralized decisions for each gNodeB's users and improves the downlink power control command data traffic.
- A PPO algorithm with a novel actor-critic architecture to train the DNN model. The proposed structure can adapt the model to the vehicular environment changes with a controlling parameter to prevent divergence, eliminating the need for a massive storage space to keep replay buffer data.
- A Design of an innovative reward function to effectively assist the PPO algorithm in satisfying the optimization function's requirements. The proposed reward function can be easily adapted to include any number of parameters effective in power control.
- An extensive numerical simulation to compare the performance of the proposed model with multiple power control models. The simulation results show that the proposed system outperforms other models and satisfies the desired objective.

The rest of this paper is organized as follows. The background of the conventional power control, DRL, and PPO algorithms is briefly investigated in Section II. The related research works in power control for 5G's uplink channel are reviewed in Section III. In Section IV, we present our system model, problem formulation, and assumptions. The PPO-aided DRL-based power control model is presented in Section V. The simulation results and comparison with the reference algorithms are studied in Section VI. Finally, the conclusion is provided in Section VII. To simplify cross-referencing, Table 1 presents a compilation of abbreviations utilized throughout this paper.

II. BACKGROUND

In this section, we offer an overview of prevalent power allocation algorithms in wireless communication, shedding light on their applicability to uplink power control and our methodology for comparative analysis. Our examination highlights why certain algorithms are unsuitable for our purposes and why direct comparisons may not be feasible. Additionally, we introduce a benchmark algorithm against which we gauge the performance of our system.

Moving forward, we delve into the power control framework within 5G networks, drawing from 3GPP standard documents as our primary reference for simulations. This framework serves as the cornerstone for evaluating the performance of our proposed algorithm, aligning with our primary objective of enhancing 3GPP-based power control (1), as shown at the bottom of the next page.

A. GENERIC POWER ALLOCATION ALGORITHMS

There are many power allocation algorithms discussed in the literature, including equal power allocation, max-min fairness, channel inversion, water-filling, proportional fairness, and many more. In this section, we briefly review the most commonly studied power allocation algorithms and go through some of their advantages and disadvantages.

- **Equal Power Allocation**: In the equal power allocation algorithm, an identical power level is allocated to all users. It is an easy algorithm to implement and has low runtime complexity, but it could not insure fairness among users. This is a suitable solution for an environment where all users experience fairly similar channels and have equal data rate requirements or low implementation complexity is important [16]. This strategy can be used in both uplink and downlink communication scenarios.
- **Max-Min Fairness**: This algorithm aims to maximize the minimum data rate among users through an iterative resource allocation process. It serves as a viable solution for environments where a baseline data rate must be guaranteed for all users while striving for data-rate fairness. However, one drawback of this algorithm is its potential to result in a lower overall network data rate. This is because more power may need to be assigned to a link with a poor channel condition, leading to unbalanced power allocation and increased interference for other links [17], [18]. Hence, the objective of

this algorithm does not align with the purpose of the proposed algorithm, rendering any comparison of their results unproductive and insignificant.

- **Channel Inversion Power Allocation**: In this algorithm, the assigned power to users is relevant to the inversion of their channel gain to keep a constant value for Signal to Noise Ratio (SNR). The advantage of this algorithm is a constant data rate at each channel according to the determined value as long as the assigned power does not exceed the maximum power limit. The disadvantages of this algorithm include the need for channel state information at the transmitter, and chances of poor power efficiency [19]. Since this algorithm relies on the channel gain, and given that the channel gain is unavailable according to the system model under consideration, it's not feasible to compare its performance with this algorithm.
- **Water-Filling Algorithm**: The Water-Filling Algorithm aims to maximize a network's total capacity by allocating power to users based on their channel conditions, considering both channel gain and noise level for power allocation. This algorithm assigns more power to a link with better channel quality to increase the total data rate, while it may allocate no power to very weak channels. The original Water-Filling Algorithm is solely based on SNR and does not consider interference. Additionally, it is designed for centralized transmission from one source to multiple receivers, making it unsuitable for an uplink scenario to control the power level of multiple distributed users with a single channel per user [18], [19]. Hence, it cannot be utilized for comparing the performance of the proposed algorithm.
- **Proportional fairness Algorithm**: This algorithm aims to provide resource allocation fairness with the goal of maximizing the system utility function. In this algorithm, each user gets a share of each resource with three conditions: 1) the share of each user from each resource is non-negative; 2) the sum of allocated shares of one resource to all users is equal or less than the capacity of that resource; and 3) the current allocated share of resources is the maximum possible share as provided by [20]. Proportional fairness can be computationally intensive, and there is a trade-off between fairness and total throughput in the system. The algorithm's formulation does not support its use in distributed transmitter applications [18], [20]. So,

$$
\begin{aligned}
P_{\text{PUSCH},b,f,c}(i,j,q_d,l) &= \min\left\{\begin{array}{c} P_{\text{CMAX},f,c}(i),\, P_{\text{O\_PUSCH},b,f,c}(j) + 10log_{10}\left(2^{\mu}.M^{\text{PUSCH}}_{RB,b,f,c}(i)\right) + \\ \alpha_{b,f,c}(j).PL_{b,f,c}(q_d) + \Delta_{TF,b,f,c}(i) + f_{b,f,c}(i,l)\end{array}\right\} \\
P_{\text{O\_PUSCH},b,f,c}(j) &= P_{O\_\text{NOMINAL,PUSCH},f,c}(j) + P_{\text{O\_UE\_PUSCH},b,f,c}(j) \\
P_{O\_\text{NOMINAL,PUSCH},f,c}(j) &= \begin{cases} P_{\text{O\_PRE}} + \Delta_{\text{PREAMBLE,Msg3}} & \text{if Type-1 RA, j=0, } \textit{P0-PUSCH-AlphaSet} \text{ not provided} \\ P_{\text{O\_PRE}} + \Delta_{\text{MsgA\_PUSCH}} & \text{if Type-2 RA, j=0, } \textit{P0-PUSCH-AlphaSet} \text{ not provided} \\ \textit{P0-NominalWithoutGrant} & \text{if j=1, } \textit{P0-NominalWithoutGrant} \text{ is provided} \\ P_{O\_\text{NOMINAL,PUSCH},f,c}(0) & \text{if j=1, } \textit{P0-NominalWithoutGrant} \text{ is not provided} \\ \textit{P0-NominalWithGrant} & \text{if } j \in \{2,\ldots,J-1\}, \textit{P0-NominalWithGrant} \text{ is provided} \\ P_{O\_\text{NOMINAL,PUSCH},f,c}(0) & \text{if } j \in \{2,\ldots,J-1\}, \textit{P0-NominalWithGrant} \text{ is not provided} \end{cases} \\
\Delta_{TF,b,f,c}(i) &= \begin{cases} 10\log_{10}\left(\left(2^{\text{BPRE}.K_s}-1\right).\beta^{\text{PUSCH}}_{\text{offset}}\right) & \text{if } K_s = 1.25 \\ 0 & \text{if } K_s = 0 \\ 0 & \text{if PUSCH transmission is over one layer} \end{cases} \\
\text{BPRE} &= \begin{cases} \sum_{r=0}^{C-1}\frac{K_r}{N_{RE}} & \text{if PUSCH transmission with UL-SCH} \\ Q_m.\frac{R}{\beta^{\text{PUSCH}}_{\text{offset}}} & \text{if CSI transmission in a PUSCH without UL-SCH data} \end{cases} \\
f_{b,f,c}(i,l) &= \begin{cases} f_{b,f,c}(i-i_0,l) + \sum_{m=0}^{\mathcal{C}(D_i)-1}\delta_{\text{PUSCH},b,f,c}(m,l) & \text{if } \textit{tpc-Accumulation} \text{ is not provided} \\ \delta_{\text{PUSCH},b,f,c}(i,l) & \text{if } \textit{tpc-Accumulation} \text{ is provided} \\ \Delta P_{\text{rampup},b,f,c} + \delta_{\text{msg2},b,f,c} & \text{if } i=l=0, \text{ and UE receives a random access response message} \\ \Delta P_{\text{rampup},b,f,c} & \text{if } i=l=0, \text{ and UE transmits the PUSCH} \end{cases}
\end{aligned}
\tag{1}
$$

it cannot be used as a benchmark to compare results with the proposed algorithm.

In summary, the discussion of various generic power allocation algorithms indicates that only Equal Power Allocation can be feasibly compared with the proposed algorithm. Practical limitations, such as unavailable channel gain information or unsuitability for distributed transmitter applications, preclude the comparison with other algorithms.

B. 3GPP-BASED POWER CONTROL ALGORITHM

This section delineates the 5G power control equation, provided in (1), derived from 3GPP standard documents. Subsequently, in the Simulation Results Section (Section VI), these equations are employed to simulate the 3GPP-based power control algorithm, enabling a comparative analysis of its performance against the proposed algorithm.

The power control algorithm in uplink transmission of 3GPP-based 5G sets the power level for Physical Uplink Shared Channel (PUSCH), Physical Uplink Control Channel (PUCCH), Sounding Reference Signal (SRS), and Physical Random Access Channel (PRACH) channels [8]. At each time-step, each active User Equipment (UE) calculates its own uplink transmission using the controlling parameters provided by the 5G network (gNodeB). The radio resource management function resides in the central unit of gNodeBs, and one of its responsibilities is to prepare all of these power control parameters and send them to UEs through the RRC messages. The main source of these power control parameters is the UE's serving gNodeB, but in some cases such as a handover process, some of the parameters might come from the other components such as from the 5G core network. The gNodeB sets these parameters based on the gathered measurement reports from the UEs under its coverage and makes decisions for all of them.

In this section, we delve into the 3GPP-based power control mechanism tailored for 5G networks. This fundamental equation serves as the cornerstone for our subsequent simulations, providing a benchmark for evaluating our experimental outcomes against established standards. According to 3GPP's Technical Specification (TS) in [8], a UE determines its power level on the PUSCH channel at time-step i on uplink Bandwidth Part (BWP) b of carrier f of serving cell c using (1). In Equation (1), the first line represents the primary equation governing power calculation, followed by subsequent lines detailing the definition of each variable therein. In this equation, j is the parameter set configuration index, q_d is the Reference Signal (RS) index for the active downlink BWP, and l is the PUSCH power control adjustment state index. The $P_{\mathrm{CMAX},f,c}(i)$ is the maximum output power configured for UE at carrier f of serving cell c at transmission occasion i which is further explained in clause 6.2.4 of [21]. The variable μ is the Sub-Carrier Spacing (SCS) as described on clause 4 of [22], and $M^{\mathrm{PUSCH}}_{RB,b,f,c}(i)$ is the bandwidth of the PUSCH resource in terms of number of resource blocks. $\mathrm{PL}_{b,f,c}(q_d)$ is the UE's estimation of downlink pathloss using RS. *BPRE* stands for Bits Per Resource Element, N_{RE} is a number of resource elements, $\delta_{\mathrm{PUSCH},b,f,c}(i,l)$ is a TPC command value included in a Downlink Control Information (DCI) format that schedules the PUSCH.

The value for each of these parameters and the other states and variables, that have not been described here as there is no title or description in [8], get their values from RRC messages. Depending on the conditions, each of them obtains its value from a different parameter. This implies that, in most cases, there are multiple conditions for each of the aforementioned states and parameters. As can be seen, the uplink power control equation in 5G is very complicated and depends on many parameters, as well as the algorithm to select the value of each of these parameters.

According to Table 6.2.1.0-1 of [21], seven UE power classes are defined for the 5G network, where class 2 is considered for vehicular UEs. The values of the minimum and maximum output power for different operating bands of class 2 users are presented in clauses 6.3.1.2 and 6.2.1.2, respectively. The Effective Isotropic Radiated Power (EIRP) has a maximum output power value of 43 dBm across all operating bands, representing both the emitted power and transmitter antenna gain.

The uplink messages on PUCCH carry the UE's measurement reports, including Signal to Interference plus Noise Ratio (SINR), Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), and Received Signal Strength Indicator (RSSI) [23]. The downlink messages on the Physical Downlink Control Channel (PDCCH) transfer the RRC messages, such as power control commands and other controlling data, to the users. To minimize the number of bits needed for transmitting these values over the air interfaces, the 5G network transfers quantized mapped values. The SINR report mapping equation, as presented in clause 10.1.16 of [24], is as follows.

$$\text{SINR Reported value} = \lfloor \text{SINR} * 2 + 47 \rfloor \tag{2}$$

where SINR is measured in dB, and the reported SINR value is limited between 0 (equivalent to values less than -23 dB) and 127 (for SINR values equal to or larger than 40 dB).

III. RELATED WORKS

In this section, we briefly review recent research on power control algorithms for 5G networks. Our main focus is on RL-based uplink power control algorithms designed for vehicular users in 5G networks with small cells, utilizing mm-wave frequency bands.

The power control algorithm has been investigated for various links involving autonomous vehicles, including links between vehicles and UAVs, as well as between vehicles themselves. In [27], researchers aim to optimize power allocation with the objective of maximizing the total data rate for each UAV in the downlink channel of a UAV-assisted vehicular network. In [28], the problem of joint power and bandwidth allocation for V2V links is investigated, employing the PPO algorithm for resource allocation, with

each V2V link or vehicle acting as an agent. Zhang et al. in [29], studied relay selection and power allocation for sub-6 GHz in multi-hop vehicular networks using a centralized hierarchical DRL in an RSU. The RSU provides support for sub-6 GHz (for broad coverage) and mm-wave (for high-bandwidth short-range communication). Two Deep Q-Network (DQN) models are employed in the RSU: one for relay selection and the other for power allocation in D2D links. In [40], authors tried to achieve the maximum weighted sum rate by proper power allocation in D2D transmission using graph neural network.

The power control algorithm for V2I links can be categorized into two main groups: downlink and uplink power controls. There is a key difference between these two types of power control algorithms. Downlink transmission originates from a centralized location, typically a gNodeB, and is directed towards multiple distributed users. Consequently, a total power limit must be equitably distributed among different links, with all processing occurring in a single central unit. Conversely, uplink transmission emanates from multiple distributed sources to a central receiving unit (gNodeB). Thus, each transmitter (or link) operates within its own transmission power limit. Furthermore, the uplink power control algorithm can be either centralized or distributed, each with its own set of advantages and disadvantages, as discussed below.

Given the predominant focus of power allocation research on downlink connections within the realm of 5G networks, we provide a brief overview of some of the most pertinent publications in downlink power allocation. The downlink power allocation algorithm is investigated in [11], [12], [30], [31], and [32]. In [30], the authors explore downlink power allocation and user association to base stations (BSs) in ultra-dense small-cell mm-wave 5G BSs using a Q-learning algorithm. Their proposed algorithm aims to maximize the sum capacity and ensure the required QoS for users by appropriately distributing power between BSs. Furthermore, the problem of power allocation and resource management in 5G's mm-wave small cells is studied in [31], where the authors propose a sub-optimal solution for maximizing the sum rate on the downlink channel. Spectrum and power allocation in ultra-dense networks are addressed in [32] with the goal of achieving a trade-off between spectral efficiency, energy efficiency, and throughput fairness using DRL. In [11], the problem of joint power control and channel allocation in downlink transmission in Wireless Local Area Networks (WLANs) with multiple access points is solved using the Q-learning algorithm. Authors in [12] solved the power allocation problem jointly with handover management with a multi-agent PPO with centralized training to maximize the throughput and reduce the handover frequency. They considered a macro BS with multiple mm-wave small BSs in downlink communication with multiple agents and centralized training at the macro BS. The power control in a vehicular network for video delivery and caching service, involving a macro BS and multiple mm-wave BSs,

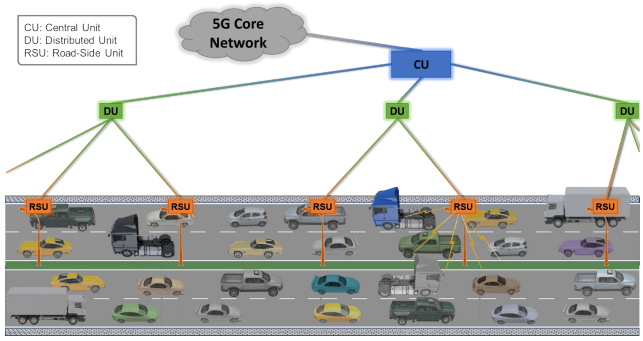

FIGURE 1. Considered 5G-connected V2I link structure for vehicular users.

is examined in [33]. The authors employed Deep Deterministic Policy Gradient (DDPG) for optimal power allocation alongside cache allocation in downlink transmission.

As mentioned above, the uplink power control for cellular users in the V2I link can be divided into two general groups: centralized and distributed management [16]. The centralized power control algorithms mainly reside in the gNodeB, where the gNodeB gathers information from the environment, makes decisions for all users, and then informs the users about their new power levels. In this method, interference between different users can be managed and reduced through proper power allocation. However, a disadvantage of this method is the extensive information exchange between users and the network, as all users need to report their measurements to the network for decision-making. This approach has been adopted for the 3GPP-based 5G network.

On the other hand, in distributed power control mechanisms, each user uses an algorithm and its locally available information to make decisions for itself. In some approaches, such as neural network-based algorithms, each user trains its algorithm with its observations. Periodically, all users send their local algorithms to a central location to merge all locally trained algorithms. Finally, the central trainer broadcasts the merged algorithm to the users. This approach reduces the transmission of measurement data over the air interfaces, but adds the periodic model transfer load, which might be significant in the case of DNN. A disadvantage of the distributed approach is its poor interference management, as each user lacks information about other users' decisions and conditions.

In [34], the problem of uplink power allocation in joint resource allocation for vehicular networks with Software-Defined Network (SDN)-assisted Mobile Edge Computing (MEC) architecture is addressed using a stateless Q-learning algorithm. The authors attempt to minimize computational overhead through transmission power control, sub-channel allocation, and optimizing offloading strategy. In their power control model, each user acts as an agent and makes decisions locally. The model of distributed training and centralized aggregation, as investigated in [35], focuses on a macro cell and multiple small cells where users train their DQN model and send it to the base station for federated learning aggregation. This approach pursues two objectives:

throughput maximization and total power consumption minimization while ensuring the required QoS. Additionally, [36] addresses the maximization of capacity for the vehicles' uplink channel, considering the reliability and latency of V2V links. The joint uplink power allocation and beamforming problem for high-speed railways in mm-wave-equipped base stations is explored in [7]. The authors decompose the problem into two separate parts and address them individually. They utilize a multi-agent DDPG algorithm with an actor-critic architecture for power allocation, aiming to maximize the achievable sum rate. To the best of our knowledge, there is no existing literature investigating centralized power control for 5G-based vehicular networks with mm-wave small cells in the uplink channel using the PPO algorithm.

IV. SYSTEM MODEL

In this section, we present the system model discussed in this paper, and we formulate the problem we aim to address. We envision a section of a road with multiple lanes in each direction covered by the 3GPP-based 5G network, as shown in Figure 1. Our focus is on vehicular users of the 5G network, while pedestrians and other cellular users utilize separate channel sets.

The 5G coverage for vehicular users is provided by small cells known as RSUs, which are installed on the median of the road. Each RSU covers all lanes in each direction and its coverage extends a few hundred meters alongside the road. RSUs offer full coverage for road users, with this paper focusing solely on the connection between RSUs and CAVs, namely the V2I link. The RSUs operate on mm-wave bands, specifically the n257 operating band within the FR2-1 frequency range, which ranges between 26,500 MHz and 29,500 MHz [21]. In this paper, we adopt the 3GPP's gNodeB split model, which includes a single Central Unit (CU or gNB-CU) and multiple Distributed Units (DU or gNB-DU). Each DU comprises multiple cells, and associated network functions are embedded in units along the road. Our focus is specifically on the small cells of the DU, also referred to as RSUs in this context, which play a crucial role in facilitating communication within the vehicular environment. It's worth noting that each gNodeB may or may not include both small cells and large cells.

The communication link between the RSU and the CAV is a direct Line-of-Sight (LoS) connection. The mobility management model discussed in this paper adheres to the scheme proposed in [23].

Our target is to optimize the uplink power control for vehicular users in the 5G network, maintaining compatibility with the 3GPP-defined network structure and signaling, while proposing minimal changes to the network. We aim to achieve the desired data rate for each user by controlling its power level at each time-step, and guaranteeing the minimum required QoS. Based on the described goal, our optimization problem for power control at each gNodeB is defined as follows. To our knowledge, this optimization problem for uplink power control appears to be novel, as it has not been previously addressed in existing literature.

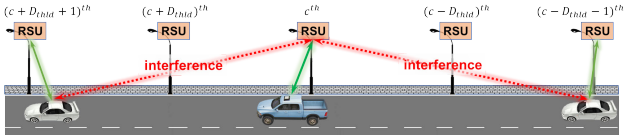

FIGURE 2. Co-channel interferences for the uplink channel in the middle RSU with D_{thld} of one.

$$\text{Minimize} \quad \sum_{u=1}^{U} \left| R^{u}_{Achieved}(i) - R^{u}_{Objective}(i) \right| \tag{3}$$

$$\text{s.t. } P^{u}_{\text{PUSCH},b,f,c}(i) \leq P_{\text{CMAX},f,c}(i), \ \forall u, i, \tag{4}$$

$$P^{u}_{\text{PUSCH},b,f,c}(i) \geq P_{\text{CMIN},f,c}(i), \ \forall u, i, \tag{5}$$

$$\gamma^{u}_{\text{PUSCH},b,f,c}(i) \geq \gamma_{CMIN}(i), \ \forall u, i. \tag{6}$$

The term $R^{u}_{Achieved}(i)$ represents the achieved data rate of the u-th user at time-step i, as defined in (7). $R^{u}_{Objective}(i)$ denotes the desired data rate for the u-th user at time-step i, determined by the network based on user requirements. $P^{u}_{\text{PUSCH},b,f,c}(i)$ indicates the power level of the u-th user at time-step i on the uplink PUSCH over BWP b of carrier f of serving cell c. $P_{\text{CMAX},f,c}(i)$ and $P_{\text{CMIN},f,c}(i)$ represent the maximum and minimum permitted uplink power, respectively, at carrier f of serving cell c at time i, as mentioned in the Background section. The terms $\gamma^{u}_{\text{PUSCH},b,f,c}(i)$ and $\gamma_{\text{CMIN},b,f,c}(i)$ denote the achieved uplink transmission SINR of the u-th user at its serving RSU and the minimum required SINR, respectively, at time i for PUSCH transmission over BWP b of carrier f of serving cell c.

The u-th user's achieved data rate at time-step i is calculated as follows.

$$R^{u}_{Achieved}(i) = B^{u}_{\text{PUSCH}}(i).\log_2\left(1+\gamma^{u}_{\text{PUSCH},b,f,c}(i)\right) \tag{7}$$

where $B^{u}_{\text{PUSCH}}(i)$ represents the uplink bandwidth of the u-th user on PUSCH transmission at time step i. The $\gamma^{u}_{\text{PUSCH},b,f,c}(i)$, as defined above, denotes the SINR at the gNodeB and is calculated using (8).

$$\gamma^{u}_{\text{PUSCH},b,f,c}(i) = \frac{P^{u}_{r,\text{PUSCH},b,f,c}(i)}{\sum_{u' \in S_{\text{IFR}}} P^{u'}_{r,\text{PUSCH},b,f,c'}(i) + n_0} \tag{8}$$

In (8), $P^{u}_{r,\text{PUSCH},b,f,c}(i)$ represents the received power at RSU c from user u via the PUSCH link, using BWP b, carrier f, at time-step i. $P^{u'}_{r,\text{PUSCH},b,f,c'}(i)$ denotes the received power at RSU c from interfering user u' located on RSU c', utilizing the same channel as user u for communication with its serving cell at time-step i. Additionally, n_0 signifies the noise power and S_{IFR} is the set of all interfering users for user u.

According to the channel allocation scheme proposed in [23], there should be a distance of D_{thld} between two RSUs utilizing a channel for two users, as depicted in Figure 2. Consequently, for channel assignment at the c-th RSU, the gNB-CU considers all channels in use at the c-th RSU and D_{thld} RSUs on each side of the c-th RSU.

TABLE 2. List of major notations.

Variable	Description
$a^u(i)$	Action (power level) for user u at time-step i
$A(i)$	Set of actions for all users at time-step i
b	Bandwidth part index
$B^u_{\dots}$	Bandwidth of the u-th user
c	RSU index
c^u	Cell ID of the u-th user
d_c^{RSU}	Coverage diameter of the c-th serving RSU
$EIRP^u$	Equivalent isotropic radiated power of the u-th user
f	Carrier index
g	Hyperparameter of the reward function
i	Time-step indicator
IFR	Interferer
$P^u_{\dots}$	Transmission power of the u-th user
$P^u_{r,\dots}$	Received power from the u-th user
PL^u_{RS,q_d}	Pathloss of the u-th user on reference signal by index q_d
$R^u_{Achieved}$	Achieved data rate of the u-th user
$R^u_{Objective}$	Objective (desired) data rate for the u-th user
r_{max}	Maximum possible reward value at each time-step for each user
r_{min}	Minimum possible reward value at each time-step for each user
Rwd^u	Achieved reward value by user u
$s^u(i)$	state vector of the u-th user at time-step i
$S(i)$	set of all state vectors at time-step i
u	Indicator for the specific u-th user
U	Total number of users
$\mathbf{X}^u_{\dots}$	Vector of considered QoS parameters ($M \times 1$)
$\gamma^u_{\dots}$	SINR of the u-th user
$\gamma^{opt}_{\dots}$	Optimal SINR value

V. PPO-DNN-BASED ADAPTIVE POWER CONTROL

This section presents our solution to the optimization problem defined in Section IV. We introduce a novel adaptive power control model for 3GPP-based 5G networks, leveraging a DNN trained using the PPO algorithm. The model is based on a centralized power control algorithm in gNodeBs for uplink transmissions, akin to traditional power control in 5G, where decisions are made at the gNodeB and transmitted to the users. Inputs to our model consist of quantized mapped data measured and transmitted by users to the serving gNodeB, utilizing the same resolution as defined for the 5G gNodeB's power control mechanism. The output power levels of our model are quantized mapped power values based on the 3GPP mapping formula, designed for efficient transmission over the air. This proposed algorithm demonstrates the capability to converge to the desired data rate in fewer number of time-steps compared to conventional 5G iterative power control algorithms.

In our system, each gNodeB is an agent, making decisions for all users connected and served by cells within that gNodeB. Each power control agent is equipped with a DNN and a PPO algorithm to gradually train and adapt the DNN to environmental changes. The PPO algorithm has the ability to train the DNN with only a few observations (mini-batch) at each time-step, which eliminates the need for a large storage space for a replay buffer, as required in some other RL algorithms like Deep Q-Learning algorithm. The environment includes the road, the RSUs, and all cellular users communicating with their RSUs to measure their signal quality and report it to the network.

A modified actor-critic architecture is used in our PPO algorithm, wherein the actor constitutes the DNN model utilized for power level predictions, while the critic is substituted with a simple fixed value generator to suit the specific nature of our problem. In this scenario, the reinforcement learning problem is a single-episode task, where each decision for every user is considered independent of previous decisions. This independence stems from the fact that the power of interfering users can be changed at each time-step, consequently affecting the objective user's SINR. Thus, a power level deemed suitable in the past might not suffice in the present moment. Consequently, decisions at each time-step are made independently of prior instances, with a single-episodic task deemed the optimal approach to achieve the best results in an interference-laden environment. Since the critic's role involves estimating the expected value at each state, in our single-episodic task, the maximum expected value at each state corresponds to the maximum reward per action. This obviates the necessity for a critic DNN model and its associated training, simplifying the network architecture while enhancing stability and processing efficiency.

It is essential to highlight that during the exploration of solutions for the optimization problem outlined in (3) to (6), we experimented with a wide range of inputs and various reinforcement learning algorithms such as Q-Learning, DQN, and Double DQN (DDQN). However, due to the intricate nature of the vehicular environment coupled with the unpredictable characteristics of wireless links,characterized by small-scale fading and interference, these algorithms failed to achieve satisfactory performance. Consequently, we opted for the PPO algorithm, and experimentally increased the number of inputs (states) to enhance the efficacy of the power control mechanism. Additionally, it's worth noting that the parameter ϵ in PPO algorithm plays a crucial role in ensuring the stability and convergence of the training process in PPO algorithms [26]. By constraining the ratio of the old and new policy probabilities within a certain range, ϵ prevents excessive updates and divergence during training, facilitating more stable and efficient convergence.

Now, let's proceed with the description of the states, actions, and reward function of our system.

States: The PPO algorithm is capable of handling both continuous and discrete input values. Since some of the power control algorithm's input data are measured and sent by users over the air interfaces in discrete format, all inputs to our algorithm are considered as discrete values. Each element of

the input state vector represents the condition vector of one user, which will initiate the generation of a power value for that user in the next time-step. The set of all state vectors at time-step i is defined in (9), which includes the state vectors of all U users in a gNodeB.

$$S(i) = \left\{ s^1(i), s^2(i), \ldots, s^u(i), \ldots, s^U(i) \right\} \tag{9}$$

In (9), $s^u(i)$ represents the state vector of user u at time-step i, while in total there are U users in this serving gNodeB. Each user's state includes its current transmission information, the interfering users' information, the serving RSU's data, and the requirements for the desired rate and QoS. To satisfy the QoS, multiple parameters such as SINR, RSRP, RSSI, latency, reliability, jitter, handover success rate, packet loss rate, and some other parameters need to be met. In this paper, we focus on SINR and satisfy the required data rate with the assumption of equal bandwidth for all users. However, the concept can be easily extended to include more parameters. Hence, the state of user u at time-step i is defined as a vector presented in (10).

$$s^u(i) = \left\{ \begin{array}{c} c^u(i), ch^u(i), \text{EIRP}^u(i), \gamma^u_{\text{PUSCH},b,f,c}(i), \\ \text{PL}^u_{\text{RS},q_d}(i), \text{EIRP}^{\text{IFR1}}(i), \text{EIRP}^{\text{IFR2}}(i), \\ \text{PL}^{\text{IFR1}}_{\text{RS},q_d}(i), \text{PL}^{\text{IFR2}}_{\text{RS},q_d}(i), \gamma^{opt}_{\text{PUSCH},b,f,c}(i), d^{\text{RSU}}_c \end{array} \right\} \tag{10}$$

In (10), c^u represents the serving Cell Identification (CID) number. In real-world scenarios with numerous heterogeneous RSUs per gNodeB, c^u may denote each unique type of RSU. The channel number for the u-th user is denoted by ch^u, while its current EIRP for PUSCH transmission is represented by $\text{EIRP}^u(i)$. The SINR received at RSU c for user u during PUSCH transmission over BWP b on carrier f is denoted by $\gamma^u_{\text{PUSCH},b,f,c}(i)$. The reported Path-Loss (PL) by user u at time-step i on downlink RS by index q_d is given by $\text{PL}^u_{\text{RS},q_d}(i)$. This PL value provide insights into the user's channel conditions, potentially impacted by weather, obstacles, or proximity to the RSU's coverage edge. Similarly, the downlink PL of the two major interferers in vicinity that are using the same channel on other cells (co-channel interference), are given by $\text{PL}^{\text{IFR1}}_{\text{RS},q_d}(i)$ and $\text{PL}^{\text{IFR2}}_{\text{RS},q_d}(i)$. Additionally, the EIRPs of the two primary interfering users nearby are denoted as $\text{EIRP}^{\text{IFR1}}(i)$ and $\text{EIRP}^{\text{IFR2}}(i)$, respectively. The desired SINR level at RSU c for successful uplink transmission decoding on PUSCH over BWP b of carrier f is quantified as $\gamma^{opt}_{\text{PUSCH},b,f,c}(i)$, which is a mapped value derived from (2). The coverage diameter of the c-th serving RSU (representing the coverage length of the road) is communicated to the DNN as d^{RSU}_c, particularly valuable for heterogeneous networks with various RSU types and coverage patterns.

Actions: The gNodeB, which hosts the power control agent for all users within its coverage, takes individual actions for users upon receiving the state vector at time-step i. The action vector for all users at time-step $i+1$ is as follows.

$$A(i+1) = \left\{ a^1(i+1), a^2(i+1), \ldots, a^u(i+1), \ldots, a^U(i+1) \right\} \tag{11}$$

In this context, the action for user u refers to the EIRP values at time-step $i+1$, denoted as $\text{EIRP}^u(i+1)$.

Rewards: One critical aspect of the reinforcement learning system is the design of the reward function. The reward function should provide the RL algorithm with enough information to guide it toward the desired output [37]. In our system, the objective is to achieve as close to the desired data rate as possible while satisfying the demanded QoS at each time-step through proper adjustment of the uplink transmission power at the user, at the same time minimizing interference to users in other RSUs. The proposed reward function to achieve the desired QoS is provided in (12).

$$\text{Rwd}^u(i) = (\text{r}_{\max} - \text{r}_{\min}) \exp^{-\frac{X^u(i)}{2g}} + \text{r}_{\min} \tag{12}$$

In (12), $\text{r}_{\min}$ and $\text{r}_{\max}$ are the minimum and maximum limits of the reward function values. Hyperparameter g controls the expansion or contraction of the multidimensional bell-shaped reward function and the equation for $X^u(i)$ is expressed in (13). The determination of this hyperparameter typically relies on assessing the disparity between the minimum acceptable SINR and the desired SINR mappings. Nevertheless, our simulations across various values of this hyperparameter have revealed minimal impact on the ultimate outcome of the achieved average data rate, primarily due to the algorithm's focus on maximizing reward points.

$$X^u(i) = \left(\mathbf{X}^u_{\text{PUSCH},b,f,c}(i) - \mathbf{X}_{\text{opt},b,f,c}(i) \right)^T \times \left(\mathbf{X}^u_{\text{PUSCH},b,f,c}(i) - \mathbf{X}_{\text{opt},b,f,c}(i) \right) \tag{13}$$

Vector $\mathbf{X}^u_{\text{PUSCH}}(i)$ represents the vector of observed parameters associated with the u-th user's QoS at time-step i, structured as an $M \times 1$ vector. $\mathbf{X}_{\text{opt},b,f,c}(i)$ denotes the vector containing the optimal values of QoS parameters at RSU c

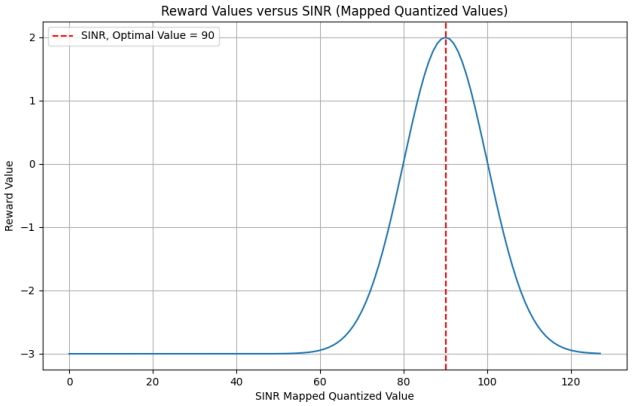

FIGURE 3. Graph of the proposed reward function for a single-variable for the range of SINR reported value with the desired value of 90 (21.5 dB).

across BWP b of carrier f at time-step i, mirroring the shape of vector $\mathbf{X}^u_{\text{opt}}(i)$. The symbol T represents the transpose operator, converting the output vector of subtraction from $M \times 1$ to $1 \times M$, where M signifies the number of QoS parameters.

In our scenario with the focus on SINR, $M = 1$ and the reward function simplifies to the below equation and depicted in Figure 3.

$$\text{Rwd}^u(i) = (\text{r}_{\max} - \text{r}_{\min}) \exp\left(-\frac{\left(\gamma^u_{\text{PUSCH},b,f,c}(i) - \gamma_{\text{opt},b,f,c}(i)\right)^2}{2g}\right) + \text{r}_{\min} \quad (14)$$

In (14), $\gamma^u_{\text{PUSCH},b,f,c}(i)$ represents the achieved SINR of the u-th user during uplink transmission at time-step i. $\gamma_{\text{opt},b,f,c}(i)$ denotes the optimal SINR at time-step i for RSU c, BWP b, and carrier frequency f.

Considering our system setup as explained above and the PPO algorithm presented in Section II, the prediction and training process of our DNN is as follows. At each time-step, the gNodeB collects all the measurement reports (including the pathloss and PUSCH SINR measurements) from all users connected to its RSUs. The gNodeB create the state vector of each user as expressed in (10). These state vectors serve as inputs to the DNN, and the outputs (actions) are then sent to the users as their power levels for the next time-step. Meanwhile, in a separate processing thread and without interrupting the real-time power control algorithm, the PPO algorithm utilizes this recent interaction with the environment to train and update the DNN for subsequent decisions. Therefore, while the DNN model is employed for power control of users at each time-step, the PPO algorithm utilizes each time-step's experiences to gradually train and enhance the performance of the DNN model. The PPO algorithm inherently possesses rapid change limitations due to the clip function and the ϵ hyper-parameter as explained earlier. Also, the bell-shaped design of the reward function encourages the algorithm to remain near optimal values of the QoS effective parameters applied in (13). Consequently, the DNN cannot change significantly at each time-step, thereby reducing the risk of algorithm divergence while enabling it to adapt the model to changes in the dynamic environment of the road.

Based on the structure of the proposed power control algorithm explained above, its distinguishing features are as follows.

- By satisfying the desired capacity instead of only focusing on maximizing it, our algorithm utilizes the minimum required power for each uplink channel, thereby automatically controlling and minimizing co-channel interference.
- Minimizing the interference will result in the improvement of network spectral efficiency and an increase in system capacity.
- The proposed algorithm can instantly adapt the user's power level based on real-time channel changes, contrasting with traditional 5G iterative algorithms that require more time to converge to the proper power level.
- The PPO's mini-batch update ability helps update the DNN model with immediate observations without requiring large storage space, unlike Deep Q-learning-based power control algorithms.
- The suggested algorithm can provide the desired QoS with multiple effective parameter for all users.
- Our algorithm uses quantized measurement data already envisioned in the network, rather than exact SINR or full Channel State Information (CSI) assumptions made by some other algorithms.
- Our algorithm does not use or rely on user location information that is not available/easily accessible in gNodeB, as assumed in some resource allocation articles.
- The proposed algorithm is adaptable to environmental changes and can learn from experiences, enabling it to make better decisions as the environment evolves over time.
- Since the algorithm provides the demanded QoS with minimum power, it improves users' energy efficiency.
- This algorithm can adjust users' power levels to achieve any demanded link capacity, including the maximum capacity if needed, providing the freedom to attain any capacity on demand.
- The proposed algorithm is specifically tailored for uplink transmission, addressing the complexities and challenges inherent in this aspect of 5G networks. However, with minor modifications, it can be adapted for downlink power control as well, extending its applicability across both uplink and downlink channels.

VI. SIMULATION RESULTS

In this section, we provide a performance comparison of the proposed PPO-DRL-based power control algorithm with other algorithms through extensive simulations conducted in the Python language.

A. PROPOSED ALGORITHM'S SIMULATION SETUP

We investigate a road segment serviced by 5G RSUs located in the road median, ensuring coverage for both directions of traffic. Our study centers on CAVs' V2I connection in uplink transmission via PUSCH, focusing on the mm-wave band FR2-1 of the 3GPP standard, specifically at 28 GHz. Each channel has a 100 MHz bandwidth with a 2.45 MHz guard-band between consecutive channels. Users maintain a direct LoS connection with the nearest RSU.

We utilize a channel allocation mechanism proposed in [23], designed for a 5G mm-wave vehicular network akin to this study's assumptions. According to this allocation, one channel can serve two CAVs with only one RSU between their serving RSUs (see Figure 2). Consequently, each user may encounter a minimum of two dominant interferers nearby (one in front and one behind the CAV). While assuming sufficient channels to accommodate all users,

TABLE 3. Assumptions of the road and network parameters.

Parameter	Value
Carrier frequency	28 GHz
Channel bandwidth	100 MHz
Channel guardband	2.45 MHz
EIRP maximum	30 dBm
EIRP minimum	-15 dBm
EIRP steps	1 dBm
Antenna gain on main-lobe	35 dBi
Antenna sidelobe gain	25 dBi
Noise figure	10 dB
Temperature (Kelvin)	300
Rician k-factor	9.0
Rician scale parameter	0.11
Rician location parameter	0
Minimum time step	1 ms
Road type	Two-way
Number of RSUs	8
Number of lanes	2
Road lane's width	3.5 m
RSU height	12 m
Avg. vehicle length	5 m
Min. distance between vehicles	10 m
Avg. vehicle speed	100 Km/h
Distance-Threshold	1 cell

we allocate the minimum required number of channels to each configuration to simulate intense interference scenarios.

In our simulations, each RSU and user is equipped with one antenna, with the RSU antenna gain set at 35 dBi. The EIRP level, determined by users' uplink transmission antenna gain and output power, is exclusively used throughout all simulations. RSUs have a 25 dBi antenna gain on the sidelobe towards interferers, while users' transmission antennas directed towards other RSUs incur a 5 dB loss.

In these simulations, one lane per direction is considered only for processing simplicity, but our algorithm can be applied to any number of lanes without negative impact. The algorithm is trained with RSU coverage lengths ranging from 50 to 500 meters, resulting in varying simulated road lengths based on the total of 8 RSUs considered. The total number of users depends on the road length and traffic intensity.

For the large scale fading model of the channel, the Free Space Path Loss (FSPL) model is considered and Rician distribution for the small scale fading model of the uplink channel [39]. Table 3 presents the network parameters assumption in our simulations and Table 4 lists the hyper-parameter values of our DNN and PPO algorithm.

The actor in the PPO algorithm utilizes a DNN model, with the inputs being the state vector for each user (As presented in (10)) and the output (action) being the user's uplink power level. Both the input features and output values are discrete.

TABLE 4. Assumptions of the DNN and PPO hyper-parameters.

Parameter	Value
DNN number of input features	11
DNN number of hidden layers	7
DNN hidden layer connection types	Dense
DNN number of hidden nodes per layer	128
DNN hidden layers activation func.	Tanh
DNN number of output nodes	46
PPO learning rate	3e-5
PPO epsilon	0.2
Optimizer	Adam
Maximum reward	2
Minimum reward	-3
Reward function g parameter	10
Minimum SINR [dB]	-10
Training iterations per batch	3
KL divergence threshold	0.01

To ensure uniformity across input values, we normalize each state vector value to a range of 0 to 1. All hidden layers are fully connected with a dense architecture and employ the hyperbolic tangent activation function. The output layer consists of 46 nodes, representing values ranging from -15 to 30 dBm, and utilizes a linear activation function.

B. REFERENCE ALGORITHM'S SIMULATION SETUP

We compare the performance of the proposed scheme with four power control algorithms. These algorithms include a simplified 3GPP-based power control for 5G [8], the Maximum-Power allocation, and the Equal-Power allocation [16].

The 3GPP-based power control algorithm for the 5G network, as briefly explained in Section II, is complicated for exact implementation and depends on many variables. The value of each of these variables is selected based on a rule or using an algorithm within a network's function. Since the algorithms for these variables are inaccessible, we employ a simplified model of the equation provided in (1), utilizing the available information. The variable $P_{\mathrm{CMAX},f,c}(i)$ is set to 30 dBm, representing the maximum EIRP value discussed in this paper. $P_{\mathrm{O\_NOMINAL,PUSCH},f,c}(j)$ dependent on frequency f, carrier c, and parameter j, all of which remain constant throughout our simulations. Therefore, $P_{\mathrm{O\_NOMINAL,PUSCH},f,c}(j)$ is set to -60 dBm, approximately the mid-range value. The variable $P_{\mathrm{O\_UE\_PUSCH},b,f,c}(j)$ ranges from -16 dBm to 15 dBm. We set it to zero for the same reason as $P_{\mathrm{O\_NOMINAL,PUSCH},f,c}(j)$. The SCS variable μ ranges from 0 to 6, representing sub-carrier spacings of 15 kHz to 960 kHz, respectively. We set it to zero, equivalent to $\Delta f = 15$ kHz. The PUSCH bandwidth variable $M^{\mathrm{PUSCH}}_{\mathrm{RB},b,f,c}(i)$ is considered equal to 50, as it is an integer ranging from 10 to 100 and depends on fixed b, f, and c parameters. The path loss gain factor $\alpha_{b,f,c}(j)$ can assume

one of the following values: zero, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, or 1. After extensive simulations to determine the optimal value for the given environment, it is set to 0.4. $\Delta_{\mathrm{TF},b,f,c}(i)$ is also set to zero, equivalent to single-layer transmission, as defined in (1). The function $f_{b,f,c}(i,l)$ is modeled with a weighted SINR error relative to the desired SINR, denoted by $\beta(\gamma_{\mathrm{opt}}-\gamma(i))$, where β is set to 0.5 due to its best performance confirmed by simulations. The value of $\mathrm{PL}_{b,f,c}$ is reported by users at each time-step as a quantized value through their measurement reports, while SINR is measured and provided by the gNodeB.

The Maximum-Power algorithm allocates a power level of 30 dBm to each user, serving as an upper limit on achievable capacity in the event that all users aim for maximum power. This results in both the signal level and interference reaching their peaks. It should be noted that this is different from achieving the maximum possible capacity, which requires effective interference mitigation strategies. Maximum link capacity can be attained through an optimal power control algorithm.

In an Equal-Power allocation algorithm, akin to the Maximum-Power algorithm, all users' power levels are set to an equal value at each time step. Within our simulations, we employ two distinct Equal-Power algorithms: Equal-Power-1 and Equal-Power-2. In Equal-Power-1, all users' power levels at time-step i are identical and adjusted to match the average of the powers allocated by our algorithm (the Proposed-Algorithm) at that time-step. Conversely, Equal-Power-2 maintains a fixed power level of 5 dBm for all users at all times. By employing both Maximum- and Equal-Power algorithms, we aim to discern the performance disparities between a straightforward equal power allocation method and more intricate algorithms such as the proposed one.

C. PERFORMANCE COMPARISONS

To demonstrate the settling time and power allocation efficiency of the algorithms, we perform 10,000 random initialization of the entire network. This process involves randomizing users' locations, channel assignments, and power allocations. At each iteration, each user takes a random power level from the valid range of -15 dB to 30 dB by a uniform distribution. This is an extreme case scenario whereas in the network all users would never restart at the same time and begin transmitting by a random power. However, it helps to evaluate our proposed algorithm's behavior in total random situation.

Figures 4, 5, and 6 depict the users' achieved SINR with different power control algorithms at 3, 5, and 10 time-steps after network initialization, respectively. The graph for each algorithm in each figure represents the results for 150 users across 8 RSUs with coverage length of 100 meters and over 10,000 random iterations.

In these figures, the optimal SINR per link is set to 21.5 dB, equivalent to a reported SINR value of 90 and a link capacity of approximately 715 Mbps. It should be noted that the SINR value of 90 is provided as an example and can be set to any arbitrary value. However, for consistency and the ability to compare results across different simulations, we maintain the same desired SINR value in our simulations. At each initialization, all algorithms experience the same user locations and environmental conditions, but the channel fading is different and randomly generated for each link with similar characteristics. Due to the random positioning, channel allocation, and power allocation of users at each initialization, each link experiences a new random SINR within a wide range of values, resulting in a significantly different state for each user. These states serve as inputs to the power control algorithms for adjusting user power levels at the next time-step.

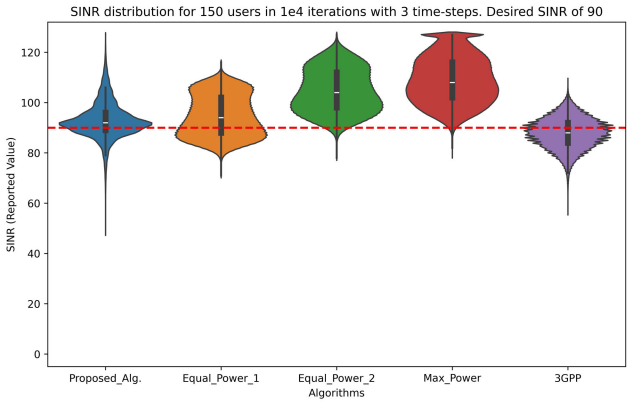

FIGURE 4. Achieved SINR of 1,500,000 data points across different algorithms, observed 3 time-steps after random initialization of the environment.

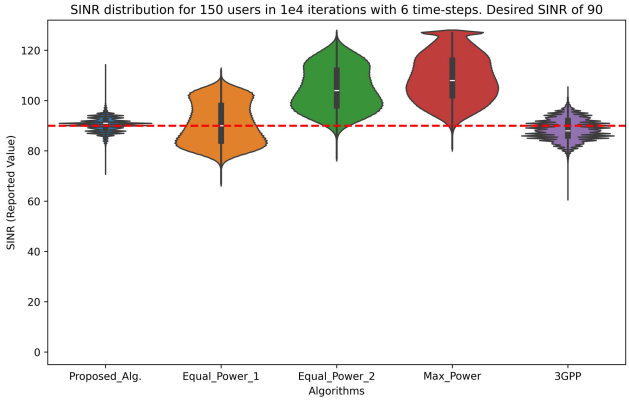

FIGURE 5. Achieved SINR of 1,500,000 data points across different algorithms, observed 6 time-steps after random initialization of the environment.

The updated SINR levels after the specified time-steps for each figure, reflecting the outcomes of the algorithms' power controls, are illustrated in Figures 4 to 6. The horizontal axis separates the algorithms, while the vertical axis represents the range of achieved reported-value SINR for each algorithm (ranging between 0 to 127). In each algorithm's graph, the width alongside the horizontal axis indicates the probability density of the data at that SINR value. The average of achieved SINR values for each algorithm is depicted with a small white dash-line on the middle vertical black straight

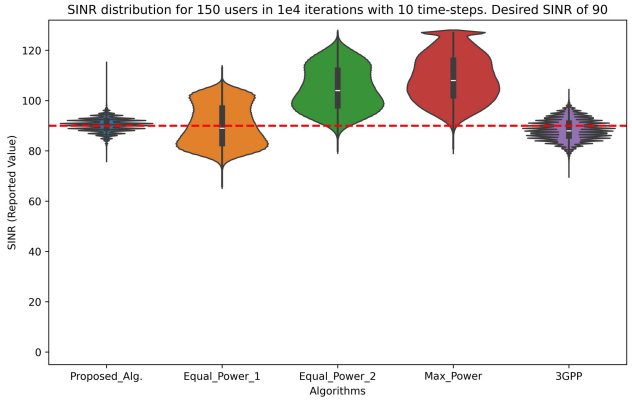

FIGURE 6. Achieved SINR of 1,500,000 data points across different algorithms, observed 10 time-steps after random initialization of the environment.

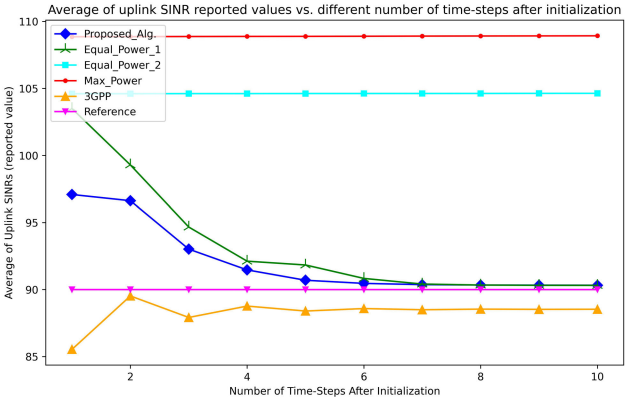

FIGURE 7. Average uplink SINR changes over multiple time-steps for 150 CAVs distributed across 8 RSUs, employing different power control algorithms.

line, while the thick black section represents the median of the data.

Based on the results in Figure 4, 3 time-steps after random initialization, the average achieved SINR of CAVs using both the proposed DNN-based and 3GPP-based algorithms is close to the desired SINR of 90 (21.5 dB). The average SINR of the Equal-Power and Max-Power algorithms is further away from the desired value. So, the performance of our algorithm and the 3GPP-based approach after three time-steps is similar, but the probability distribution of SINRs differs, as illustrated in Figure 4. The standard deviation of achieved SINR values for the proposed algorithm is approximately 7.1 units, while for the 3GPP-based algorithm, it is about 6 units. Other algorithms exhibit a range between 8.2 and 9.2 units as it can be observed in Figure 8.

At 4 time-steps after the initialization, the statistics (average and the standard deviation) of the users' SINR with the proposed algorithm is better than the 3GPP-based algorithm. 6 time-steps after the random initialization, users' SINR values with both the proposed DNN-based and the 3GPP-based algorithms are very close to their long term statistics. At this stage and after it, the proposed algorithm's statistics and the distribution of the SINRs is considerably better than the 3GPP-based algorithm as shown in Figure 5.

The average SINR values of the proposed algorithm at 6 time-steps demonstrate stabilization of users' powers close to the desired SINR. The average SINR value of the proposed algorithm is 90.7 with a standard deviation of 3 at 5 time-steps, while the average is 90.4 with standard deviation of 2.7 nearing the desired SINR value of 90 and very close to its results at 10 time-steps. Considering the noisy wireless channel with large-scale and small-scale fading, the achieved result is close enough to be acceptable.

Comparing the results of the 3GPP-based algorithm in Figure 5 with those in Figure 4, we observe improvements over time in both the average and standard deviation of users' SINRs. As expected, the Equal- and Maximum-Power allocation algorithms exhibit constant SINR statistics once all users' power levels are allocated, a process occurring within a single time step. The only variation lies in the average SINR of the Equal-Power-1 algorithm, as users' power levels are set equal to the average allocated power of the proposed algorithm.

The results depicted in Figure 6 allow us to observe the long-term changes in probability density of the algorithms. Both the proposed algorithm and the 3GPP-based algorithm have shown slight improvements in their average and standard deviation of SINRs beyond 6 time-steps, although these improvements are not very significant. However, the probability density of the uplink SINR values for the proposed algorithm at 10 time-steps more closely resembles a normal distribution with the average over the desired SINR level.

In Figure 7, we present the average SINR levels on uplink transmission, utilizing various power control algorithms across 8 RSUs covering a diameter of 100 meters length of a road (per RSU) on a 2-lane road. The horizontal axis represents the number of time-steps after network initialization, while the vertical axis denotes the average SINR values computed from 1.5 million data points (150 users over 10,000 iterations). To facilitate comparison, a magenta horizontal line with a ▼ symbol is added that indicate the reference desired SINR level of 90 (21.5 dB).

Figure 8 illustrates the standard deviation of the achieved SINRs for different algorithms across 1 to 10 time-steps following initialization under the same environmental conditions depicted in Figure 7. Analysis of the graphs in Figures 7 and 8 suggests that all algorithms achieve stability after 6 time-steps in the simulated environment, with SINR statistics showing negligible variation thereafter. Consequently, it can be inferred that user power levels stabilize at 6 time-steps. However, individual user power levels may change due to relocations and channel changes, while total network power statistics remain stable. Comparison of the proposed algorithm with others beyond 6 time-steps, as presented in Figures 7 and 8, indicates superior performance of the proposed algorithm.

Figure 9 depicts the average of uplink capacities per user over 10,000 iterations in a channel with 100 MHz bandwidth

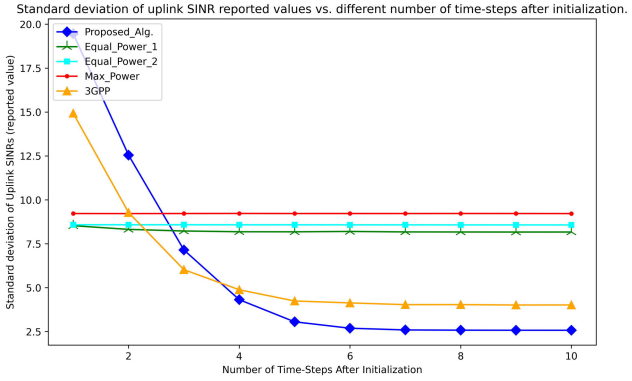

FIGURE 8. Standard deviation of uplink SINR changes over multiple time-steps for 150 CAVs distributed across 8 RSUs, employing different power control algorithms.

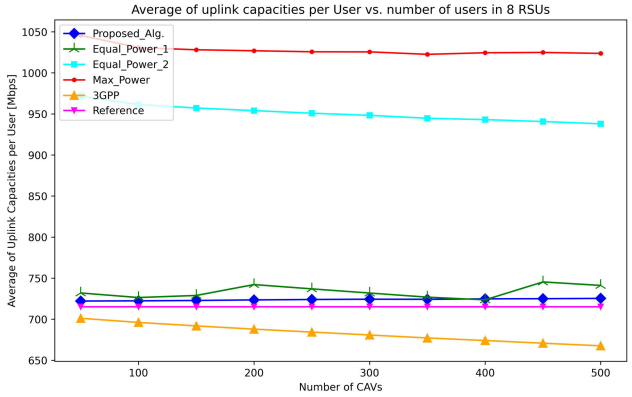

FIGURE 9. Average uplink capacity per channel across 8 RSUs for a diverse range of traffic (number of CAVs) with a desired capacity of 715 Mbps.

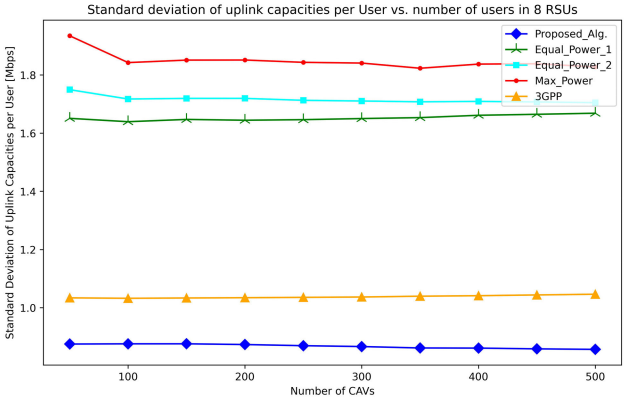

FIGURE 10. Standard deviation of uplink capacity per channel across 8 RSUs for a diverse range of traffic (number of CAVs) with a desired capacity of 715 Mbps.

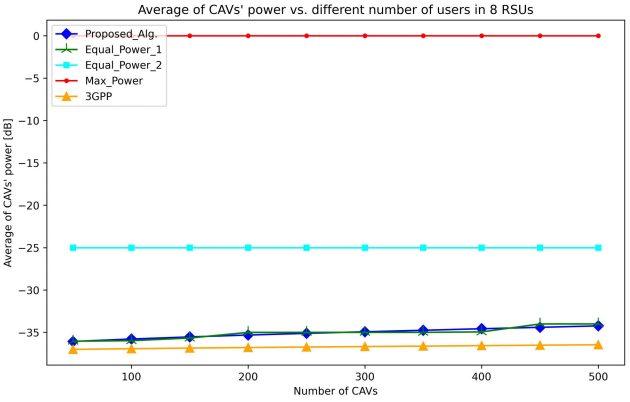

FIGURE 11. Average power [dB] of CAVs across 8 RSUs under varying traffic loads, with a desired SINR of 90.

while their uplink transmission power is controlled by the considered algorithms. These graphs are the results of power control on 6 time-steps after the network initialization which users' power levels are stabilized. Similar to previous figures, in this figure also a section of a road equivalent to coverage area of 8 RSUs with 100 meter coverage length alongside the road is investigated. The considered section of the road has one lane on each direction and this figure covers different number of users on the road ranging between 50 to 500 users which are simulating very low traffic to very high traffic on the road. The reference average capacity line for a channel with 100 MHz bandwidth and SINR of 90 (21.5 dB) is added to this figure in magenta color and ▼ symbol.

By comparing the results of different algorithms in Figure 9, it is evident that the proposed algorithm achieves results closest to the desired channel capacity and maintains this level across varying traffic loads within the network. The Equal-Power-1 algorithm also yields results close to the desired capacity since its power levels are equal to the average of the proposed algorithm. However, its stability falls short compared to the proposed algorithm, exhibiting fluctuations across different numbers of users on the road. The 3GPP-based algorithm's achieved average channel capacities are almost close to the reference line (desired capacity), but its worse than the proposed algorithm, and its performance decreases with an increase in the number of users on the road.

Figure 10 illustrates the standard deviation of the uplink capacities of CAVs using different power control algorithms over 10,000 iterations under the same channel and environmental conditions as described for Figure 9. According to the graphs in Figure 10, the proposed algorithm exhibits the best performance in terms of achieved capacity stability. Our algorithm demonstrates the lowest variation in achieved capacity across various traffic loads on the road, with the standard deviation even slightly decreasing as the traffic load increases. The 3GPP-based algorithm presents a steady standard deviation that is higher than that of the proposed algorithm. The Equal- and Maximum-Power algorithms exhibit the highest standard deviation for users' capacities, as these algorithms employ fixed power levels for all users, leading to high variation in channel capacities due to channel variations.

Figure 11 illustrates the average allocated power to users employing different power control algorithms with the achieved capacity statistics depicted in Figures 9 and 10. Based on the graphs in Figure 11, both the 3GPP-based and the proposed algorithms have an average power level

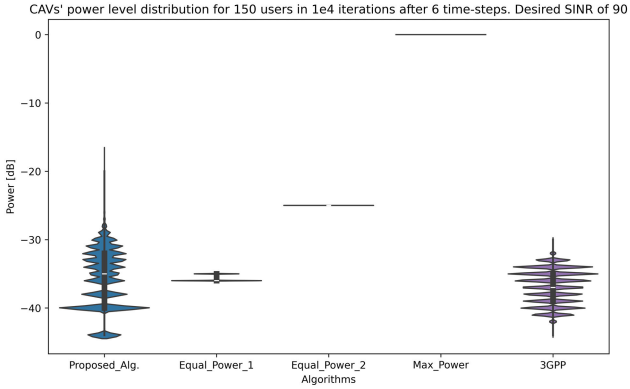

FIGURE 12. Distribution of allocated power to 150 users over 10,000 iterations by various power control algorithms across 8 RSUs, with a target SINR of 90.

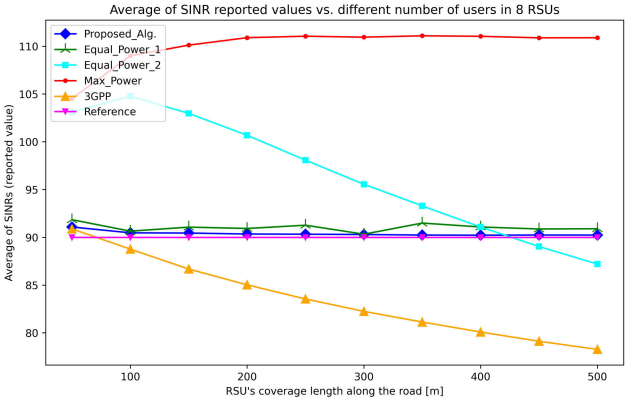

FIGURE 13. Average SINRs on uplink channels across varying RSU coverage lengths along the road for 8 RSUs in 10,000 iterations with a fixed traffic intensity.

ranging between −36 dB and −34 dB. The average power of the 3GPP-based algorithm remains relatively constant across varying traffic loads, resulting in a decline in the average achieved capacity of users as the number of users increases (as shown in Figures 9). In contrast, the proposed algorithm's average allocated power gradually increases with the rise in the number of users to mitigate the growing interference in the network. Consequently, our algorithm maintains a stable average capacity at demanded level for users under various traffic conditions.

In Figure 11, we observed the average allocated power levels for different number of users. To gain insight into the distribution of allocated power to 150 CAVs across 8 RSUs over 10,000 iterations, at 6 time-steps after initialization, Figure 12 is provided. Since the Equal-Power-2 (−25 dB) and Max-Power (0 dB) algorithms allocate a fixed power level to all users, their power level distributions appear as flat lines at equivalent power levels.

The power levels of CAVs in the Equal-Power-1 algorithm change according to the average of the proposed algorithm. Consequently, the distribution of power values of 150 CAVs utilizing the Equal-Power-1 algorithm over 10,000 iterations is concentrated around two power levels. In contrast, the proposed algorithm employs a broader range of power levels to adapt to varying channel conditions and achieve the desired channel capacity. Meanwhile, the 3GPP-based algorithm exhibit an almost uniform distribution of power values ranging between −41 and −32 dB.

It is worth noting that the distributions of power values in the 3GPP-based and proposed algorithms are jagged due to quantized integer power levels available to algorithms to allocate to users, ranging between −45 dB and 0 dB with steps of 1 dB.

In order to observe the impact of RSU coverage length (along the road) on algorithm performance, Figure 13 has been generated. This figure illustrates a section of a 2-lane two-way road serviced by 8 RSUs, each with equal coverage length and traffic intensity. The horizontal axis displays the range of coverage length per RSU, varying from 50 to 500 meters in increments of 50 meters. For instance, a 200-meter coverage length corresponds to a section of 1600 meters of a road. To maintain uniform traffic load across all coverage lengths, the average distance between consecutive CAVs remains fixed at 10 meters throughout the simulation. Consequently, the number of CAVs on the road for a coverage length of 200 meters is almost twice that of the coverage length of 100 meters.

The results of the average SINRs of the proposed algorithm across 8 RSUs with varying coverage lengths, as depicted in Figure 13, demonstrate the robustness of our power control algorithm across various RSU coverage diameters. The average SINRs of uplink channels, when using the 3GPP-based algorithm, gradually decrease with increasing RSU coverage length. This performance degradation of the 3GPP-based algorithm in Figure 13 is attributed to the limited time available for this algorithm to adjust the power levels, as results in this figure obtained after 6 time-steps post-initialization. As, the 3GPP-based algorithm requires additional time in larger cell sizes, because at each time-step, it can only increment user power levels by a restricted amount. Consequently, for users near the cell edge, the disparity between the random initial power and the desired power level may be too substantial to rectify within 6 time-steps. Thus, iterative algorithms like the 3GPP-based algorithm necessitate more time to stabilize user power levels within larger RSUs.

The average SINRs on the Equal-Power-2 algorithm are dropping rapidly with the increase of RSUs' coverage area, as users on the cell edge in larger RSUs require more power to overcome interference and noise. Therefore, a constant power level (−25 dB in the Equal-Power-2 algorithm) is not sufficient in larger RSUs. However, conditions for the Max-Power algorithm are different, as the allocated power of 0 dB in this algorithm is adequate for each user to communicate with its RSU even at cell edge, while the increase of RSUs' coverage area will reduce interference. As a result, in the Max-Power algorithm, with the increase of RSUs' coverage area, the average SINR gradually increases until it reaches a balance between the power of desired signal and interference on RSUs' coverage length above 200 meters.

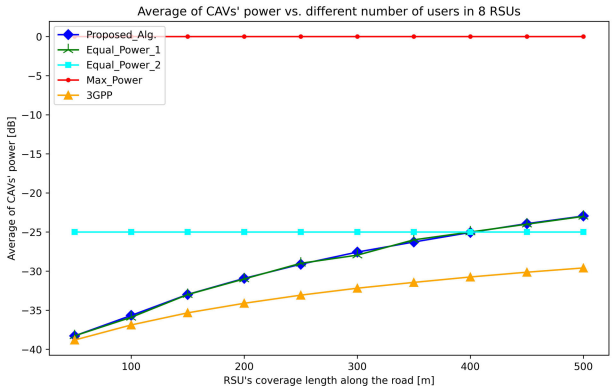

FIGURE 14. Average allocated power [dB] to CAVs in each algorithm across various RSUs' coverage lengths, evaluated over 10,000 iterations with fixed traffic intensity and involving 8 RSUs.

Figure 14 displays the average allocated power to CAVs across varying RSU coverage lengths to achieve the desired SINR of 90 (21.5 dB). This figure illustrates the allocated power to CAVs by different algorithms, while the achieved SINRs by applying these power levels are demonstrated in Figure 13.

In Figure 14, we observe a gradual increase in average allocated power levels with the enlargement of RSU sizes to compensate for the elongated distance between the CAVs and their serving RSUs, in both the proposed algorithm and the 3GPP-based algorithm. However, the proposed algorithm demonstrates agility in terms of power level adjustment, enabling rapid stabilization of users' power levels irrespective of the environment size. In contrast, the 3GPP-based algorithm requires more time to reach the final desired power level in larger RSUs, explaining the significant gap between its average power and our algorithm at higher RSU coverage lengths.

As we observed in above simulations, the proposed power control algorithm outperforms the 3GPP-based algorithm in achieving the desired link capacity for all users in various conditions. This advantage is particularly notable in high dynamic environments with battery powered users like in 5G-based vehicular networks.

VII. CONCLUSION

In this paper, we introduced a novel DRL-based power control algorithm tailored for vehicular users in 5G networks. Our focus was on optimizing PUSCH transmissions between vehicular users and mm-wave band RSUs of the 5G gNodeB infrastructure. Our objective centered on improving the performance of power control algorithm for CAVs connection to the 5G small-cells, aiming to achieve desired uplink channel capacities while maintaining QoS standards, considering the individual transmitters' power limits, and staying compatible with the 3GPP-based 5G network architecture. That means, the environment state information will be limited to the quantized data reported by users on their measurement reports.

To address this challenge, we proposed a centralized power control algorithm integrated into the gNodeB's RRC function. Our approach leverages a DNN architecture with normalized quantized input states and quantized power level outputs. The DNN model is trained and continually updated using PPO with a modified actor-critic architecture, incorporating recent interaction data to enhance model performance.

Our extensive simulations demonstrate that the proposed algorithm consistently outperforms existing 3GPP-based power control methods, achieving desired capacities within comparable time intervals or even faster in certain scenarios. Furthermore, our algorithm ensures demanded SINR level across varying traffic loads and RSU coverage areas, with minimal variance and average error. By optimizing transmission power, our approach mitigates co-channel interference, enhances spectrum efficiency, and promotes energy efficiency among users.

In conclusion, our research contributes a robust solution for optimizing power control in 5G vehicular networks, aligning with the demands of future CAV deployments while upholding network performance standards and compatibility with existing 3GPP frameworks.

REFERENCES

[1] L. Wang, H. Liang, and D. Zhao, "Deep-reinforcement-learning-based computation offloading and power allocation within dynamic platoon network," *IEEE Internet Things J.*, vol. 11, no. 6, pp. 10500–10512, Mar. 2024, doi: 10.1109/jiot.2023.3327712.

[2] B. P. Nayak, L. Hota, A. Kumar, A. K. Turuk, and P. H. J. Chong, "Autonomous vehicles: Resource allocation, security, and data privacy," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 1, pp. 117–131, Mar. 2022, doi: 10.1109/TGCN.2021.3110822.

[3] P. Qian, L. Wang, Y. Lin, J. Du, and X. Dong, "Joint power allocation and task offloading in NOMA enhanced MEC for ABS-assisted ITS," *IEEE Commun. Lett.*, vol. 27, no. 9, pp. 2403–2407, Sep. 2023, doi: 10.1109/LCOMM.2023.3295057.

[4] A. A. Alkheir, M. Aloqaily, and H. T. Mouftah, "Connected and autonomous electric vehicles (CAEVs)," *IT Prof.*, vol. 20, no. 6, pp. 54–61, Nov. 2018, doi: 10.1109/MITP.2018.2876977.

[5] *Service Requirements for V2X Services; Stage 1*, Standard TS 22.185, V17.0.0, Release 17, 3GPP, Mar. 2022, pp. 1–14.

[6] Y. Yang, T. Lv, Y. Cui, and P. Huang, "MADRL based uplink joint resource block allocation and power control in multi-cell systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2023, pp. 1–6, doi: 10.1109/WCNC55385.2023.10119106.

[7] J. Xu and B. Ai, "Experience-driven power allocation using multi-agent deep reinforcement learning for millimeter-wave high-speed railway systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5490–5500, Jun. 2022, doi: 10.1109/TITS.2021.3054511.

[8] *NR; Physical Layer Procedures for Control*, Standard TS 38.213, V18.0.0, Release 18, 3GPP, Sep. 2023, pp. 12–85.

[9] F. Fang, Z. Ding, W. Liang, and H. Zhang, "Optimal energy efficient power allocation with user fairness for uplink MC-NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1133–1136, Aug. 2019, doi: 10.1109/LWC.2019.2908912.

[10] Y. Kwon, H. Baek, and J. Lim, "Uplink NOMA using power allocation for UAV-aided CSMA/CA networks," *IEEE Syst. J.*, vol. 15, no. 2, pp. 2378–2381, Jun. 2021, doi: 10.1109/JSYST.2020.3028884.

[11] G. Zhao, Y. Li, C. Xu, Z. Han, Y. Xing, and S. Yu, "Joint power control and channel allocation for interference mitigation based on reinforcement learning," *IEEE Access*, vol. 7, pp. 177254–177265, 2019, doi: 10.1109/ACCESS.2019.2937438.

[12] D. Guo, L. Tang, X. Zhang, and Y.-C. Liang, "Joint optimization of handover control and power allocation based on multi-agent deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13124–13138, Nov. 2020, doi: 10.1109/TVT.2020.3020400.

[13] X. He, Y. Mao, Y. Liu, P. Ping, Y. Hong, and H. Hu, "Channel assignment and power allocation for throughput improvement with PPO in B5G heterogeneous edge networks," *Digit. Commun. Netw.*, vol. 10, no. 1, pp. 109–116, Feb. 2024, doi: 10.1016/j.dcan.2023.02.018.

[14] Y. Liu, M. Derakhshani, and S. Lambotharan, "Outage analysis and power allocation in uplink non-orthogonal multiple access systems," *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 336–339, Feb. 2018, doi: 10.1109/LCOMM.2017.2769088.

[15] S. Alenezi, C. Luo, and G. Min, "Energy-efficient power control and resource allocation based on deep reinforcement learning for D2D communications in cellular networks," in *Proc. 20th Int. Conf. Ubiquitous Comput. Commun. (IUCC/CIT/DSCI/SmartCNS)*, Dec. 2021, pp. 76–83, doi: 10.1109/IUCC-CIT-DSCI-SmartCNS55181.2021.00026.

[16] E. Yaacoub and Z. Dawy, "A survey on uplink resource allocation in OFDMA wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 2, pp. 322–337, 2nd Quart., 2012, doi: 10.1109/SURV.2011.051111.00121.

[17] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57–73, Feb. 2008, doi: 10.1109/JSTSP.2007.914876.

[18] T. Bonald, L. Massoulié, A. Proutière, and J. Virtamo, "A queueing analysis of max-min fairness, proportional fairness and balanced fairness," *Queueing Syst.*, vol. 53, nos. 1–2, pp. 65–84, Jun. 2006, doi: 10.1007/s11134-006-7587-7.

[19] K. Xiong, P. Fan, Y. Zhang, and K. B. Letaief, "Towards 5G high mobility: A fairness-adjustable time-domain power allocation approach," *IEEE Access*, vol. 5, pp. 11817–11831, 2017, doi: 10.1109/ACCESS.2017.2712710.

[20] H. Shi, R. V. Prasad, E. Onur, and I. G. M. M. Niemegeers, "Fairness in wireless networks: Issues, measures and challenges," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 5–24, 1st Quart., 2014, doi: 10.1109/SURV.2013.050113.00015.

[21] *NR; User Equipment (UE) Radio Transmission and Reception; Part 2: Range 2 Standalone*, Standard TS 38.101-2, V18.3.0, Release 18, 3GPP, Sep. 2023, pp. 12–28.

[22] *NR; Physical Channels and Modulation*, Standard TS 38.211, 3GPP, V18.0.0, Release 18, Sep. 2023, pp. 11–59.

[23] M. Raeisi and A. B. Sesay, "Handover reduction in 5G high-speed network using ML-assisted user-centric channel allocation," *IEEE Access*, vol. 11, pp. 84113–84133, 2023, doi: 10.1109/ACCESS.2023.3297982.

[24] *NR; Requirements for Support of Radio Resource Management*, Standard TS 38.133, V18.3.0, Release 18, Sep. 2023.

[25] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, no. 3, pp. 237–252, Mar. 1998, doi: 10.1057/palgrave.jors.2600523.

[26] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*.

[27] M. Hosseini and R. Ghazizadeh, "Stackelberg game-based deployment design and radio resource allocation in coordinated UAVs-assisted vehicular communication networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 1196–1210, Jan. 2023, doi: 10.1109/TVT.2022.3206145.

[28] X. Hu, S. Xu, L. Wang, Y. Wang, Z. Liu, L. Xu, Y. Li, and W. Wang, "A joint power and bandwidth allocation method based on deep reinforcement learning for V2V communications in 5G," *China Commun.*, vol. 18, no. 7, pp. 25–35, Jul. 2021, doi: 10.23919/JCC.2021.07.003.

[29] H. Zhang, S. Chong, X. Zhang, and N. Lin, "A deep reinforcement learning based D2D relay selection and power level allocation in mmWave vehicular networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 416–419, Mar. 2020, doi: 10.1109/LWC.2019.2958814.

[30] R. Amiri and H. Mehrpouyan, "Self-organizing mm wave networks: A power allocation scheme based on machine learning," in *Proc. 11th Global Symp. Millim. Waves (GSMM)*, May 2018, pp. 1–4, doi: 10.1109/GSMM.2018.8439323.

[31] J. Shi, H. Pervaiz, P. Xiao, W. Liang, Z. Li, and Z. Ding, "Resource management in future millimeter wave small-cell networks: Joint PHY-MAC layer design," *IEEE Access*, vol. 7, pp. 76910–76919, 2019, doi: 10.1109/ACCESS.2019.2920745.

[32] X. Liao, J. Shi, Z. Li, L. Zhang, and B. Xia, "A model-driven deep reinforcement learning heuristic algorithm for resource allocation in ultra-dense cellular networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 983–997, Jan. 2020, doi: 10.1109/TVT.2019.2954538.

[33] D. Kwon, J. Kim, D. A. Mohaisen, and W. Lee, "Self-adaptive power control with deep reinforcement learning for millimeter-wave Internet-of-vehicles video caching," *J. Commun. Netw.*, vol. 22, no. 4, pp. 326–337, Aug. 2020, doi: 10.1109/JCN.2020.000022.

[34] H. Zhang, Z. Wang, and K. Liu, "V2X offloading and resource allocation in SDN-assisted MEC-based vehicular networks," *China Commun.*, vol. 17, no. 5, pp. 266–283, May 2020, doi: 10.23919/JCC.2020.05.020.

[35] Q. Guo, F. Tang, and N. Kato, "Federated reinforcement learning-based resource allocation in D2D-enabled 6G," *IEEE Netw.*, vol. 37, no. 5, pp. 89–95, Sep. 2023, doi: 10.1109/MNET.122.2200102.

[36] C. Guo, L. Liang, and G. Y. Li, "Resource allocation for high-reliability low-latency vehicular communications with packet retransmission," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6219–6230, Jul. 2019, doi: 10.1109/TVT.2019.2919181.

[37] J. D. Sorg, "The optimal reward problem: Designing effective reward for bounded agents, dissertation of doctor of philosophy," Ph.D. dissertation, Dept. Comput. Sci. Eng., Univ. Michigan, 2011.

[38] *NR; Base Station (BS) Radio Transmission and Reception*, Standard TS 38.104, 3GPP, V18.3.0, Release 18, Sep. 2023, pp. 13–96.

[39] M. K. Samimi, G. R. MacCartney, S. Sun, and T. S. Rappaport, "28 GHz millimeter-wave ultrawideband small-scale fading models in wireless channels," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–6, doi: 10.1109/VTCSPRING.2016.7503970.

[40] H. Yang, N. Cheng, R. Sun, W. Quan, R. Chai, K. Aldubaikhy, A. Alqasir, and X. Shen, "Knowledge-driven resource allocation for wireless networks: A WMMSE unrolled graph neural network approach," *IEEE Internet Things J.*, vol. 11, no. 10, pp. 18902–18916, 2024, doi: 10.1109/JIOT.2024.3368516.

MOSTAFA RAEISI received the B.S. degree in electrical engineering from KIAU, Karaj, Iran, in 2009, and the M.S. degree in electrical engineering from Iran University of Science and Technology (IUST), Tehran, Iran, in 2013. He is currently pursuing the Ph.D. degree in electrical and software engineering with the University of Calgary, Calgary, Alberta, Canada.

From 2009 to 2018, he was in telecommunication and electrical engineering in Tehran, Iran. His research interests include the wireless communications, autonomous vehicles, machine learning, big data, data science, and cloud computing.

ABU B. SESAY (Life Senior Member, IEEE) received the Ph.D. degree in electrical engineering from McMaster University, Hamilton, ON, Canada, in 1988. From 1979 to 1984, he worked on various International Telecommunications Union Projects. From 1986 to 1989, he was a Research Associate with McMaster University. In 1989, he joined the University of Calgary, Calgary, AB, Canada, where he is currently a Full Professor with the Department of Electrical and Computer Engineering and was the Department Head, from 2005 to 2011, and the Acting Associate Dean of Graduate Studies. From 1989 to 2005, he was a TRLabs Adjunct Scientist, where he conducted wireless research for various sponsors, including Nortel and Lucent. He spent sabbatical visits with Nortel Networks in Ottawa and Calgary. He has received numerous best paper awards with his students most notably the "1996 Neal Shepherd Memorial Best Propagation Paper Award" for the paper "Effects of Antenna Height, Antenna Gain, and Pattern Downtilting for Cellular Mobile Radio," published in IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (Vol. 45, No.2, May 1996). His current research interests include cooperative cellular wireless networks, orthogonal frequency-division multiple-access and code-division multiple-access systems, multiple-input-multiple-output systems, equalization, adaptive signal processing, heterogeneous wireless network resource and mobility management, advanced signal processing (including machine learning) for unmanned aerial vehicles (UAV) using GNSS and 5G assisted autonomous vehicles, and relay networks for LTE and 5G.