SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I5P107

Volume 12 Issue 5, 80-93, May 2025

Original Article

Enhancing Spectral Efficiency in Vehicular Optical Camera Communications Using Multi-Agent Deep Reinforcement Learning

A. Kondababu[1], S. Vinaya Kumar[2], K.H.K. Prasad[3], Kasetty Lakshminarasimha[4], U.S.B.K. Mahalaxmi[5]

[1,2,3,5]*Department of ECE, Aditya University, Surampalem, Andhra Pradesh, India.*
[4]*Department of ECE, SVR Engineering College, Nandyal, Andhra Pradesh, India.*

[1]*Corresponding Author : kondababu.amaradi@adityauniversity.in*

Received: 04 March 2025 Revised: 06 April 2025 Accepted: 08 May 2025 Published: 27 May 2025

Abstract *- Vehicular communication systems are very important for modern transportation. These systems allow vehicles to share data and improve road safety. Optical Camera Communication (OCC) is a new method that helps vehicles communicate using visible light. This method has many advantages over traditional Radio Frequency (RF) communication. It offers a larger spectrum, lower cost and better security. This paper focuses on optimizing spectral efficiency in vehicular OCC. It proposes a new approach using multi-agent Deep Reinforcement Learning (DRL). This method helps vehicles decide their speed and modulation to maximize spectral efficiency. The system also ensures a low Bit Error Rate (BER) and ultra-low latency. The main goal is to find the best modulation order and vehicle speed to increase spectral efficiency. This needs to be done while maintaining reliability and low latency. The problem is difficult to solve with traditional methods. The reason is that it is a mixed-integer programming problem with nonlinear constraints. A solution is proposed using Reinforcement Learning (RL). In this case, each vehicle acts as an autonomous agent. The vehicles learn the best way to adjust their speed and modulation order. This is done using a technique called Q-learning. However, since the problem is large and complex, DRL is used to improve learning efficiency. This paper presents a new way to improve spectral efficiency in vehicular OCC. It uses DRL to optimize speed and modulation order. The system meets reliability and latency constraints. The results show that this method is more effective than existing approaches.*

Keywords *- Deep Reinforcement Learning, Optical Camera Communication, Open car control, Markov decision process, Multi-agent reinforcement learning.*

1. Introduction

Vehicular communication plays a key role in modern Intelligent Transportation Systems (ITS) [1]. It improves road safety and traffic management by allowing vehicles to share data. RF communication [2] is used in traditional vehicular networks. Nevertheless, this increase in vehicles causes the network to become congested. [3] The RF spectrum is extremely limited and cannot meet the increasing demand for data exchange. Now, OCC has taken birth as an alternative technology to solve this problem. OCC is a form of Visible Light Communication (VLC) where LEDs act as transmitters and cameras act as receivers [4]. Benefits of OCC has a few benefits. Enhanced Security: The enhanced security it offers relative to RF systems, including lower costs and lower energy consumption. OCC works in the unlicensed spectrum band, which makes it suitable for vehicular communication [5]. OCC does have a few challenges despite its benefits. The system should provide high spectral efficiency, low BER, and ultra-low latency [6]. This problem is difficult to solve with traditional optimization algorithms. Hence, RL is a potential technique that can be used to enhance OCC systems' efficiency. Reinforcement learning-based methods are challenging to use in vehicular OCC due to the complex nature of the problem [7]. It is brave to train a centralized RL agent. The agent collects vehicle data, processes it, and sends optimal policies back. This results in high latency, congestion and sub-optimal decision-making. This also adds complexity to the state-action space as the vehicle count in the network rises. To address these problems, the problem is modeled as a Multi-Agent Reinforcement Learning (MARL) framework. Vehicles are agents that observe only local information [8]. Each vehicle independently discovers its policies without requiring a central agent. This approach minimizes communication overhead and enhances scalability.

In this work, a multi-agent DRL framework is introduced to maximize spectral efficiency in vehicular OCC. The aims of the proposed approach are to optimize spectral efficiency

through vehicular speed and maximum modulation order adjustment. This guarantees that the system satisfies BER and latency constraints [9]. Decrease global communications to add efficiency. It employs deep reinforcement learning for solving complex decision-making problems [10]. Formulation as a POMDP The problem formulation comprises a Partially Observable Markov Decision Process (POMDP). The system is based on Q-learning, in which each vehicle learns independently [11].

Meanwhile, Q-learning suffers from slow convergence and poor performance across large state spaces. DRL is used to approximate the state-action value function to address this issue. The system also employs a Lagrange relaxation method to simplify the constrained optimization problem [12]. This study mostly makes contributions to the following:

- Develop a DRL-based approach to optimize spectral efficiency in vehicular OCC.
- Modeling the problem as a partially observable MDP and designing a reward function to meet system constraints.
- Transforms the constrained problem into an unconstrained one using Lagrange relaxation.
- Implements independent learning and solving the problem with deep Q-learning.
- Evaluate the proposed method through simulations and compare it with traditional RF-based communication.

In this paper, a DRL-Based vehicular OCC scheme was proposed to maximize spectral efficiency. The proposed method allows for high-speed vehicle modulation and minimises the delayed BER required to be ultra-low. The system addresses the issue of centralized RL through independent learning [13]. The experimental evaluation demonstrates the superiority of DRL-based solutions over conventional RF systems as well as other vehicular OCC approaches. In the future, Further attention can be given to enhancing learning efficiency for this system and implementing its simulation in real-time environments.

2. Background Works

The vehicular OCC has been extensively researched because of its potential for both high spectral efficiency and security. A multitude of researchers have resorted to various techniques to improve the efficiency of OCC and vehicular networks. This section presents the relevant work in the fields of multi-agent DRL, reinforcement learning in vehicular networks, and OCC. Data is also studied for transmission, such as OCC systems, as high-speed data devices have the potential for an LED light source to an image sensor receiver. Owing to the rising demands, different methods have been suggested to enhance the data rate and reliability of OCC systems. For example, an automotive VLC system is designed in [14]. It employed an optical communication image sensor. The system will produce Results: A data rate of 55 Mbps for an average BER of less than. As for the automotive application, work in [15] also analyzed image sensor-based visible light communication. The study demonstrated that VLC is an alternative to RF communication. These systems showed good performance, but they did not take into account the mobility limitations of real-time systems or the need to derive spectral efficiency. The present study addresses this challenge by enhancing these efforts with DRL to optimize the performance of OCC in changing vehicular environments.

Vehicular networks have been actively studied by researchers to optimize resource allocation using DRL [16]. These approaches adapt RL techniques to improve the performance of spectrum sharing and power allocation. An RF-based system with vehicles communicating over V2V links was proposed, in which each V2V link acted as an agent in a deep reinforcement learning framework for spectral sharing. They applied the concept of MARL to decide how to allocate spectrum and power. While RF-based networks were examined in the study, OCC systems were not covered. In [16], channel allocation in vehicular networks was studied via a multi-agent DRL framework. Both methods outperform the traditional resource allocation methods. However, they were not designed to consider the challenges specific to OCC, such as visible light noise and limited field-of-view constraints [20].

Recently, given its effectiveness in optimizing communication performance, Multi-Agent Reinforcement Learning has recently been investigated for OCC and may potentially alleviate some of the difficulties in OCC. This work focuses on single-agent RL methods, while in [17], independent-learning approaches are considered for vehicular OCC systems. Spectral crowding leads to a high level of interference, which is one of the main issues for the existing RF-based MARL approaches. OCC can spatially separate multiple transmitter sources (as opposed to RF-based systems) to minimize interference. An RL-based method was proposed in [18] for network selection in VLC/RF heterogeneous networks. However, these studies focused solely on photodiode-based optical wireless communications over Visible Light Channels (VLC) since they did not consider the intrinsic characteristics of OCC.

The work proposed here expands on this previous research by utilizing an independent learning multi-agent reinforcement learning framework for optimizing spectral efficiency in vehicular OCC. The independent Q-learning method has lower communication overhead and is more scalable [19]. In this section, the related works in vehicular OCC and DRL-based resource optimization were surveyed. Although the potential for OCC for vehicular communication was indicated by previous works, the mobility constraints and latency requirements were not addressed sufficiently. Likewise, though DRL has been used in vehicular networks, current approaches do not address the specific challenges of OCC. To overcome this, the proposed research proposes a framework based on MARL to optimize the spectral efficiency for vehicular OCC.

3. Proposed Vehicular OCC Modelling

OCC is an advanced communication technique that utilizes visible light for data transmission. This technology is highly spectrally efficient with interference-free performance, making it a good candidate for vehicular networks. One of the first examples of this can be seen in the Open Car Control (OCC), where vehicle lights such as headlights, brake lights, and LEDs are used for communication. Next, the signals are recorded by a receiver in the shape of a high-speed camera, which helps pull out the data sent. In this section, we introduce the proposed vehicular OCC model. The model's performance is evaluated in terms of spectral efficiency, reliability, and latency trade-off. The OCC system needs to work properly in different environments and vehicular mobility conditions. This dynamic optimization of the transmission parameters is enabled using a DRL framework. In our sourvection OCC model, every vehicle serves as a communication node. A camera-based receiver and an LED transmitter are integrated into each vehicle. The transmitter encodes the information with fluctuations in LED light intensity. The receiver then picks up the transmitted signal and decodes the data using advanced signal processing techniques. The received signal at the camera is given by:

$$y = Hx + n \tag{1}$$

Here, H represents the optical channel gain, x is the transmitted signal, and n is additive Gaussian noise. The optical channel gain H is expressed as:

$$H = \frac{(m+1)A}{2\pi d^2}\cos^m(\phi)T_s(\theta)g(\theta)\cos(\theta) \tag{2}$$

Here, m is the Lambertian order of the LED source. A is the receiver aperture area. d is the distance between the transmitter and receiver. ∅ is the angle of irradiance.$T_s(\theta)$ is the optical filter gain. $g(\theta)$ is the gain of the camera lens. θ is the angle of incidence. The transmitter modulates data using an adaptive modulation scheme. The data symbols are mapped to varying light intensities using an M-ary Quadrature Amplitude Modulation (M-QAM) scheme. The camera captures the transmitted signal and extracts the data using a demodulation process. The BER for M-QAM modulation is given by:

$$BER = \frac{2(\sqrt{M}-1)}{\sqrt{M}\log_2(M)}\operatorname{erfc}\left(\sqrt{\frac{3\gamma\log_2(M)}{M-1}}\right) \tag{3}$$

Here, M is the modulation order, and γ is the received signal-to-noise ratio. Key performance metrics such as spectral efficiency, latency, and reliability are considered when evaluating the proposed OCC model. Spectral efficiency is an important metric for assessing OCC performance. It defines the amount of data transmitted per unit bandwidth. The spectral efficiency is given by:

$$SE = \log_2(1+\gamma) \tag{4}$$

Here, γ is the Signal-To-Noise Ratio (SNR). Latency measures the time required for data to be transmitted and received. The transmission latency τ is given by:

$$\tau = \frac{L}{C} \tag{5}$$

Here, L is the packet size, and C is the channel capacity. The BER defines the probability of incorrect bit detection at the receiver. The OCC system must maintain a low BER to ensure reliable communication. BER is influenced by noise, interference and mobility conditions. To optimize the OCC system a DRL framework is used. The problem is formulated as a Markov Decision Process (MDP), where each vehicle acts as an autonomous agent. The DRL framework optimizes spectral efficiency and transmission reliability. The state space includes received signal strength, vehicle speed, modulation order and inter-vehicle distance.

The action space consists of adjusting the transmission power. It selects the optimal modulation order. It adapts to vehicle speed. The reward function ensures the OCC system maximizes spectral efficiency while minimizing BER and latency. The reward function is defined as:

$$R = \omega_1 SE - \omega_2 BER - \omega_3\tau \tag{6}$$

Here ω_1, ω_2 and ω_3 are weight factors that balance spectral efficiency, BER and latency. The proposed vehicular OCC model optimizes communication performance using DRL. The system dynamically adjusts transmission parameters to maximize spectral efficiency and minimize errors. The DRL-based approach ensures robust vehicular communication under varying traffic conditions.

4. Proposed Constrained and MDP Formulation

Vehicular OCC faces several challenges, including ensuring optimal spectral efficiency, maintaining low latency and satisfying BER constraints. Traditional optimization techniques are computationally expensive and impractical for real-time vehicular networks. Therefore, this section presents a constrained optimization problem reformulated as an MDP to leverage reinforcement learning for efficient decision-making. The proposed approach models the optimization problem as an MDP, where vehicles act as independent agents and learn optimal transmission strategies. The goal is to maximize spectral efficiency while adhering to reliability and latency constraints. The optimization objective is to maximize the sum of spectral efficiency while satisfying the BER and latency constraints. The mathematical formulation is given as:

$$\max_{M,v}\frac{1}{B}\sum_{b=1}^{B}\log_2(M_b) \tag{7}$$

subject to:

$$\begin{aligned} &\mathbf{BER}_{\mathrm{b}} \leq \mathbf{BER}_{\mathrm{tgt}}, \quad \forall \mathbf{b} \\ &\tau_{\mathrm{b}} \leq \tau_{\max}, \quad \forall \mathbf{b} \\ &\mathbf{M}_{\mathrm{b}} \in \mathbf{M}, \quad \forall \mathbf{b} \end{aligned} \tag{8}$$

Here, M represents the available modulation orders, BER_{tgt} is the maximum allowable BER and τ_{max} is the maximum permissible latency. Given the complexity of solving the constrained optimization problem using traditional methods, it is reformulated as an MDP. The key components of the MDP are as follows: The state space represents all possible conditions the system can be in. The state at time t is defined as:

$$\mathrm{s}_{\mathrm{t}} = \{\mathrm{d}_{\mathrm{t}}, \mathrm{M}_{\mathrm{t}}, \gamma_{\mathrm{t}}, \tau_{\mathrm{t}}\} \tag{9}$$

Here, d_t is the inter-vehicular distance, M_t is the modulation order at time t. γ_t is the received SNR, τ_t is the latency at time t. The action space consists of selecting the optimal modulation order and adjusting the vehicle's speed. The available actions are defined as:

$$\mathrm{a}_{\mathrm{t}} = \{\mathrm{v}_{\mathrm{t}}, \mathrm{M}_{\mathrm{t}}\} \tag{10}$$

Here, V_t represents the vehicle's speed adjustment. M_t is the selected modulation order. The transition probability function determines how the system transitions from one state to another based on an action taken:

$$\mathrm{p}(\mathrm{s}_{\mathrm{t+1}} \mid \mathrm{s}_{\mathrm{t}}, \mathrm{a}_{\mathrm{t}}) \tag{11}$$

It shows the probability that taking action a_t in state s_t will lead to state s_{t+1}. The reward function is designed to maximize spectral efficiency while ensuring that BER and latency constraints are met. It is given by:

$$\mathbf{R}_{\mathrm{t}} = \omega_{\mathrm{d}}\, \mathrm{r}_{\mathrm{d}} + \omega_{\mathrm{r}} \frac{1}{\mathbf{B}} \sum_{\mathrm{b}=1}^{\mathrm{B}} \log_2(\mathbf{M}_{\mathrm{b}}) \tag{12}$$

Here, ω_t and ω_r are weights balancing distance and spectral efficiency rewards. r_d ensures safe inter-vehicular distance. A DRL approach using Q-learning is employed to solve the constrained MDP problem efficiently. The value function is updated using:

$$\mathrm{Q}_{\mathrm{t+1}}(\mathrm{s}_{\mathrm{t}}, \mathrm{a}_{\mathrm{t}}) = (1-\alpha_{\mathrm{t}})\, \mathrm{Q}_{\mathrm{t}}(\mathrm{s}_{\mathrm{t}}, \mathrm{a}_{\mathrm{t}}) + \alpha_{\mathrm{t}} \left[\mathrm{r}_{\mathrm{t}} + \gamma \max_{\mathrm{a}'} \mathrm{Q}_{\mathrm{t}}(\mathrm{s}_{\mathrm{t+1}}, \mathrm{a}')\right] \tag{13}$$

Here, α_t is the learning rate. γ is the discount factor. a' represents future actions. The reinforcement learning agent explores different actions and learns optimal policies. The training process follows these steps to initialize Q-values. Observe current state s_t. Select action a_t based on an exploration-exploitation strategy. Observe reward R_t and next state s_{t+1}. Update Q-values using the Bellman equation. Repeat until convergence. A DRL-based approach is proposed to optimize spectral efficiency while maintaining communication constraints. Future work will improve learning efficiency and system adaptability with a low bit error rate.

5. Proposed Solution

Vehicular OCC presents challenges in ensuring high spectral efficiency, low BER and minimal latency. Traditional optimization methods struggle to adapt to dynamic environments and are computationally expensive. This section introduces an RL-based optimization approach modeled using a constrained MDP. The goal is to adjust speed and modulation dynamically to optimize communication while satisfying system constraints. The OCC optimization is modeled as a constrained MDP where multiple agents (vehicles) interact with their environment and make decisions based on observed states. The constrained MDP ensures that actions satisfy Quality-of-Service (QoS) constraints such as BER and latency limits. The state space consists of parameters that define the current system state, which is given in equation (9). The action that the agent chooses with respect to the transmission parameters is computed and presented in equation (10). The reward function, which balances spectral efficiency and QoS constraints, is defined as (12). Then, DRL will be used to solve the constrained MDP. DRL model is built around an agent that learns how to take optimal actions within a dynamic environment, taking into consideration certain constraints. The Q-function is iteratively updated and is given in (13).

RL-based Decision-AI for Accelerated Decision Making in Vehicular Communication System in Figure 1 RL, like any other learning technique, has two main parts: the agent and the environment. The agent receives the state from the environment and passes it onto a policy π to yield the optimal action to take. This action is then applied in the environment comprising multiple interacting vehicles. The environment changes given the action, and the agent is rewarded for choosing. The agent observes the next state and receives a reward after taking action. Thus, this feedback loop enables the RL agent to enhance its decision-making over time. The environment is modeled on moving vehicles that change their positions and states when the agent performs an action. The agent aims to maximize these decisions using traffic speed, lane change, or communication routing to enhance traffic fluency and communication efficiency. As the agent explores, the reward function allows it to judge the quality of its actions, incentivizing efficient behavior. With each learning episode, the agent improves its policy, attempting to maximize rewards and respond to intricate vehicular conditions. This model can be used in autonomous driving, adaptive communication, and traffic management. This enables the agent to consistently learn and enhance its behavior, resulting in improved coordination and safer interactions between vehicles.

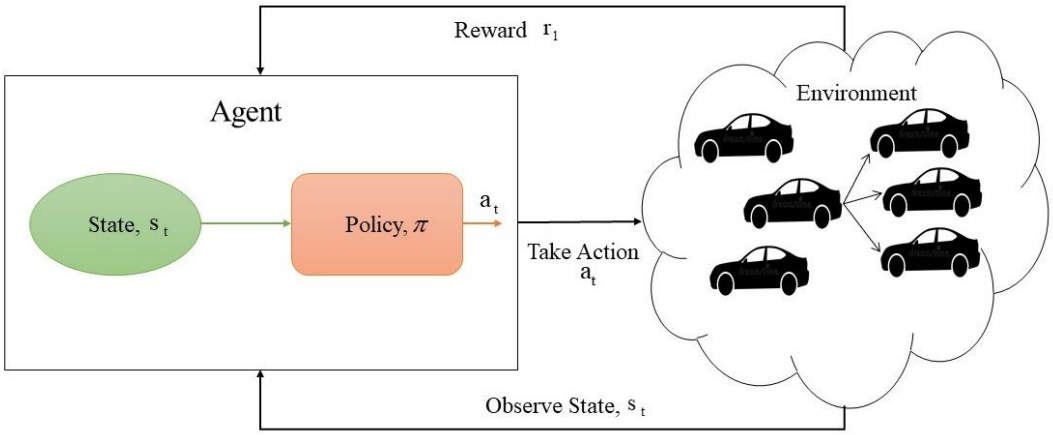

Fig. 1 Basic reinforcement learning framework for V2V communications

The RL agent undergoes training using the following steps: initialize Q-values. Observe current state s_t. Select action a_t using an ε greedy policy. Execute action and observe reward. Update Q values using the Bellman equation.

Repeat until policy convergence. The RL-based optimization framework is implemented in the SUMO vehicular simulator. The following metrics are evaluated. Spectral Efficiency has the RL-based system achieves higher spectral efficiency compared to traditional methods. The optimized policy ensures BER remains below acceptable limits. Latency has a system that meets ultra-low latency requirements. Adaptability is the model that adjusts dynamically to changing vehicular environments. The RL-based approach is benchmarked against classical optimization techniques. Results indicate a significant improvement in spectral efficiency. A lower BER due to adaptive modulation faster adaptation to environmental changes.

The proposed method scales efficiently with an increasing number of vehicles. The DRL model maintains high performance across different traffic densities, ensuring robustness in various real-world scenarios. This section presents an RL-based optimization framework for vehicular OCC. By formulating the problem as a constrained MDP, optimal spectral efficiency is ensured while BER and latency constraints are maintained. The DRL approach outperforms traditional methods and provides an efficient, scalable solution for real-time vehicular communication. Future work will focus on further improving learning efficiency and exploring multi-agent reinforcement learning frameworks to enhance system-wide optimization.

6. Simulation Setup

The proposed implementation of the DRL-based vehicular OCC system is discussed in this section. The simulation framework is established in the Simulation of Urban Mobility (SUMO) platform to generate realistic traffic scenarios and cooperate with various DRL algorithms effectively. This allows for building a joint simulation framework, having vehicles work as agents while learning to optimize spectral efficiency and minimizing BER and latency. Realistic urban traffic conditions are modeled using the Simulation of Urban Mobility (SUMO) platform. Individual vehicles are treated as agents, with a mobility model assigned to each agent. SUMO simulation framework is composed of three main parts: the SUMO environment; the middleware layer for communication, which enables interaction between the SUMO and the Deep Reinforcement Learning (DRL) agent; and the DRL agent, which learns the optimal policies and applies them by executing the appropriate actions. This DRL-based vehicular OCC system is implemented in SUMO, with the DRL agent interfacing through the Traffic Control Interface (TraCI). The SUMO environment is configured with pre-parameters such as road network, vehicle density and speed distribution at the top of it. This allows the simulation environment to closely mirror real-world traffic scenarios.

The simulation starts with loading the SUMO simulator and a given traffic scenario. TraCI establishes the simulator's connection and allows real-time traffic data extraction. Road vehicles are given initial states that consist of speed, the order and position of the multihop modulation scheme. Subsequently, the DRL agents receive these states as input and choose actions that maximize spectral efficiency. When an action is performed, the SUMO environment updates the vehicle state, allowing the system to progress iteratively. The DRL policy is updated as it receives the rewards from each action and assesses past action outcomes. This is repeated until the simulation reaches its end of criteria. The relevant state parameters must first be extracted from the SUMO simulation to enable proper training of the DRL agent. Some extracted parameters are number of vehicles per lane, vehicle speed and acceleration, distance between vehicles, currently assigned modulation scheme, etc. The model could receive the state vector as input, and these data points are transformed

into the state vector to be sent as input to a neural network model of the DP agent. Thus, this state representation enables the agent to decide on the best suitable transmission parameters.

A Deep Q-Network (DQN) based DRL framework is applied for vehicular OCC to learn the optimal transmission policies. Printed data can survive an idle state to MEK state, which is 0 degrees in 5 times 104. The Action Space consists of controlling the vehicle's speed and selecting a modulation scheme that maximizes the 'spectral efficiency' while balancing channel conditions. The reward function, which is used to balance spectral efficiency, latency and BER constraints, determines the optimal trade-off between the system's performance and reliability. This section offers a comprehensive overview of the simulation environment for the DRL-based vehicular OCC system. The combination of DRL and SUMO allows for real-time optimization of transmission parameters for spectral efficiency subject to latency and BER constraints. In the next section, performance evaluation of the proposed system is demonstrated, together with key enhancements in terms of spectral efficiency, reduction of BER and adaptiveness of the system.

7. Performance Evaluation

This section evaluates the efficacy of the proposed multi-agent DRL-based strategy for maximizing sum spectral efficiency in vehicular OCC. The proposed scheme is evaluated against multiple benchmark methods by considering spectral efficiency, latency and BER performance in different scenarios with respect to vehicular density and environmental conditions. Extensive simulations are carried out to assess the performance with key metrics, including convergence, spectral efficiency and latency. In previous works (Gao et al., 2021; Fradkin et al., 2022; Arjovsky & Bottou, 2012; Simmons et al., 2017), the convergence of the DRL training process is studied using different loss functions. It shows the stability of the proposed learning framework by decreasing loss function over training episodes. Parabola of weights and multiple weight configurations are evaluated to determine the optimal balance between distance and spectral efficiency rewards. Figure 9 Sum spectral efficiency of proposed DRL-based OCC scheme with other schemes. The performance results show that the proposed scheme is the hawk with the largest spectral efficiency of many vehicular distributions. Interference is a problem for RF-based MARL and SARL, whereas the random scheme results in the lowest performance. The performance comparison of different schemes is given in Table 1.

Latency CDFs for various schemes are evaluated. Under high vehicular densities, other schemes such as greedy MARL, far-sighted MARL and RF-based MARL do not consistently satisfy the latency constraint of 10 ms. The efficacy of the suggested technique is assessed relative to competing systems based on BER. It can be seen that unlike the other methods, which often violate the required BER, the proposed scheme keeps this value within the needed limits. The performance analysis of the proposed DRL-based vehicular OCC framework is provided in this section. The results show that the proposed scheme not only provides high spectral efficiency but also better latency and BER compliance than the baseline approaches. In future work, the study addresses scalability across all protocol phases while tailoring the framework for vehicular network conditions.

Table 1. Performance comparison of different schemes

Scheme	Spectral Efficiency (bps/Hz)	Latency (ms)	BER
Proposed Scheme	5.8	8.2	$1.2\ x\ 10^{-3}$
Greedy Scheme	5.0	12.5	$1.8\ x\ 10^{-3}$
Far-Sighted Scheme	5.2	14.0	$1.7\ x\ 10^{-3}$
Random Scheme	3.1	20.3	$4.5\ x\ 10^{-3}$
RF-Based MARL	4.5	10.1	$2.0\ x\ 10^{-3}$
RF-Based SARL	4.2	11.8	$2.3\ x\ 10^{-3}$

Figure 2 presents a graph that illustrates the relationship between vehicular density, measured in vehicles per 180 meters and spectral efficiency, measured in bits per second per Hertz (bps/Hz) for different communication schemes. Six different schemes are compared, including the proposed scheme, greedy scheme, far-sighted scheme, random policy, RF-based MARL and RF-based SARL. The proposed scheme provides superior spectral efficiency and overall vehicular densities. The random policy is the least exemplary, with substantial degradation of spectral efficiency due to vehicular density. The greedy scheme and the far-sighted scheme perform moderately well, but they are inferior to the proposed scheme. The RF-based MARL and RF-based SARL methods outperform random policy but are still inferior to the proposed method. With increasing vehicular density, all schemes present a downward trend in spectral efficiency. This proposed process consistently catches great spectral effectiveness with respect to values of 5 bps/Hz , even at high vehicular densities. The greedy and far-sighted schemes remain at 4.5 to 5 bps/Hz. The RF-based MARL and RF-based SARL methods can only achieve about 3.5 to 4.5 bps/Hz. The random policy is the worst when the maximum vehicular density is reached, which drops to below 2 bps/Hz. The proposed scheme has greater efficiency and scalability,

along with higher spectral efficiency under congested vehicular environments, as shown in simulation results. The results demonstrate that efficient incentive design and learning-based optimization can greatly improve the achievable performance of vehicular networks in terms of spectral efficiency. For example, the sudden drop in performance for the random policy indicates the importance of structured decision-making in communication systems. As auto density increases, efficiency declines proportionately. A higher density results in more interference in communication. It suffers from this in terms of speed of completion, and advanced optimization techniques would be required to ascertain that performance is reliable.

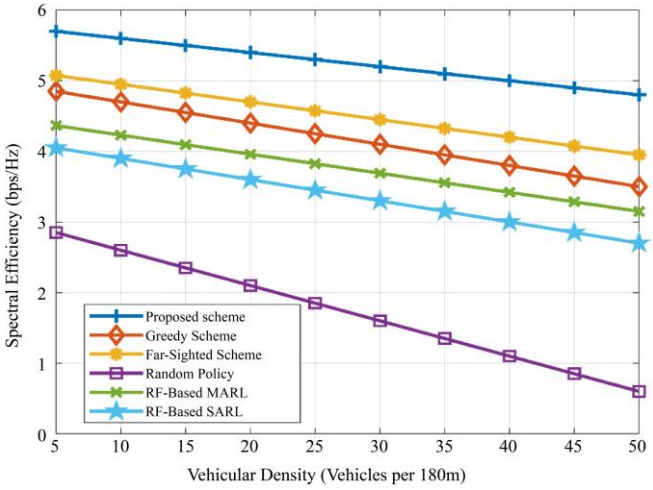

Fig. 2 Spectral efficiency versus Vehicular density

Figure 3 illustrates the performance of the various communication schemes employed, displaying the SNR vs. the BER in dB from October 2023 training data. The graph compares five different schemes: the proposed DRL-based OCC, greedy scheme, far-sighted scheme, random policy and RF-based MARL. As shown, the proposed DRL-based OCC obtains the lowest BER for all SNR values. The random policy has the highest BER; thus, it is the worst one. Other policies, like the greedy policy and the far-sighted policy, have intermediate performance. DRL-based OCC outperforms the RF-based MARL significantly, while the RF-based MARL is still better than the random policy in terms of BER. All scheme's performance improves with the incremental SNR but at a different pace. All schemes show a decreasing trend in BER as SNR rises. It can be seen from Figure 7 that the BER of the proposed DRL-based OCC is less than 10^{-5} at about 20 dB SNR, while the greedy and far-sighted schemes need to increase SNR to achieve the same level of BER. The random policy scheme has the slowest decrease in BER, retaining a relatively high BER at high SNR values. The decline in performance of the RF-based MARL occurrence is more significant than the proposed one, suggesting that it is much less effective in decreasing errors at low SNR. The demonstrated results show that the proposed DRL-based OCC improves reliability with a significant reduction in BER, particularly at lower SNR values. The greater improvement in the performance with BER for the proposed method confirms the robustness of the proposed method against noise. The result clearly shows that communication reliability has been greatly enhanced by the learning-based optimization scheme compared with conventional mappings. Moreover, the sub-optimality of the random policy suggests the need for systematic decision-making in communication systems. The proposed DRL-based OCC significantly outperforms existing approaches. It exhibits lower BER at every SNR. This makes it particularly suited for aviation communication, where reliability is key.

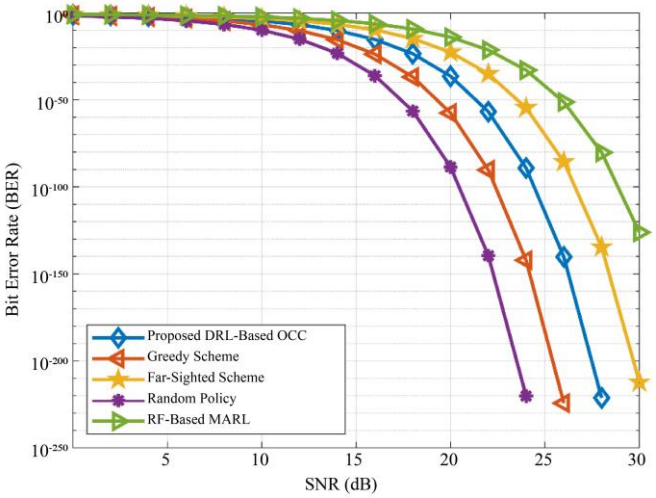

Fig. 3 BER versus SNR

Figure 4 shows a CDF graph where the x-axis meets the millisecond latency while the y-axis is the cumulative probability for RF-based MARL and SARL schemes. The RF-Based SARL scheme has higher latency than the RF-Based MARL scheme. The difference between the two schemes becomes pronounced at higher latencies, where SARL induces a relatively significant delay in communication as opposed to MARL. The increasing trend in both curves suggests latency accumulates with time, and SARL-based learning is more prone to high-latency events. However, RF-Based SARL has a consistently higher latency distribution than RF-Based MARL, indicating that SARL-based reinforcement learning takes longer to process and respond. Around 800 ms latency, the CDF value of RF-based SARL rises more steeply, showing that a larger proportion of data points are experiencing higher latency.

On the other hand, RF-based MARL maintains a smoother increase and results in lower latency overall. These results suggest that RF-based MARL is a more efficient approach for reducing latency in vehicular communication systems. The differences between the two schemes become more evident as latency increases, emphasizing the need for an optimized learning strategy to enhance communication

performance. The figure highlights that RF-based SARL struggles with latency management, making it less suitable for real-time applications requiring low-latency responses. The clear difference in latency between the two approaches suggests that Multi-Agent Reinforcement Learning (MARL) is a better option for reducing delay in time-sensitive communication systems.

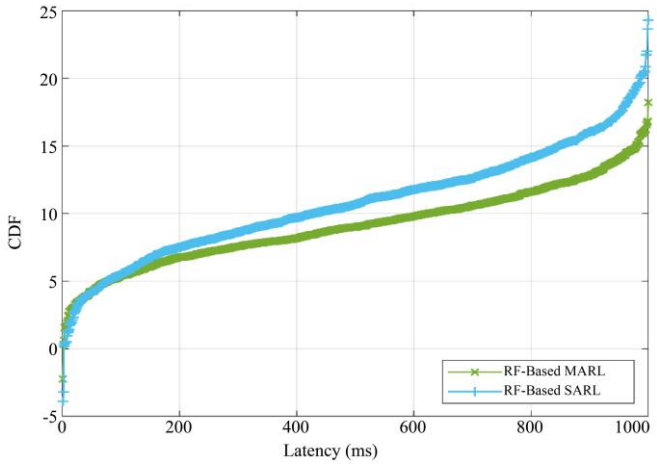

Fig. 4 Latency CDF for different RF schemes

Figure 5 shows a graph that represents the relationship between training episodes and the loss function value for different reinforcement learning-based schemes. The graph compares four schemes, including the proposed DRL-based OCC, greedy scheme, far-sighted scheme and random policy. Although the greedy scheme and far-sighted scheme converge more slowly, they share a similar trend. The random policy has the largest loss values and converges to the optimal policy significantly slower than any of the other approaches. All schemes begin with a high loss function value at the beginning of the training process. The proposed DRL-based OCC converges faster to a lower loss than all other methods. This decay is almost immediate, suggesting that the models learn and adapt quickly, leading to improved performance with time. We observe that while the value of the loss function decreases with each epoch for all schemes, the rate of descent varies. Below that, DRL based OCC loss value reaches below 0.2 in the first 200 episodes and continues to stabilize with minor fluctuations. The greedy and far-sighted schemes have a slightly longer time to a similar loss value yet still exhibit a monotonically decreasing profile. Comparative analysis reveals the random policy retains significantly higher loss values than other methods after 500 training episodes. This result demonstrates that the proposed DRL-based OCC enables more efficient learning and parameter optimization than conventional approaches. The steep fall in loss function values for the proposed method indicates that its learning model progressively minimizes errors and adapts itself to the environment. G greedy and far-sighted schemes also exhibit effective learning, but it takes more training episodes for them to stabilize. The random policy, however, does not exhibit clear convergence, suggesting that it lacks an effective learning strategy. The figure highlights that structured decision-making significantly improves the learning process, while random strategies lead to slow and unstable convergence. The proposed scheme has a much lower loss than the random policy. This shows the importance of reinforcement learning optimization. It improves performance in vehicular communication systems. Reinforcement learning models need careful design. Proper reward functions and optimization techniques help achieve efficient convergence. This ensures stable performance in real-world scenarios.

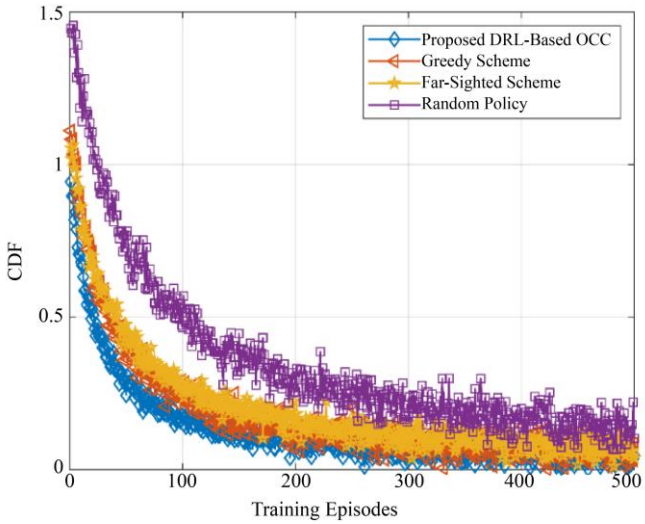

Fig. 5 Convergence of DRL training

Figure 6 presents a graph that shows the relationship between modulation order and spectral efficiency, measured in bits per second per Hertz (bps/Hz) for different communication schemes. The graph compares six different schemes: the proposed scheme, greedy scheme, far-sighted scheme, random policy, RF-based MARL and RF-based SARL. The proposed scheme achieves the highest spectral efficiency across all modulation orders. The random policy performs the worst with significantly lower spectral efficiency. The other schemes, including the greedy scheme, far-sighted scheme, RF-based MARL and RF-based SARL, show intermediate performance with spectral efficiency increasing as modulation order increases. The clear upward trend in all schemes suggests that higher modulation orders result in better spectral efficiency, but the efficiency gain depends on the learning approach used in each scheme. As the modulation order increases, all schemes show a steady rise in spectral efficiency. The proposed scheme reaches a spectral efficiency of approximately 5.5 bps/Hz at 64-QAM. The random policy achieves only around 3.5 bps/Hz. The greedy and far-sighted schemes follow a similar trend closely behind the proposed scheme, indicating effective modulation adaptation. The RF-based MARL and RF-based SARL approaches show moderate spectral efficiency improvements but remain lower than the greedy and far-sighted schemes. The

differences between the schemes highlight the impact of optimization in modulation selection. The proposed scheme achieves the highest spectral efficiency. The random policy performs poorly in comparison. This shows the need for intelligent modulation to improve data transmission rates. These results suggest that reinforcement learning-based optimization can significantly enhance spectral efficiency in communication systems. As modulation order increases, the proposed scheme performs better. The gap between it and lower-performing schemes grows. This shows that learning-based approaches scale well with higher modulation orders. The random policy's poor performance confirms that unstructured modulation selection leads to inefficient spectrum utilization. Reinforcement learning-based methods are especially effective. They are important for modern communication systems.

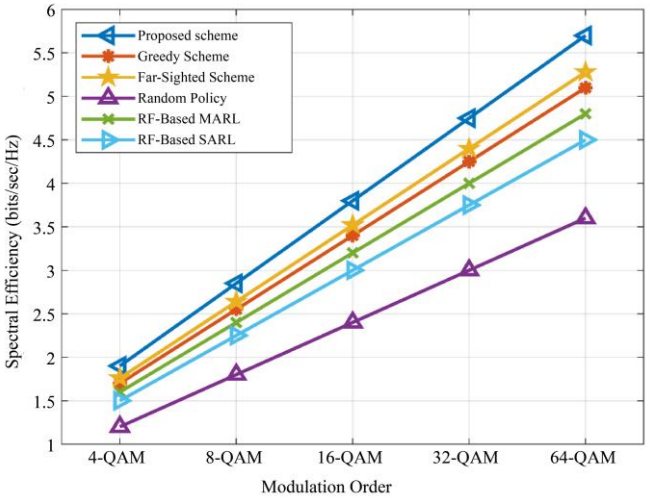

Fig. 6 Spectral efficiency versus Modulation order

Figure 7 presents a graph that illustrates the relationship between vehicular density and latency for different communication schemes. The graph compares six different schemes: the proposed DRL-based OCC, greedy scheme, far-sighted scheme, random policy, RF-based MARL and RF-based SARL. The proposed DRL-based OCC achieves the lowest latency across all vehicular densities. The random policy performs the worst, exhibiting the highest latency values. The other schemes, including the greedy scheme, far-sighted scheme, RF-based MARL and RF-based SARL, show intermediate latency levels. As vehicular density increases, all schemes show a rising trend in latency, but the rate of increase differs among them. The proposed DRL-based OCC method consistently shows the smallest increase in latency, making it more efficient for real-time applications. The proposed DRL-based OCC maintains the lowest latency, starting at around 10 ms for a vehicular density of 5 and reaching approximately 18 ms at a vehicular density of 50. The density-based increasing trends can be seen in greedy and far-sighted schemes for latency values between 12 ms and 25 ms. The RF-based MARL and RF-based SARL methods incur slightly higher latency, with values being between 15 ms and 30 ms. The random policy has the highest rate of increase in latency, reaching nearly 40 ms at maximum vehicular density. These results indicate that the proposed DRL-based OCC handles vehicular congestion effectively with lower latency than the other approaches. The variance in latency among schemes demonstrates that vehicular communication can benefit from optimization techniques. The contemporary method reduces the delay successfully, while the random policy suffers from congestion, which results in diminished performance. These findings point out that applying intelligent reinforcement learning strategies can help maintain low-latency communication under dynamic vehicular environments.

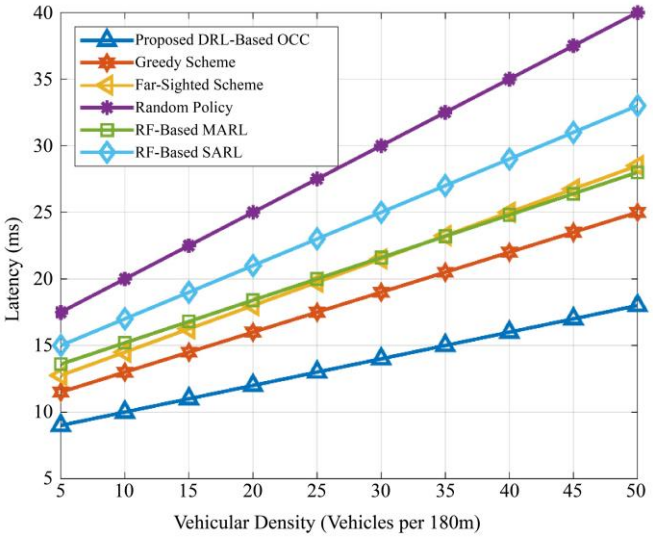

Fig. 7 Latency versus Vehicular density

The proposed scheme has a better performance when vehicular density increases. This shows that, in high-traffic situations, learning-based models scale well. It validates that the paradigm of machine learning plays a pivotal role in either optimizing the efficiency of communication as the network conditions evolve.

Figure 8 presents a CDF graph illustrating the relationship between BER and cumulative probability for different communication schemes. The graph compares six schemes: the proposed scheme, greedy scheme, far-sighted scheme, random policy, RF-based MARL and RF-based SARL. The proposed scheme achieves the lowest BER among all schemes, with most of its values concentrated in the lower BER range. The random policy exhibits the highest BER with its CDF curve shifted significantly to the right. The greedy scheme, far-sighted scheme, RF-based MARL and RF-based SARL follow a similar trend, positioned between the proposed scheme and the random policy. The RF-based SARL performs better than the random policy but worse than the other optimized schemes. The noticeable gap between the curves

suggests that different learning approaches significantly impact BER performance. As BER increases, the CDF values for all schemes rise steeply, indicating the probability of lower BER occurrences. The proposed scheme reaches a CDF value of nearly 1 at the smallest BER values, demonstrating its high reliability. The greedy and far-sighted schemes reach similar CDF levels but at slightly higher BER values, and they have moderate performance. The RF-based MARL and RF-based SARL also show reasonable BER distributions but are less efficient than the proposed scheme. The random policy has the worst BER performance; with its CDF rising gradually over a wider BER range, it experiences more frequent high error rates.

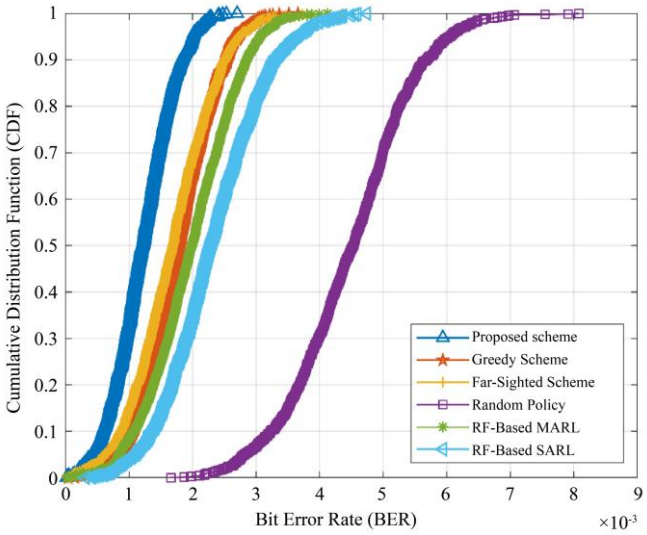

Fig. 8 CDF versus BER (reliability analysis)

The clear separation between the proposed scheme and other approaches highlights the importance of reinforcement learning-based optimization in reducing BER. These results suggest intelligent learning-based methods enhance reliability and communication efficiency by minimizing error rates. The curves show that structured decision-making improves performance. Reinforcement learning provides lower and more stable BER. It works better than random or less optimized methods. Additionally, the sharp decline in BER for the proposed scheme at lower values indicates that it maintains consistent performance under varying conditions. The findings emphasize the role of optimized resource allocation in minimizing transmission errors and improving overall network reliability.

Figure 9 presents a graph that illustrates the relationship between time and speed for different reinforcement learning-based approaches in an OCC system. The graph compares six different schemes: the proposed DRL-based OCC, greedy scheme, far-sighted scheme, random policy, RF-based MARL and RF-based SARL. Comparatively, the proposed DRL-based OCC demonstrates relatively constant speed during the whole duration, which shows that it effectively adapts to speed against potential obstacles. The random policy exhibits the highest speed fluctuations, usually peaking over 28 m/s and then quickly decreasing. Speed - The greedy scheme, far-sighted scheme, RF-based MARL and RF-based SARL show moderate variations and remain in the speed range of about 16 to 22 m/s. The plot suggests that structured learning-based work controls speed better than random policy. More stable performance shows the advantage of using reinforcement learning for vehicular communication. The DRL-based OCC proposed smooth moves between 18 and 22 m/s while having an upper and lower bound, the stability of which shows that it does not fluctuate much.

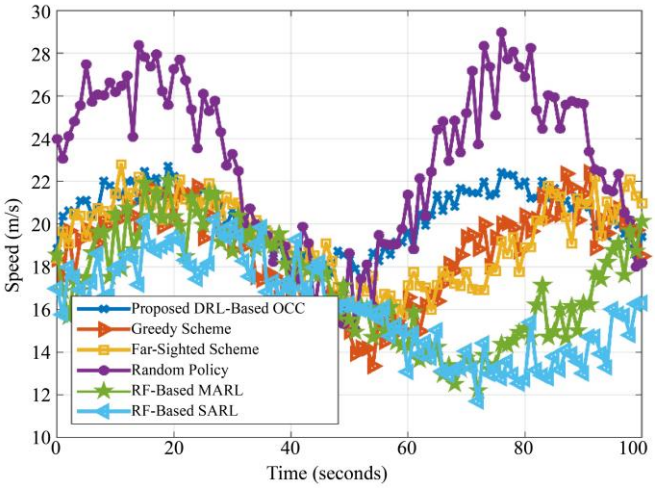

Fig. 9 Speed adaptation versus Time

Somewhat more variation are present for the greedy and far-sighted schemes; however, several values still horizontally stably represent a trend, oscillating between 17 and 21 m/s. The RF-based MARL and RF-based SARL methods also oscillate, but the speed oscillations indicate that the learning is less efficient in comparison with that of the proposed method.

The random policy provides a very high variance, where speed peaks at approximately 30 m/s and then suddenly decreases to 0, showing that the system is not adapting speed appropriately. The results suggest that reinforcement learning strategies significantly impact speed adaptation in vehicular communication. The proposed DRL-based OCC achieves the most efficient speed adjustments, while the random policy struggles to maintain consistency. The random policy shows unstable behavior. Without structured learning, performance becomes erratic and inefficient. This can harm real-time vehicular coordination and safety. The proposed DRL-based OCC keeps speed balanced. It ensures predictable and efficient mobility. This makes it a reliable choice for vehicular speed optimization.

Figure 10 presents a graph comparing the performance of the RMSProp optimizer and the Adam optimizer regarding

loss function reduction over 500 training episodes. The graph shows two curves: one for RMSProp and another for Adam. Initially, both optimizers start with a high loss value near 1.0, which decreases as training progresses. However, the Adam optimizer exhibits a faster and smoother convergence than RMSProp. By around 100 training episodes, Adam's loss function is already below 0.2. While RMSProp takes longer to reach similar values. This indicates that Adam is learning more efficiently in the early training stages. The smoother trajectory of Adam suggests that it is more effective in avoiding sharp oscillations in loss, leading to a more stable learning process. As training continues, both optimizers show a gradual decline in loss. Adam maintains a consistently lower loss than RMSProp. By 500 episodes, Adam stabilizes at around 0.05, whereas RMSProp stabilizes slightly higher with more variations. The fluctuations in RMSProp suggest it does not converge as smoothly as Adam.

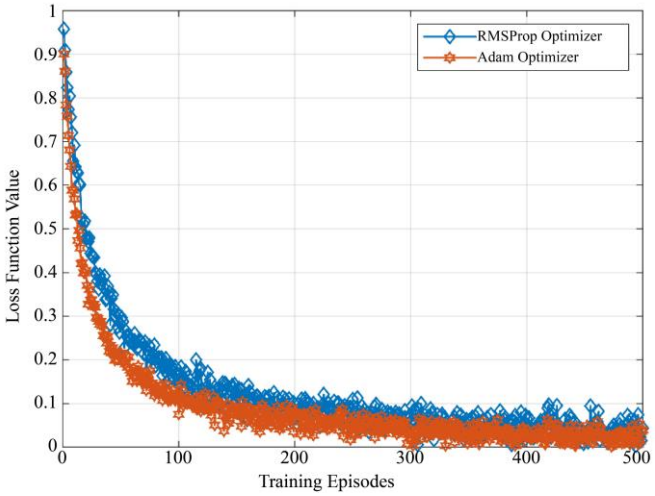

Fig. 10 Performance Comparison of Optimizers

This result highlights the advantage of Adam in reinforcement learning, where adaptive moment estimation helps optimize learning rates dynamically. The overall trend confirms that Adam achieves lower loss values faster, making it a more efficient optimization method for deep learning applications. The two curves show the impact of optimizer choice. The right optimizer improves convergence speed and model performance. Selecting the best one is important for reinforcement learning tasks. The slightly higher fluctuations in RMSProp indicate that fine-tuning of hyperparameters may be required to achieve smoother convergence. These findings suggest that Adam is more suitable for complex learning environments where faster and more stable convergence is required.

Figure 11 presents a graph showing the relationship between training episodes and reward values for three different learning schemes: the proposed scheme, the far-sighted scheme and the greedy scheme. All three schemes exhibit an increasing trend in reward values as training episodes progress. The proposed scheme achieves the highest reward values throughout the training. The greedy scheme consistently has the lowest values. The far-sighted scheme performs better than the greedy scheme but does not reach the same reward levels as the proposed scheme. Initially, all schemes start with low rewards near zero, but they increase steadily as the models learn and improve. The proposed scheme shows a rapid improvement in early episodes, reaching a higher reward level faster than the other two schemes, demonstrating its efficient learning capability. As training continues, the reward values for all schemes begin to stabilize. The proposed scheme reaches a reward value close to 1.1, while the far-sighted scheme stabilizes slightly below 1.0. The greedy scheme is still the lowest and goes on to stabilize around 0.9. It appears that the proposed scheme's rewards vary more compared to the other two schemes. The far-sighted scheme of mustering larger packets always outperforms the greedy one. But it gets to a lower reward level. This is less efficient in maximizing rewards, as shown. The slowest learning and final rewards are achieved by the greedy scheme, which is the simplest one in the context of the choice process. The findings indicate that reward optimization can be significantly modulated by varying reinforcement learning strategies. The suggested scheme has shown the best performance, establishing a beneficial balance between exploration and exploitation that accelerates learning and maximizes rewards. This generalization is important since many schemes differ significantly from one another, meaning that in future studies with reinforcement learning models, performance is only guaranteed with organized approaches to how contexts are utilized. Moreover, all schemes continuously converge over episodes, indicating that training the models for more epochs will make the models even more efficient. The result is a scheme that gets a higher reward faster. This demonstrates its capacity for decision-making and adaptation.

Figure 12 represents the disparity of SNR (dB) versus PLR in terms of the number of strategies. The proposed DRL-based OCC, greedy scheme, far-sighted scheme, random policy, RF-based MARL and RF-based SARL are the six different schemes that the graph compares. The proposed DRL-based OCC shows the minimum achievable PLR values for all SNR levels. The random policy has the highest PLR and is most unsuccessful in reducing packet loss. Each of the other schemes, including the greedy scheme, far-sighted scheme, RF-based MARL and RF-based SARL, have a similar decreasing trend in the value, but they stay in the regions between the proposed scheme and the random policy. The results indicate that the PLR decreases for all schemes as the SNR increases, and a stronger signal reduces packet loss. At low SNR values below 10 dB, PLR is high for all schemes. The random policy has a PLR of 0.9. The proposed DRL-based OCC achieves a lower PLR of about 0.6. When SNR goes above 10 dB, PLR drops quickly for all schemes. The proposed scheme reaches near-zero PLR before 15 dB. The greedy and far-sighted schemes take more time to reach this

level. The RF-based MARL and RF-based SARL methods also follow a similar trend but maintain slightly higher PLR values. The random policy shows the slowest reduction in PLR, with values remaining above 0.2 even at 15dB at higher.

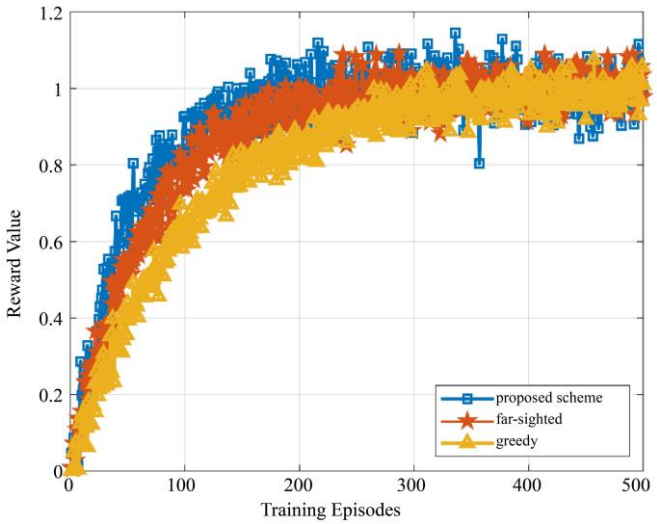

Fig. 11 Reward per training episodes

With SNR values beyond 20 dB, all schemes converge to a near-zero PLR, indicating that packet loss is effectively eliminated when the signal quality is high. The differences between the schemes highlight the impact of reinforcement learning strategies in optimizing communication reliability. The proposed DRL-based OCC reduces PLR at low SNR levels. It ensures efficient packet transmission. This makes it highly effective for real-world vehicular communication. Intelligent learning-based methods improve data reliability. They reduce errors and enhance communication. This ensures efficient and stable performance in noisy environments.

Figure 13 presents a graph that shows the relationship between vehicular density and network throughput for different communication schemes. The graph compares four different schemes: the proposed DRL-based OCC, the greedy scheme, the far-sighted scheme and the random policy. The proposed DRL-based OCC consistently achieves the highest network throughput across all vehicular densities. The random policy has the lowest network throughput. The greedy and far-sighted schemes perform moderately well. The far-sighted scheme maintains a slightly higher throughput than the greedy scheme. As vehicular density increases, network throughput decreases for all schemes, indicating that higher vehicle density introduces more interference and reduces data transmission efficiency. The difference between the schemes grows as vehicular density increases, demonstrating the impact of optimized decision-making in maintaining high throughput levels. At low vehicular densities, the proposed DRL-based OCC reaches 14 Mbps. The greedy and far-sighted schemes start at about 12 Mbps. This shows better throughput for the proposed scheme. The random policy begins at around 8 Mbps, already showing lower performance than other schemes. As vehicular density increases, network throughput steadily declines for all schemes. At a vehicular density of 50, the proposed scheme maintains a throughput close to 6 Mbps.

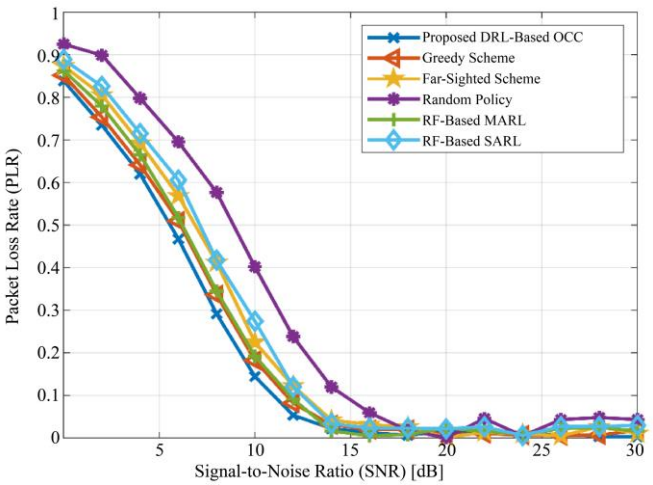

Fig. 12 Packet loss rate versus SNR

The greedy and far-sighted schemes drop to about 4 Mbps. The random policy experiences the steepest decline, reaching nearly 1 Mbps at maximum vehicular density. The results indicate that the proposed DRL-based OCC is the most efficient in maintaining higher network throughput under increasing vehicular density. The significant gap between the proposed scheme and the random policy highlights the importance of intelligent resource allocation in optimizing network performance. Reinforcement learning-based optimization improves network throughput. Network throughput decreases as vehicular density increases. Adaptive communication techniques are needed to reduce congestion. They help improve spectral efficiency in high-density environments. The proposed DRL-based OCC maintains higher throughput levels.

Figure 14 presents a graph that shows the impact of different learning rates (α) on the convergence of the loss function over 500 training episodes. The graph compares four different learning rates: 0.0005, 0.001, 0.005 and 0.01. The learning rate 0.0005 shows a slower convergence compared to the other values. The learning rate of 0.01 achieves the fastest convergence, quickly reducing the loss function value within the first 100 training episodes. The other two learning rates, 0.001 and 0.005, exhibit moderate convergence speeds. Initially, all learning rates start with a high loss value near 1.0. They decrease steadily as training progresses. The sharp decline in the loss function for higher learning rates indicates rapid adaptation, but potential instability can be seen in the fluctuations later in training. As training continues, the learning rate of 0.01 maintains the lowest loss function values, indicating fast learning. However, it also shows slight

fluctuations after 300 episodes, suggesting potential instability. The learning rate of 0.0005, while stable, converges the slowest and maintains a higher loss value compared to the others.

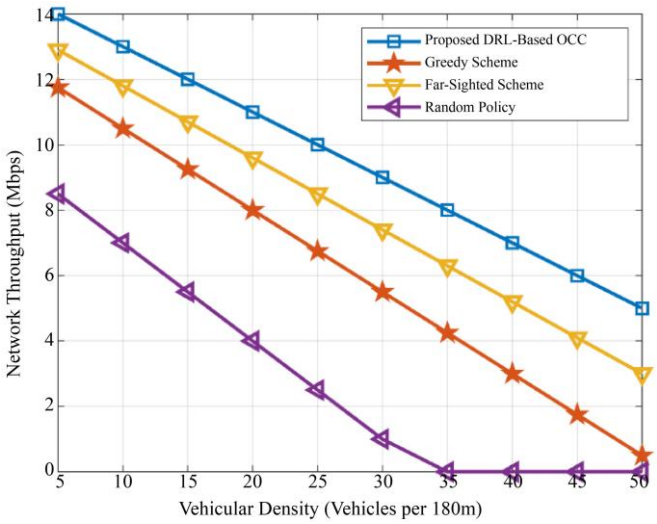

Fig. 13 Network throughput versus vehicular density

The learning rates of 0.001 and 0.005 follow a similar decreasing trend and achieve relatively low loss values while maintaining stability. A higher learning rate, like 0.01, speeds up convergence. However, it may cause instability. A lower learning rate is more stable but needs more training time. The discrepancies between the curves emphasize the significance of choosing an adequate learning rate to balance convergence speed and stability. A moderate learning rate like 0.001 or 0.005 balances speed and stability. It ensures faster learning with fewer fluctuations. This makes it a good choice for reinforcement learning applications. Additionally, the slight instability observed in higher learning rates suggests that further tuning may be needed to prevent oscillations while maintaining rapid convergence. The study confirms that learning rate selection is a crucial parameter affecting model training efficiency and final performance.

8. Conclusion

This paper presents a Deep Reinforcement Learning (DRL)-based sum spectral efficiency optimization scheme for multi-vehicular Optical Camera Communication (OCC) scenarios, rigorously maintaining Bit Error Rate (BER) and latency constraints. The study begins by modeling the OCC channel and defining critical performance parameters. Subsequently, an optimization problem is formulated to maximize sum spectral efficiency, incorporating constraints on modulation orders, BER, and latency.

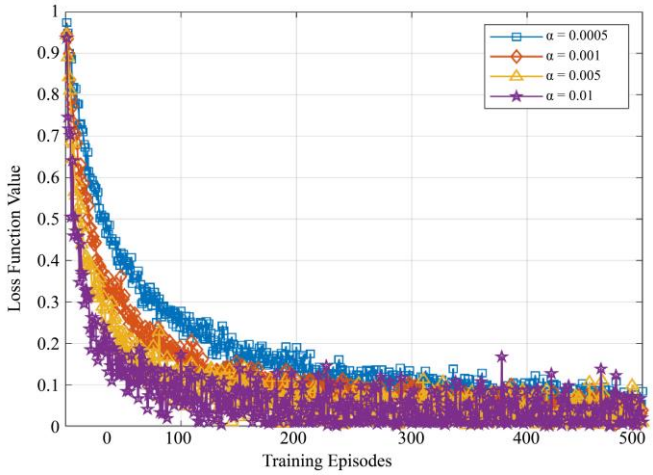

Fig. 14 Impact of learning rate on convergence

Due to the inherent NP-hard complexity, the problem is reformulated as an MDP to enable the tractable solution. The reward function was designed to reflect the optimization objectives. To address the complexity of the constrained problem, The Lagrangian relaxation method was applied to transform the constrained optimization problem into an unconstrained formulation by relaxing both BER and latency constraints. Deep Q-learning was then employed to solve the problem efficiently. This allowed for intelligent decision-making regarding vehicle speed and modulation order selection. Extensive simulations were conducted to evaluate the performance of the proposed scheme. The results demonstrated that the proposed approach significantly improves sum spectral efficiency while achieving lower average latency compared to alternative schemes. By analyzing the CDF of experienced latency and BER, Experimental results confirm that the proposed system satisfies ultra-low latency communication requirements while consistently maintaining BER constraints.

In contrast, competing schemes failed to consistently satisfy these constraints over extended periods. In summary, this study highlights the effectiveness of using DRL to optimize vehicular OCC performance. Future research directions include further improvements in learning efficiency, incorporating more complex mobility models and developing adaptive strategies for real-world deployment.

References

[1] Felipe Cunha et al., *Vehicular Networks to Intelligent Transportation Systems*, Emerging Wireless Communication and Network Technologies, Springer, Singapore, pp. 297-315, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[2] Mario Gerla, and Leonard Kleinrock, "Vehicular Networks and the Future of the Mobile Internet," *Computer Networks*, vol. 55, no. 2, pp. 457-469, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[3] You Han et al., "Spectrum Sharing Methods for the Coexistence of Multiple RF Systems: A Survey," *Ad Hoc Networks*, vol. 53, pp. 53-78, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[4] Oluwaferanmi Oluwatosin Atiba, "*Optical Wireless and Visible Light Communication Techniques*," Master's Thesis, Tampere University, pp. 1-67, 2023. [Google Scholar] [Publisher Link]

[5] Amirul Islam, "*Machine Learning Assisted Ultra Reliable and Low Latency Vehicular Optical Camera Communications*," Ph.D. Thesis, University of Essex, pp. 1-168, 2022. [Google Scholar] [Publisher Link]

[6] He Chen et al., "Ultra-Reliable Low Latency Cellular Networks: Use Cases, Challenges and Approaches," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 119-125, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[7] Amirul Islam, Nikolaos Thomos, and Leila Musavian, "Achieving uRLLC with Machine Learning Based Vehicular OCC," *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, Rio de Janeiro, Brazil, pp. 4558-4563, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Ana L.C. Bazzan, "A Distributed Approach for Coordination of Traffic Signal Agents," *Autonomous Agents and Multi-Agent Systems*, vol. 10, pp. 131-164, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[9] Guillermo Pocovi et al., "Achieving Ultra-Reliable Low-Latency Communications: Challenges and Envisioned System Enhancements," *IEEE Network*, vol. 32, no. 2, pp. 8-15, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[10] Zidong Zhang, Dongxia Zhang, and Robert C. Qiu, "Deep Reinforcement Learning for Power System Applications: An Overview," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 1, pp. 213-225, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[11] Xin Xu et al., "A Reinforcement Learning Approach to Autonomous Decision Making of Intelligent Vehicles on Highways," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 10, pp. 3884-3897, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[12] Mikhail A. Bragin et al., "Convergence of the Surrogate Lagrangian Relaxation Method," *Journal of Optimization Theory and Applications*, vol. 164, pp. 173-201, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[13] Laetitia Matignon, Guillaume J. Laurent, and Nadine Le Fort-Piat, "Independent Reinforcement Learners in Cooperative Markov Games: A Survey Regarding Coordination Problems," *The Knowledge Engineering Review*, vol. 27, no. 1, pp. 1-31, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[14] Yuki Goto et al., "A New Automotive VLC System Using Optical Communication Image Sensor," *IEEE Photonics Journal*, vol. 8, no. 3, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[15] Takaya Yamazato et al., "Image-Sensor-Based Visible Light Communication for Automotive Applications," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 88-97, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[16] Anitha Saravana Kumar, Lian Zhao, and Xavier Fernando, "Multi-Agent Deep Reinforcement Learning-Empowered Channel Allocation in Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 1726-1736, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Amirul Islam, Nikolaos Thomos, and Leila Musavian, "Multi-Agent Deep Reinforcement Learning for Spectral Efficiency Optimization in Vehicular Optical Camera Communications," *IEEE Transactions on Mobile Computing*, vol. 23, no. 5, pp. 3666-3679, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Zhiyong Du et al., "Context-Aware Indoor VLC/RF Heterogeneous Network Selection: Reinforcement Learning With Knowledge Transfer," *IEEE Access*, vol. 6, pp. 33275-33284, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[19] Anastasios Giannopoulos et al., "Deep Reinforcement Learning for Energy-Efficient Multi-Channel Transmissions in 5G Cognitive HetNets: Centralized, Decentralized and Transfer Learning Based Solutions," *IEEE Access*, vol. 9, pp. 129358-129374, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Willy Anugrah Cahyadi et al., "Optical Camera Communications: Principles, Modulations, Potential and Challenges," *Electronics*, vol. 9, no. 9, pp. 1-44, 2020. [CrossRef] [Google Scholar] [Publisher Link]