smart cities

MDPI

Article

An Intelligent Arterial Traffic Control Framework for Visible Light-Connected Vehicles

Gonçalo Galvão [1,2,*], Manuela Vieira [1,2,3,*], Manuel Augusto Vieira [1,2], Mário Véstias [1,4] and Paula Louro [1,2]

[1] Instituto Superior de Engenharia de Lisboa (ISEL), Instituto Politécnico de Lisboa (IPL), 1959-007 Lisbon, Portugal; mvieira@deetc.isel.ipl.pt (M.A.V.); mario.vestias@isel.pt (M.V.); paula.louro@isel.pt (P.L.)
[2] UNINOVA-CTS and LASI, Quinta da Torre, Monte da Caparica, 2829-516 Caparica, Portugal
[3] NOVA School of Science and Technology, Quinta da Torre, Monte da Caparica, 2829-516 Caparica, Portugal
[4] INESC INOV, 1000-029 Lisbon, Portugal
* Correspondence: goncalo.galvao@isel.pt (G.G.); manuela.vieira@isel.pt (M.V.)

Highlights

What are the main findings?

- Deployment of smart traffic management approaches that dynamically prioritize certain routes or arterials according to real-time traffic conditions.
- Development of the SAPA module, designed to modify signal timings in response to vehicle queue lengths and the traffic load of adjacent intersections.

What are the implications of the main findings?

- Showcases the effectiveness of adaptive deep reinforcement learning methods in optimizing traffic flow across complex multi-intersection networks.
- Emphasizes that incorporating SAPA block further boosts system efficiency, leading to lower congestion and shorter delays for both vehicles and pedestrians.

Abstract

Inefficient urban traffic management remains a critical challenge, as conventional signal controllers—built on fixed timing plans—cannot cope with the dynamic nature of modern city traffic. This study addresses this limitation by developing a decentralized MARL-based framework capable of coordinating five interconnected intersections as a unified traffic cell. Central to the proposed solution is the Strategic Anti-Blocking Phase Adjustment (SAPA) module, which enables intersections to autonomously modify phase durations in response to real-time traffic conditions. The framework is designed to handle heterogeneous demand patterns, with particular emphasis on arterial corridors connecting urban centers to peripheral zones. Integration of a Visible Light Communication (VLC) network allows continuous monitoring of key variables, including vehicle kinematics and pedestrian activity, feeding the agents with rich environmental feedback. Experimental evaluation confirms the effectiveness of the approach: the SAPA-augmented DQN achieves roughly 33% shorter vehicle queues and a ~70% reduction in pedestrian waiting counts relative to a standard DQN baseline. Remarkably, these gains bring the value-based method to a performance level comparable to MAPPO, a considerably more complex multi-agent policy optimization algorithm, establishing SAPA as an efficient and scalable enhancement for intelligent urban traffic control.

Check for updates

Academic Editors: Chaoxu Mu, Anguo Zhang, Qichun Zhang, Malu Zhang, Xingshuo Han and Pierluigi Siano

Received: 18 February 2026
Revised: 1 April 2026
Accepted: 15 April 2026
Published: 20 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: deep reinforcement learning (DRL); visible light communication (VLC); multi-agent systems; urban traffic management; autonomous vehicles; traffic management and efficiency

1. Introduction

Traffic congestion remains one of the most pressing challenges in modern urban environments, generating daily delays that impose significant socioeconomic and environmental burdens on cities and their inhabitants. The continuous growth of urban populations has driven a sharp increase in vehicle numbers, overwhelming road infrastructures and resulting in deteriorating travel times, rising fuel consumption, higher greenhouse gas emissions, and reduced productivity.

Despite the scale of these problems, traffic management technologies have largely failed to evolve accordingly. In many cities, Lisbon being a representative example, signal control systems continue to operate on pre-programmed, cyclic timing plans that are indifferent to real-time conditions, rendering them incapable of responding to accidents, demand surges, or temporal fluctuations in traffic load. The sensing infrastructure further compounds this limitation, as fixed-point technologies such as inductive loops and stationary cameras lack the spatial coverage and resolution required for data-driven decision-making. Consequently, signal control logic remains largely pre-scheduled or weakly reactive, rather than anticipatory and adaptive [1,2].

Intersections are particularly critical in this context, as they represent the primary sources of queue formation and congestion propagation within urban networks. Poorly coordinated intersections degrade network-wide throughput and introduce safety risks for both drivers and pedestrians, underscoring the need for intelligent, real-time-aware control frameworks.

The main contributions of this work focus on the integration of advanced vehicular communication technologies with intelligent control methods to improve urban traffic management.

First, this research proposes an integrated framework that combines Connected Vehicles (CVs) and VLC to enable real-time, high-resolution traffic information acquisition. By leveraging the dual functionality of LED-based infrastructure for both illumination and communication, the system enhances sensing and data exchange capabilities in urban traffic environments, particularly at signalized intersections.

Second, unlike our previous studies that focused on isolated intersections, this work considers a coordinated network-level scenario composed of five interconnected intersections. This enables the analysis of spatial interactions between intersections and captures inter-agent dependencies, providing a more realistic evaluation of urban traffic dynamics.

Third, the study introduces a novel adaptive traffic control approach based on DRL, incorporating the SAPA module developed in this work. The SAPA mechanism dynamically adjusts signal phase durations based not only on local queue conditions but also on downstream lane occupancy, enabling anticipatory and context-aware behavior that improves traffic flow while preventing congestion propagation.

Fourth, a composite reward function is designed to jointly optimize vehicular flow and pedestrian mobility, ensuring that the learned control policies balance the operational requirements of both traffic participants and enhance the realism of the system in urban environments.

Fifth, this work provides a systematic and controlled comparison between two prominent DRL approaches, Deep Q-Learning (DQL) and Multi-Agent Proximal Policy Opti-

mization (MAPPO), under identical traffic conditions. This comparison offers new insights into the relative strengths and limitations of value-based and policy-based methods in multi-agent traffic signal control.

Finally, the proposed system integrates VLC-enabled communication within a multi-agent reinforcement learning framework, unifying communication, control, and learning layers into a single architecture. The system is evaluated under realistic urban traffic conditions, demonstrating its potential to improve intersection performance, coordinate traffic flows, and enhance overall efficiency through the synergy of communication technologies and intelligent control strategies.

The remainder of this paper is structured as follows: Section 2 reviews relevant literature and foundational concepts concerning VLC, DRL, and connected vehicles in the context of traffic management. Section 3 describes the urban traffic scenario, the proposed MARL framework, the DRL algorithms employed, the traffic strategies considered, and the SAPA module. Section 4 presents and analyzes the performance of the trained models across the different traffic strategies. Section 5 concludes the paper with a discussion of key findings, limitations, and potential directions for future work.

2. Literature Review

2.1. Urban Traffic Management, Challenges and New Perspectives

Traffic management is concerned with organizing, arranging, guiding, and controlling both moving and stationary traffic, including vehicles, cyclists, and pedestrians. Traffic Management Systems aim to ensure safety, orderliness, and efficiency in the movement of people and goods, while simultaneously enhancing the environmental quality of traffic-affected areas. Nowadays, urban traffic management faces increasing challenges due to congestion, delays, and safety concerns. With the rapid growth and technological advancement of the automotive market, the number of vehicles circulating on city roads continues to rise. However, traffic infrastructures and control technologies have not evolved at the same pace. This imbalance highlights the inadequacy of traditional strategies, such as road expansion or fixed-timed signals, which remain reactive, inflexible, and ultimately unsustainable in addressing the complexity of contemporary urban mobility [3].

One of the most critical aspects of traffic control is the collection and prediction of traffic data. Traditionally, this information was gathered through sensors. However, such systems often suffer from environmental interference, inaccuracies, inconsistency of data, or financial infeasibility for large-scale deployment. With technological progress, more recent studies have proposed exploiting vehicles themselves and traffic infrastructures as sources of reliable data [4]. Vehicular communication has thus emerged as one of the key enabling technologies for future ITS.

The term connected vehicles refers to applications, services, and technologies that allow vehicles to communicate with their surroundings. The main objective of ITS technologies is to optimize traffic safety and efficiency by improving situational awareness and mitigating accidents through V2V, I2V, and V2I communications. With V2V, vehicles can exchange real-time information such as location, speed, and direction, which is crucial for cooperative maneuvers and safety functionalities. I2V provides connected vehicles with valuable traffic control information, such as traffic light status or warnings regarding signal violations, weather conditions, and driving recommendations, all delivered in real time. Similarly, V2I enables vehicles to share and receive detailed traffic information (e.g., speed, volume, travel time, queue length, and stop frequency) as well as road surface condition data (e.g., roughness or slipperiness). Such data exchange allows for more accurate traffic predictions and supports the activation of rerouting strategies, enabling traffic to flow through alternative routes and reducing congestion.

In these communication systems, transmitters and receivers are strategically placed to ensure seamless data transmission. On the vehicle side, headlights can serve as transmitters, while photodetectors mounted at the rear or on the roof function as receivers for signals transmitted by other vehicles or infrastructures. Infrastructures are likewise equipped with sensors placed near their light emitters to facilitate communication.

These connected-vehicle communications rely on various wireless technologies. Initially, mobile radio communications were considered sufficient. However, the diversity of requirements has since revealed the limitations of radio spectrum availability, which is becoming increasingly saturated. This has motivated research into alternative communication technologies to support ITS [5].

Another promising approach in traffic management involves the use of neural networks, which exhibit high learning capacity, robust pattern recognition, and strong modeling capabilities for traffic dynamics. Nevertheless, many existing algorithms remain predominantly vehicle-centric, often neglecting the dynamics of pedestrians and cyclists. Addressing this gap requires reinforcement learning-based traffic control approaches that explicitly incorporate pedestrian behavior. Key challenges include accommodating bidirectional pedestrian flows, accounting for differences in movement dynamics between vehicles and pedestrians, ensuring a balanced trade-off between efficiency and safety.

In reinforcement learning-based traffic control, agents observe intersection states, such as waiting times and queue lengths, to determine optimal signal phases [6]. However, single-agent approaches face scalability limitations in multi-intersection scenarios, motivating research into collaborative strategies that account for interdependencies among neighboring intersections to achieve efficiency at network scale.

Building on these advances, the proposed framework exploits real-time traffic data within Vehicle-to-Everything (V2X) communication environments to enable adaptive signal control, improving both operational efficiency and safety. Unlike conventional systems that rely on fixed detectors with limited occupancy and flow measurements, V2X-enabled adaptive systems capture richer information, including vehicle position, speed, queue length, and stop duration. While V2V communication is indispensable for safety applications such as pre-crash sensing and collision avoidance, I2V/P is particularly important for delivering timely, context-aware information to both connected vehicles and pedestrians, thereby enhancing safety and mobility.

Furthermore, for the traffic control of the considered environment, intelligent traffic management strategies were developed in order to obtain trained models for various traffic scenarios. These scenarios simulate realistic traffic fluctuations, such as morning rush hours, characterized by a strong inbound flow of vehicles entering the city, and evening rush hours, during which a higher volume of vehicles travels outward [7]. In addition, an analysis of arterial traffic control is conducted, enabling the system to manage traffic across multiple arterial roads by prioritizing one direction over another based on prevailing traffic conditions.

2.2. Recent Advances and Challenges in Vehicular Visible Light Communication

For ITS to be effective, a large-scale deployment of intelligent vehicles and infrastructures is required to ensure comprehensive data acquisition and dissemination. Although Radio Frequency (RF) communication has traditionally supported these systems, the rapid growth of mobile data traffic has exposed the limitations of the RF spectrum [8]. In contrast, the visible light spectrum offers vast, unlicensed bandwidth capable of supporting high-capacity communication networks.

VLC, enabled by the advancement of LED technology, utilizes light intensity modulation for data transmission and photodetectors or camera sensors for signal reception.

When integrated with V2V and V2I/I2V frameworks, VLC has the potential to substantially enhance vehicular communication performance, particularly in dense traffic environments, thereby improving road safety and reducing collision risk [9]. The advantages of VLC are further emphasized, including its high spectral efficiency, immunity to RF interference, and the dual use of existing lighting infrastructure. Furthermore, VLC offers enhanced security due to the confined propagation of light, which limits susceptibility to eavesdropping, as well as low-latency communication that is well-suited for safety-critical vehicular applications. Its directional nature enables high spatial reuse, supporting multiple simultaneous transmissions in dense traffic scenarios, while also contributing to reduced electromagnetic pollution and a more interference-resilient communication environment.

While VLC offers compelling advantages, its deployment in outdoor settings is subject to a range of environmental constraints. Atmospheric phenomena such as sunlight interference, fog, rain, and snow can attenuate or saturate optical signals, compromising link reliability and overall system performance [10]. For this reason, VLC is more appropriately regarded as a complement to RF communication than a replacement. A hybrid architecture combining both technologies can yield a more robust solution, with VLC delivering high-throughput line-of-sight links and RF ensuring connectivity when optical paths are obstructed or weather conditions are adverse.

Within the CAV ecosystem, embedding VLC capabilities into both vehicles and roadside infrastructure opens new possibilities for traffic coordination [11]. Real-time exchange of vehicle dynamics and road state information through VLC-based V2V and V2I/I2V links enables more responsive signal control and smoother traffic flow. An additional benefit emerges in congested conditions, where shorter inter-vehicle gaps favor chain-like optical communication, reinforcing the connectivity fabric of cooperative transportation systems.

The behavior of outdoor VLC under varying environmental and geometric conditions has been the subject of recent investigations [12], with BER serving as a common performance indicator. One such study [13] addresses tunnel scenarios, developing an I2V communication model that accounts not only for direct LoS paths from infrastructure transmitters but also for NLoS contributions arising from wall reflections. Another work [14] focuses on V2V links, systematically examining the effects of ambient light, lateral vehicle separation, and daytime versus nighttime operation. While the findings broadly confirm VLC's robustness and capacity potential, they also identify performance boundaries under extreme conditions, notably dense fog, high ambient illumination, and large inter-vehicle distances, that must be considered in system design.

Building upon current advancements, this paper introduces a novel framework that establishes a synergistic integration of VLC-based localization and communication services with a learning-driven traffic signal control paradigm. The core objective is to optimize both vehicular and pedestrian mobility across multi-intersection networks under realistic operating conditions.

The proposed V-VLC architecture, depicted in Figure 1, is designed within a hybrid communication framework that combines VLC and RF technologies. In this architecture, VLC serves as the primary high-bandwidth, low-latency channel, while RF communication (e.g., IEEE 802.11p or 5G) operates as a complementary fallback mechanism to ensure robustness under non-ideal conditions, such as signal attenuation due to adverse weather (e.g., fog, rain), line-of-sight blockages, ambient light interference, and potential packet loss or communication delays.

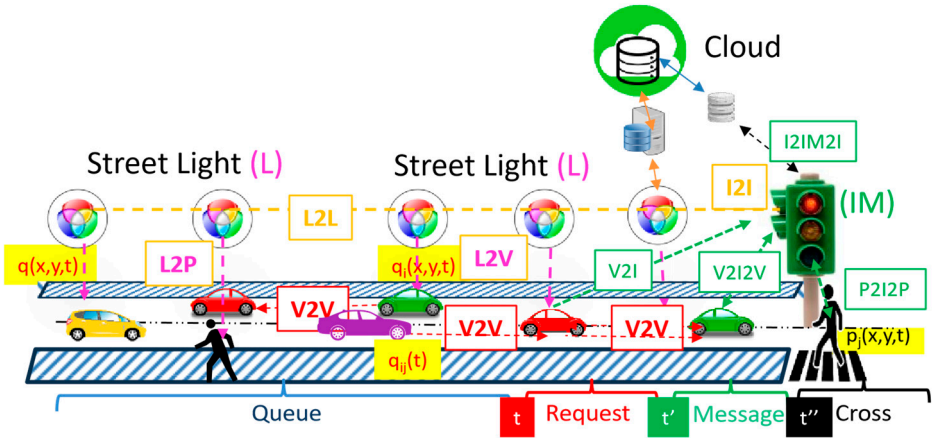


Figure 1. Schematic overview of the V-VLC architecture.

The system integrates both vehicles and infrastructure components within a unified V2X ecosystem, explicitly accounting for heterogeneous traffic environments. This includes VLC-enabled connected vehicles as well as non-connected vehicles, for which baseline system functionality is maintained through infrastructure-based sensing (e.g., traffic lights and detectors).

Communication protocols are embedded within a MARL framework, enabling adaptive and decentralized traffic management. At each intersection, a queue–request–response protocol is implemented: approaching vehicles and pedestrians issue crossing requests (V/P2I), and traffic signals, acting as intelligent agents, respond accordingly (I2V/P).

While prior studies focus primarily on the physical-layer performance of vehicular VLC systems, their integration with data-driven control strategies remains largely unexplored. In contrast, this research investigates a framework that combines VLC-based traffic sensing with reinforcement learning to enable intelligent and adaptive traffic signal control.

2.3. Intelligent Urban Traffic Management Through Deep Reinforcement Learning

The architectural complexity of DRL-based traffic control systems varies considerably depending on the scale and structure of the problem. At one end of the spectrum, isolated intersections have been addressed using single-agent RL formulations [15], while more elaborate urban networks demand the coordination of multiple agents operating simultaneously under a MARL framework [16]. The design of such systems involves several fundamental choices: whether decision-making authority is centralized in a single controller or distributed across independent agents [17], whether each agent perceives the full environment or only a local subset of it, and whether agents collaborate toward a common objective or compete to maximize individual rewards [18,19]. These architectural decisions profoundly influence system scalability, robustness, and the nature of inter-agent interactions, shaping the overall effectiveness of the traffic management solution.

Recent research in the field has increasingly focused on hybrid or hierarchical MARL architectures, in which a central agent is responsible for learning high-level coordination strategies while distributed local agents execute context-aware actions. These architectures seek to balance scalability, coordination, and robustness by integrating multi-objective reward functions that simultaneously consider efficiency and safety criteria.

Broadly, MARL algorithms can be divided into two main families: value-based and policy-based approaches. Value-based techniques, such as DQL, estimate a value function representing the expected return of each possible action and select the one with the highest predicted value [20,21]. These methods are computationally efficient and particularly suitable for independent or parameter-sharing settings. Moreover, they tend to be easier to

implement and interpret, performing well in environments characterized by a finite and discrete action space.

In contrast, policy-based approaches directly learn a policy that maps states to actions, often employing an actor–critic framework [22,23]. Within this structure, the actor determines the actions according to the learned policy, while the critic assesses their performance, thereby reducing variance and improving the stability of the learning process. Although value-based algorithms retain advantages in terms of efficiency, policy-based methods, especially those using actor–critic mechanisms, frequently demonstrate superior performance in complex and large-scale networks. The critic's capacity to evaluate joint observations and actions contributes to more coherent and coordinated phase transitions across multiple intersections, ultimately promoting greater adaptability and stability in global traffic flow dynamics.

Although previous studies have explored various reinforcement learning approaches for traffic signal optimization, the present work distinguishes itself by leveraging vehicular connectivity through VLC to gather high-resolution traffic data in real time. Beyond simply controlling intersections, this study emphasizes arterial traffic patterns, making a clear distinction between radial and circular roads and incorporating a wide range of traffic scenarios that reflect realistic urban conditions, including both vehicular and pedestrian flows. By doing so, it allows the training of DQN and PPO-based neural networks capable of learning complex spatial and temporal traffic patterns across the city. Both models are designed not only to optimize overall traffic flow and reduce congestion but also to ensure pedestrian safety and prioritize the protection of critical urban zones, providing a more resilient and adaptive city-wide traffic management framework compared to existing approaches. Another important aspect addressed in this research is the study of the influence of phase durations when activated. To this end, a SAPA algorithm was developed, which, through a spatial and temporal perception of the environment at each intersection, is able to dynamically adjust the phase duration according to the number of vehicles waiting to cross the intersection and the occupancy rate of the downstream road segment of the neighboring intersection. To evaluate the impact of this approach, an ablation study is also conducted, where a state-of-the-art DQN is compared with a DQN incorporating the proposed SAPA algorithm.

This comparative analysis aims to assess the behavioral and performance distinctions between the two algorithms, with emphasis on their capacity to optimize traffic flow across a network of interconnected intersections. The intelligent VLC-enabled control model is evaluated using SUMO, an agent-based urban traffic simulation platform [24].

3. Urban Traffic Scenario and Proposed MARL Framework

3.1. Traffic Environment

The modeled network in Figure 2 comprises five urban intersections configured to reflect peak-hour city traffic, accommodating inbound flow from C3 to C4 and outbound flow in the reverse direction. Each intersection features two lanes per approach, with the right lane serving both through and right-turn movements and the left lane reserved exclusively for left turns. Pedestrian crossings are also incorporated, addressing an aspect frequently overlooked in comparable studies. The network topology is defined by two intersecting arterials: a circular route spanning C0, C1, and C2, and a radial corridor connecting C3, C1, and C4, converging at the central junction C1.

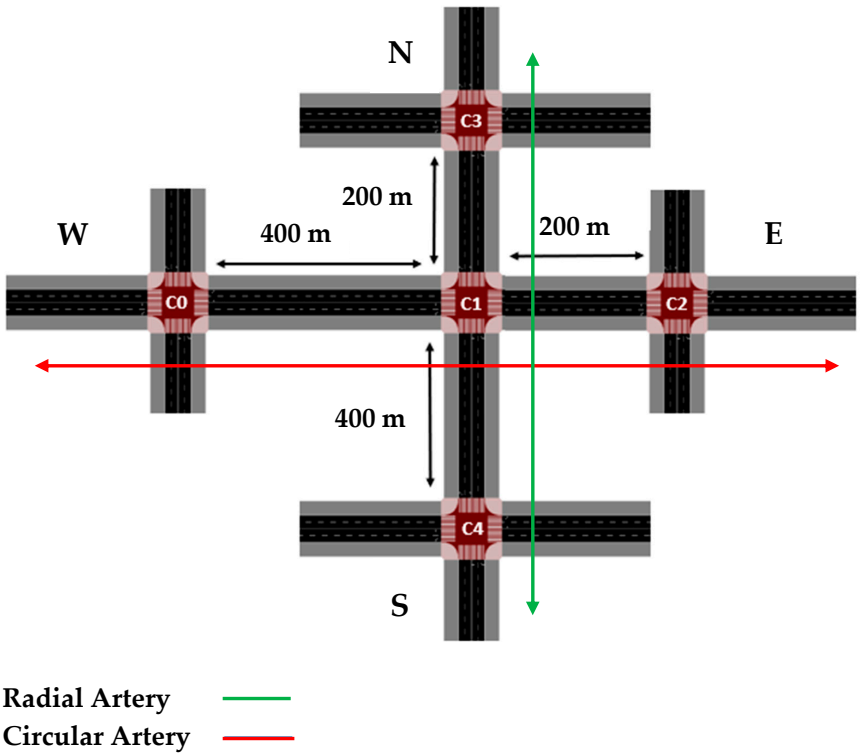


Figure 2. Traffic Scenario consisting of 5 homogeneous intersections.

Intersection C1 serves as the central coordination node connecting the horizontal and vertical arterial axes. Unlike neighboring junctions, it generates no traffic demand of its own, receiving all incoming flows from adjacent intersections (C0, C2, C3, and C4). Consequently, its traffic volume depends entirely on the phase decisions of neighboring agents, making effective synchronization at C1 critical for overall network stability. This work therefore investigates how different priority configurations assigned to C1 influence system-wide performance.

The road segments connecting the intersections vary in length, reflecting real urban heterogeneity: the C0–C1 and C4–C1 links span approximately 400 m, while the C2–C1 and C3–C1 connections are around 200 m long.

Each intersection is regulated by a signalized controller operating nine distinct phases, eight for vehicular movements and one exclusively for pedestrian crossings, as illustrated in Figure 3, during which all vehicular movements are halted to ensure safe and conflict-free crossing conditions.

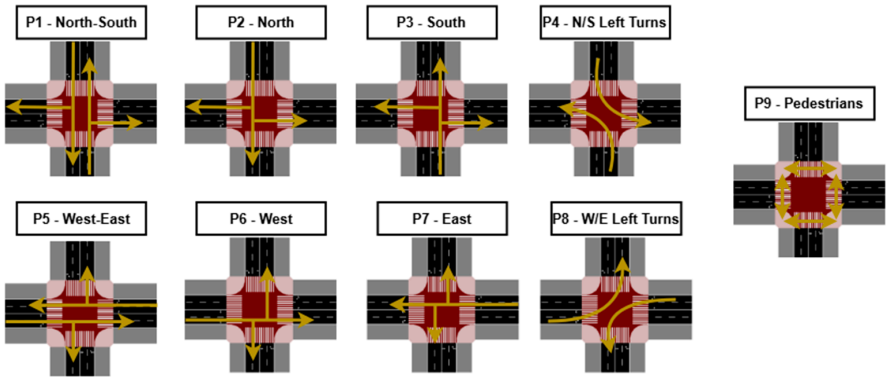


Figure 3. Illustration of the nine signal phases considered: eight vehicular phases (P1–P8) representing permitted movement directions, and one dedicated pedestrian phase (P9).

3.2. MARL Deep Q-Learning Algorithm

A MARL framework is proposed to coordinate the five intersections of the urban network, as depicted in Figure 4a. The system assigns a dedicated agent to each intersection, with all agents sharing a single neural network. Each agent observes the local traffic conditions of its incoming lanes and selects from nine predefined signal phases.

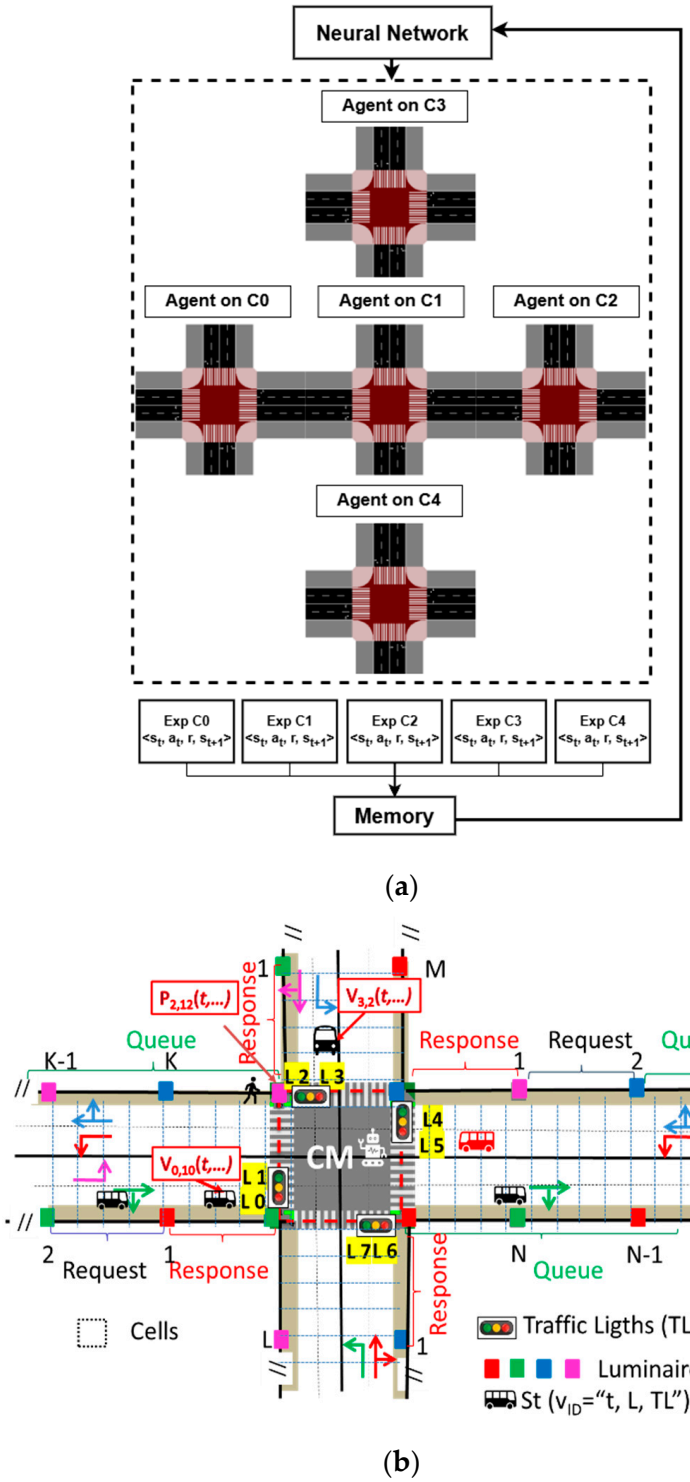


Figure 4. Multi-Agent Reinforcement Learning System with a CTDE architecture (**a**) and the corresponding cell-based representation of each lane approaching the intersection (**b**).

Agent experience is encoded as a tuple comprising the current state, the action taken, the reward received, and the resulting next state. Experiences collected across all agents are stored in a shared replay buffer and used jointly to train the common network, allowing each agent to benefit from the collective knowledge of the entire system.

The framework follows the Centralized Training with Decentralized Execution (CTDE) paradigm, which has been shown in prior work to improve network-wide performance through indirect inter-agent cooperation. Centralized training enables the shared network to leverage aggregated experiences, promoting convergence and flow coordination across the arterial network, while decentralized execution preserves scalability. Compared to a fully centralized architecture, this approach avoids combinatorial action space explosion and reduces computation overhead, while remaining straightforwardly extensible to larger intersection networks.

Each intersection is managed by a dedicated MARL agent responsible for monitoring local traffic conditions through VLC-based sensing, capturing real-time data on both vehicles and pedestrians. This information is shared among neighboring agents to support cooperative decision-making across the network.

The collected data feeds a DQL neural network trained to identify the optimal signal phase at each intersection by estimating expected cumulative rewards. In contrast to conventional tabular Q-Learning, which becomes intractable in large state-action spaces, the DQL approach employs neural networks to handle the high dimensionality characteristic of urban traffic environments. The architecture adopted is a FCLN with weights adjusted to approximate the Q-value function, representing the expected return of executing action in state.

The state representation integrates three information layers, as illustrated in Figure 4b. The first two layers each contain 80 cells, ten per lane, encoding vehicle presence as binary occupancy values and normalized speed, respectively. The third layer comprises four cells reflecting pedestrian counts in designated waiting zones. This yields an input vector of 164 neurons. Two hidden layers, each containing 400 neurons with ReLU activation, capture the underlying complexity of traffic dynamics. The output layer produces nine Q-values, one per candidate signal phase. Network training is performed by minimizing the MSE loss function. Table 1 shows more hyperparameters of the DQL network.

Table 1. Main parameters used in the DQN model for both main and target networks.

Parameter	Value
Input Layer	164
Hidden Layers	2
Width Hidden Layers	400
Batch Size	128
Discount Factor	0.75
Output Layer	9
Training Epochs	100
Learning Rate	0.0001
Optimizer	Adam
Activation Function	ReLu

The neural network training employs two separate networks with identical architectures: a primary network responsible for predicting Q-values (Q_{pred}) and a target network (Q_{target}), updated less frequently, which is used to calculate the target Q-values as defined in Equation (1).

$$Q_{target} = r_t + \Upsilon \cdot \max\left[Q_{pred}\left(s_{t+1}, a', \theta_k\right)\right] \quad (1)$$

Within this framework, the main network produces Q-value predictions, while a structurally identical target network provides stable reference values for training. Rather than being updated through backpropagation, the target network's weights are periodically synchronized with those of the main network. This asynchronous update mechanism reduces temporal correlations between predicted and target Q-values, contributing to improved training stability and convergence. A discount factor is additionally applied to future reward estimates, controlling the trade-off between immediate and long-term returns.

The reward function guides agent behavior by quantifying the effectiveness of each action. As a novel contribution, it combines two weighted components, one capturing vehicular performance and the other reflecting pedestrian experience, as defined in Equations (2) and (3), allowing the system to prioritize vehicle flow, pedestrian service, or a balanced trade-off between both.

In this study, these weights were configured to achieve an equal 50/50 balance between vehicle and pedestrian considerations. Previous experiments were conducted using different weight configurations, and the selected balance yielded the best overall performance. This configuration allows the system to remain responsive to both vehicular traffic and pedestrian demands, ensuring a more balanced and efficient operation of the intersection.

$$\mathrm{r_t} = p_{veh}(\mathrm{atwt}_{veh,t-1} - \mathrm{atwt}_{veh,t}) + p_{ped}\left(\mathrm{atwt}_{ped,\,t-1} - \mathrm{atwt}_{ped,t}\right) \tag{2}$$

The reward considers both vehicle and pedestrian average total accumulated waiting times ($atwt$). $wt_{veh,t}/wt_{ped,t}$ is the amount of time in seconds a vehicle/a pedestrian has a speed of less than 0.1 m/s at t since the spawn into the environment, and n represents the total number of vehicles/pedestrians in the environment in t per intersection.

$$\begin{aligned} atwt_{veh,t} &= \textstyle\sum_{veh=1}^{n} wt_{(veh,t)} \\ atwt_{ped,t} &= \textstyle\sum_{ped=1}^{n} wt_{(ped,t)} \end{aligned} \tag{3}$$

3.3. Multi Agent Proximal Policy Proximization Algorithm

MAPPO is an on-policy, policy-gradient algorithm tailored for cooperative multi-agent environments. It builds upon the standard Proximal Policy Optimization (PPO) by employing a centralized critic, which leverages information from the global state, alongside decentralized actor policies for each agent. This combination preserves scalability while enhancing learning stability, making MAPPO particularly suitable for controlling traffic signals across multiple intersections.

In contrast to DQN, which is an off-policy method that learns from a replay buffer, MAPPO updates agent policies exclusively using current trajectories, thereby avoiding the instability that may arise from outdated experiences. The algorithm optimizes a clipped surrogate objective, defined at Equations (4) and (5):

$$L_{\mathrm{CLIP}}(\theta) = \hat{E}_t[\min(r_t(\theta)\hat{A}_t, \mathrm{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)], \tag{4}$$

$$r_t(\theta) = \frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{\mathrm{old}}}(a_t \mid s_t)} \tag{5}$$

$r_t(\theta)$ is the probability ratio, and $\hat{A}_t$ represents the estimated advantage. The clipping operation constrains policy updates, preventing excessively large gradient steps and thereby stabilizing training. The operator $\hat{E}_t[\cdot]$ denotes the empirical expectation over the collected batch of timesteps, corresponding to the average across sampled trajectories. The

advantage is computed using Generalized Advantage Estimation (GAE), represented in Equation (6), where γ and λ control discounting and bias–variance trade-off.

$$\hat{A}_t = \sum_{l=0}^{\infty} \Big(\gamma\lambda)^l [r_t + \gamma V(s_{t+1}) - V(s_t)], \tag{6}$$

MAPPO optimizes its performance by minimizing two separate loss functions: the actor loss and the critic loss, as defined in Equations (7) and (8), respectively. An additional entropy term (H) is included in the actor loss to encourage exploration, promoting more robust policy learning and preventing premature convergence to suboptimal strategies.

$$L_\pi(\theta) = -\hat{E}_t[L_{\text{CLIP}}(\theta) + c_{\text{ent}}\, H[\pi_\theta(\cdot \mid s_t)]] \tag{7}$$

$$L_V(\phi) = c_{vf}\, \hat{E}_t[(R_t - V_\phi(s_t))^2], \tag{8}$$

An additional enhancement implemented in this algorithm is the integration of One-Hot Encoding (OHE) to represent the identity of each agent. This binary vector, containing a single active element per agent, is concatenated with the actor's observation input, enabling the shared network to distinguish between intersections and adapt its behavior to local traffic dynamics. The use of OHE helps mitigate policy aliasing, a common issue in shared-network architectures such as DQN, where similar observations may correspond to distinct optimal actions. By combining centralized training with a global critic and decentralized execution, MAPPO achieves coordinated yet autonomous decision-making across intersections, resulting in more stable and adaptive traffic signal control policies.

The MAPPO framework employs a centralized critic and decentralized actors. The critic network, composed of two fully connected layers with 256 neurons and ReLU activations, receives a global observation vector of dimension 820, obtained by concatenating the local observations of all five agents, each represented by a 164-dimensional state vector. The critic outputs a scalar state-value estimate $V(s)$, providing a comprehensive evaluation of the global traffic state and supporting coordinated learning across intersections. Each actor network processes its local observation concatenated with a one-hot encoded agent ID through two hidden layers of 128 neurons, producing nine action logits converted to probabilities via a softmax function.

Training uses Adam optimization (learning rate 0.002) with hyperparameters tuned for stability and efficiency: a discount factor $\gamma = 0.75$, a GAE parameter $\lambda = 0.75$, and a clipping threshold $\epsilon = 0.2$. Each training cycle comprises rollouts of 4096 steps, updated over four PPO epochs with matching minibatch size. Exploration is promoted through entropy regularization ($c_{ent} = 0.01$), while the critic loss is weighted by $c_{vf} = 0.5$. Gradients are clipped to a norm of 0.5, advantages are standardized, and invalid actions are masked to prevent infeasible decisions.

3.4. Strategic Anti-Blocking Phase Adjustment (SAPA)

One of the key enhancements introduced in this architecture is the SAPA module, a mechanism designed to dynamically adjust the duration of green phases after they are selected by the neural network. While the MARL agent determines which phase to activate, SAPA refines how long that phase should remain active by using real-time data obtained from VLC-enabled sensing. SAPA integrates information on upstream queue lengths and downstream lane occupancy, allowing the controller to prevent spillback, reduce blocking, and maximize throughput across the network.

SAPA operates by comparing upstream and downstream occupancy levels to predefined thresholds. When downstream occupancy remains below the designated threshold,

the system proportionally extends the green phase to clear accumulated queues. However, if the occupancy approaches critical levels, SAPA restricts the green duration to the minimum time required, 8 s, preventing oversaturation of the downstream link. This mechanism ensures that intersections cooperate implicitly by avoiding blockages that could propagate through the network, while also preventing excessively long green phases by limiting their duration to a maximum of 40 s.

To illustrate its operation, consider a corridor of five intersections where traffic flows from C0 to C1 are routed through a 400 m link. When the occupancy of this link remains below 40%, SAPA proportionally extends the green phase at C0, enabling more vehicles to move toward C1. Conversely, when vehicles progress from a 400 m link into a shorter 200 m section, the occupancy threshold is reduced to 35% to prevent overflow in the more constrained downstream segment. These dynamically adjusted thresholds ensure stability and prevent congestion under heterogeneous geometric conditions.

SAPA also incorporates an adaptive priority-weighting scheme that reflects distinct demand levels on radial and circular arterials. A balanced configuration distributes influence evenly (50–50%), whereas scenarios with dominant inbound or outbound flows increase the weight of one corridor up to 65%, allocating additional capacity where it is most needed. Low-priority movements, such as left turns, are assigned reduced weights (25%), limiting their impact on green-phase extensions. The central intersection (C1) plays a coordinating role, mediating the timing adjustments between radial and circular flows according to the selected priority strategy, thereby maintaining consistent operation across the network.

3.5. Traffic Control Strategies

Effective urban traffic management must account for several dynamic factors, among which temporal demand variations—particularly morning and afternoon rush hours—are especially significant. During morning peaks, traffic predominantly converges toward the city center, while afternoon peaks are characterized by outbound flow. An intelligent control system should therefore be capable of detecting these directional patterns and prioritizing high-volume streams accordingly, minimizing delays and maintaining safe, fluid circulation throughout the network.

Beyond temporal awareness, efficient control requires two complementary coordination mechanisms. The first operates at the inter-agent level, where neighboring intersection controllers exchange information about local conditions such as queue lengths and demand fluctuations. This cooperative exchange allows agents to collectively adjust phase priorities in response to localized congestion, promoting network-wide adaptability. The second mechanism acts at the micro-control level, focusing on the dynamic adjustment of phase durations. When downstream capacity permits, green phases can be extended proportionally to approach density, increasing throughput and reducing waiting times. This functionality is central to the role of the SAPA module. Both mechanisms must nonetheless be applied with care to prevent demand from cascading into downstream intersections.

To evaluate the proposed framework under diverse conditions, three traffic demand strategies were defined, each implemented as a distinct neural network. The first is a balanced configuration in which total demand is distributed equally, 50/50, between the circular artery (C0–C1–C2) and the radial artery (C3–C1–C4). The remaining two strategies simulate peak-hour conditions by assigning 65% of demand to the radial artery and 35% to the circular artery. One replicates the morning peak, emphasizing inbound flow along the radial corridor, while the other captures the afternoon peak, prioritizing outbound movement. Together, these scenarios allow the system to be assessed under realistic directional traffic imbalances, as summarized in Table 2.

Table 2. Overview of the traffic control strategies evaluated in this study.

Network	Strategy	Priority Artery	Direction Focus	Description
1	Standard	None	W–E N–S	Arteries and directions treated equally.
2	Radial + Outbound Radial	Radial	N–S Northbound (S to N)	Morning peak
3	Radial + Inbound Radial	Radial	N–S Southbound (N to S)	Evening peak

The mobility model captures realistic urban traffic dynamics by accounting for vehicle speed, acceleration, deceleration, and lane-changing behavior, alongside environmental factors such as pedestrian flows, signal timings, road geometry, and variable demand levels, providing a comprehensive basis for evaluating the DRL agents and the SAPA module under dynamic and varied traffic scenarios.

For each strategy, two neural networks are trained using both DQL and MAPPO algorithms, and the performance of the resulting networks is compared. Specifically, we evaluate key traffic efficiency metrics, including vehicle queue lengths, pedestrians waiting, and phase activation patterns executed by the agents.

4. Results and Discussion

4.1. Real-Time Simulation of Vehicular Visible Light Communication Using SUMO

The proposed intelligent traffic management system leverages VLC technology to enable real-time monitoring and adaptive control of urban mobility under realistic communication conditions. To this end, a laboratory-scale VLC-based platform was developed to emulate a V2X communication environment, encompassing V2V and V2I/I2V links, as illustrated in Figure 1, while also considering its integration within a hybrid communication framework (VLC + RF). Importantly, this laboratory implementation serves as proof of concept, validating the feasibility and practical integration of the proposed approach.

Traffic scenarios are generated using the SUMO simulator, which provides essential parameters such as vehicle position, speed, queue length, and waiting time. These data are then encoded according to a dedicated communication protocol (Table 3) and transmitted via VLC through an interface with the LED controller of the laboratory setup. This configuration emulates optical data transmission through vehicle lighting systems, traffic signals, and public illumination infrastructure.

To reflect non-ideal operating conditions, the system design accounts for potential impairments such as signal attenuation (e.g., due to fog or rain), line-of-sight obstructions, ambient light interference, and packet loss or communication delays. In such cases, RF-based communication (e.g., IEEE 802.11p or 5G) is assumed to provide a complementary fallback channel, ensuring communication reliability.

The transmitted optical signals are captured by photodetectors within the laboratory prototype, emulating receivers embedded in both vehicles and roadside infrastructure. The recovered traffic information is subsequently processed by an intelligent agent within a MARL framework, which determines optimal control actions at each intersection to improve overall traffic flow efficiency.

Table 3. Message protocol defined for each of the V-VLC systems.

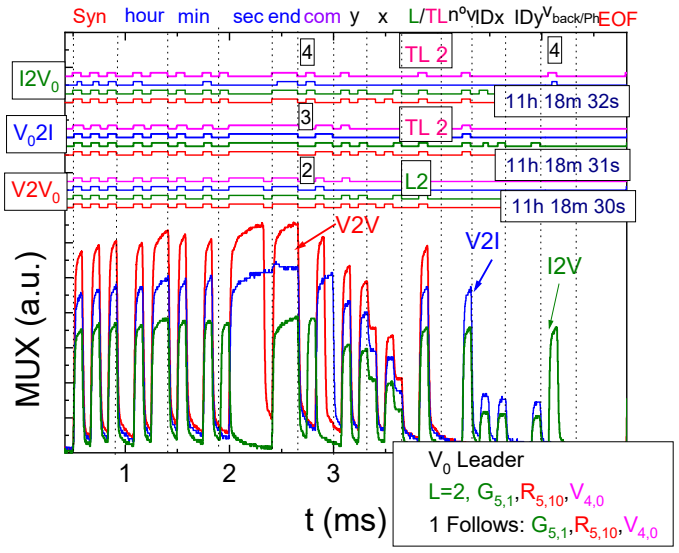
Type	Sync	Hour	Min	Sec	END	COM	Position		Lane/TL	ID/Info	Payload		EOF
L2V	Sync	Hour	Min	Sec	END	1	x	y	Payload				EOF
V2V	Sync	Hour	Min	Sec	END	2	x	y	Lane (0–7)	N° Veic	Car ID/ Behind		EOF
V2I	Sync	Hour	Min	Sec	END	3	x	y	TL (0–15)	N° Veic	Car ID/ Behind		EOF
I2V	Sync	Hour	Min	Sec	END	4	x	y	TL (0–15)	ID veic	Car ID/ Behind	Phase	EOF
P2I	Sync	Hour	Min	Sec	END	5	x	y	TL (0–15)	Direct	Payload		EOF
I2P	Sync	Hour	Min	Sec	END	6	x	y	TL (0–15)	Phase	Payload		EOF

Furthermore, the system assumes a heterogeneous traffic environment, comprising both VLC-enabled connected vehicles and non-connected vehicles. For the latter, baseline functionality is ensured through infrastructure-based sensing (e.g., traffic lights and detectors), allowing the system to maintain operational effectiveness even under partial connectivity scenarios.

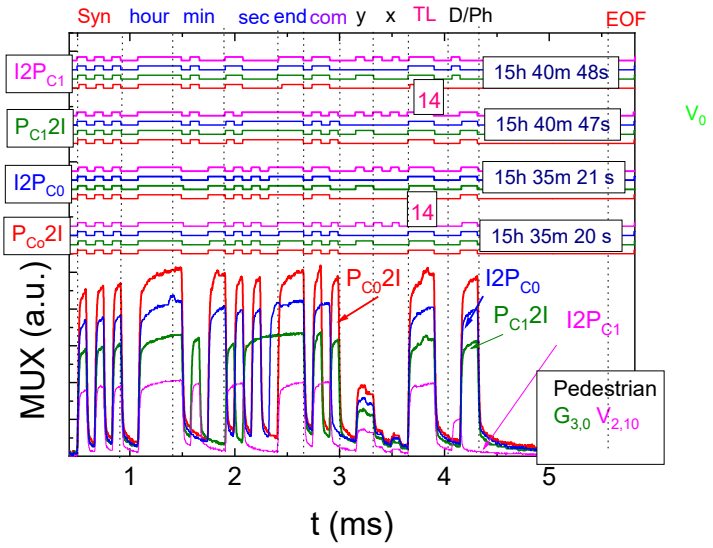
The communication protocol governs information exchange through a structured frame format comprising synchronization, identification, and payload fields. Each frame opens with a 5-bit synchronization block [10101] marking the Start of Frame, followed by a 12-bit TIME block encoding hours, minutes, and seconds. A [1111] flag signals the beginning of the ID blocks, each 4 bits wide, starting with the communication type code (L, V, P, or I). Subsequent fields specify transmitter location (x, y), lane (0–7), requested traffic light (0–15), number of following vehicles, assigned ID, cardinal direction, and active phase, varying according to whether the frame is a request or a response. Traffic-related payloads additionally carry vehicle identifiers, road conditions, waiting times, and weather data. Each frame concludes with a 4-bit End of Frame marker [0000].

Figure 5a,b illustrate the MUX signal and the decoded messages exchanged between vehicles, pedestrians, and traffic lights. The results confirm that VLC enables real-time tracking of V2I, V2V, P2I, and I2P communications across multiple intersections, establishing a structured coordination layer for both vehicular and pedestrian traffic. The integration of VLC with the MARL framework yields a decentralized yet coherent approach to intelligent traffic management. Vehicles continuously transmit position, velocity, and movement data to the infrastructure, alongside current signal phase information, enabling MARL agents to derive optimal control policies while maintaining phase synchronization across intersections. Pedestrians, in turn, submit crossing requests via VLC, receiving trajectory and phase assignments in response. These inputs are dynamically incorporated by the MARL agents to adapt signal timings, improving pedestrian crossing efficiency and reinforcing overall network safety.

Moreover, coordinated phase management is achieved through the synchronization of intersections via VLC data streams. MARL agents optimize signal transitions, mitigate conflict points, and improve the continuity of traffic flows across dense urban networks. In conclusion, VLC provides high-resolution, low-latency communication capabilities, while MARL facilitates adaptive, decentralized decision-making. The combined use of these technologies fosters safer, more efficient, and resilient intersections, advancing the state of intelligent urban traffic control systems.



(**a**)



(**b**)

Figure 5. MUX signal visualization and decoded communication frames for: (**a**) V2I and I2V exchanges; (**b**) P2I and I2P exchanges.

4.2. Performance Evaluation of DQL-SAPA and MAPPO Networks

This study aims to compare two distinct neural network configurations, each employing different learning approaches, by analyzing their performance across the three previously defined control strategies. The first configuration focuses on achieving arterial balance, requiring the network to distribute green phases in a way that equally satisfies traffic demands from both the N–S and W–E directions. The other two configurations implement radial priority, emphasizing dominant inbound and outbound traffic movements into and out of the city, respectively. These inbound and outbound patterns are simulated to represent morning and evening peak periods. The inbound scenario is modeled as increased traffic flowing from N to S, with a higher concentration of vehicles traveling from C3 to C4. Conversely, the outbound scenario reflects the opposite movement, from S to N, corresponding to traffic flowing from C4 to C3. The experimental setup is represented in Table 4. From a software perspective, the proposed system was developed and executed using Python 3.10.15 in conjunction with TensorFlow 2.10.0.

Table 4. SUMO simulation parameters for training and testing scenarios.

Parameter	Value
Number of vehicles	1800
Number of pedestrians	2000
Simulations Steps	3600
Number of Training Episodes	200

Figure 6 presents the number of vehicles queued along the radial artery comprising intersections C3–C1–C4. Each column corresponds to a different traffic control strategy, beginning with the arterial-balancing strategy, followed by the outbound-priority strategy, and concluding with the inbound-priority strategy.

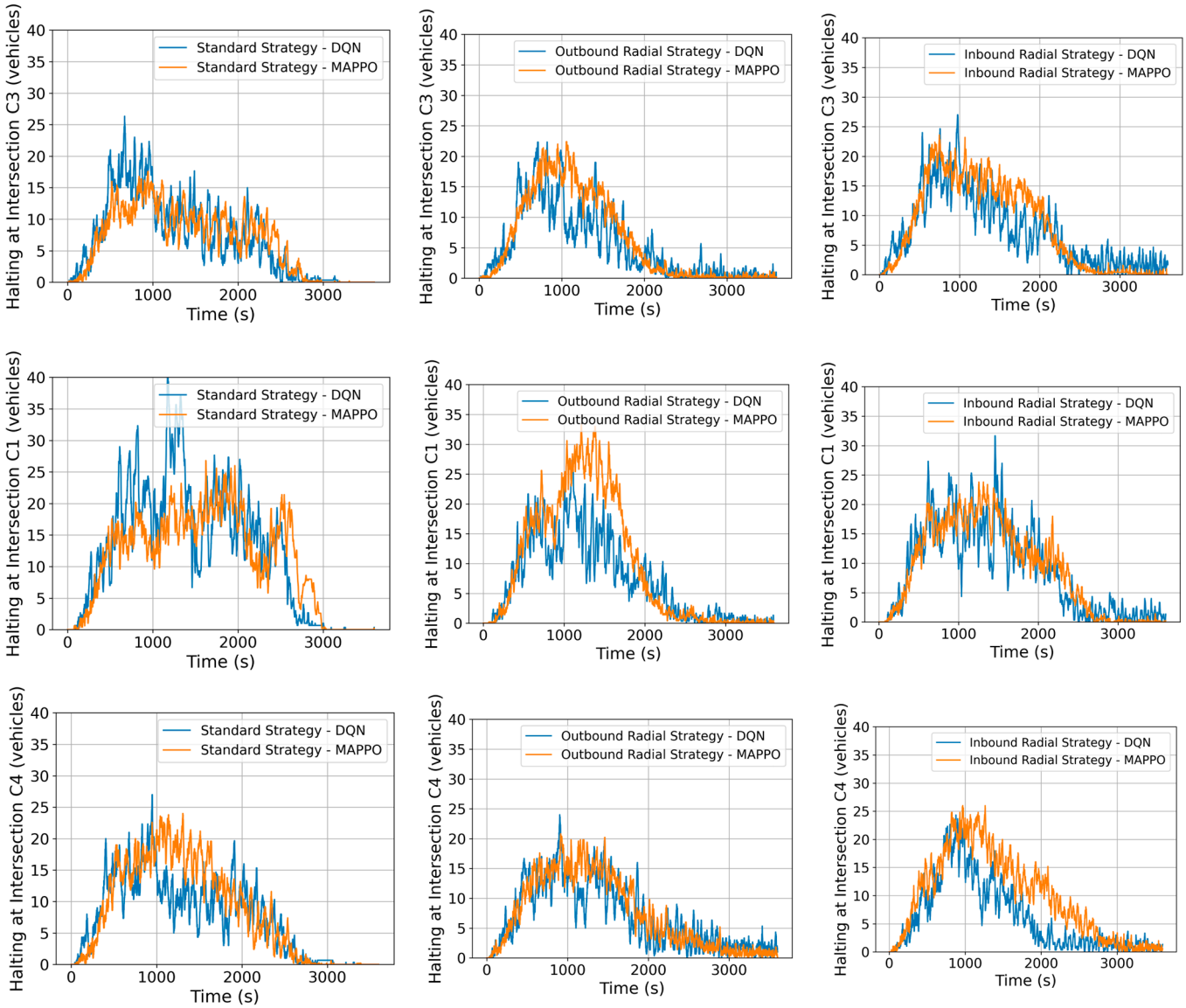


Figure 6. Vehicle halting along the radial artery at each intersection, for the strategies under consideration.

When comparing intersection C3 across the evaluated strategies, it is evident that the outbound approach yields the lowest number of vehicles at the intersection over time. This occurs because despite prioritizing the radial artery, there is a greater volume of vehicles traveling from C4 toward C3. As a result, the queue dissipates more rapidly compared

to the other two strategies, which exhibit a plateau at approximately 10 to 15 vehicles, respectively. Regarding intersection C4, all strategies display a similar trend, with vehicle queues gradually decreasing over time.

Intersection C1 represents the most critical point in the network, as it is geographically located at the convergence of both the circular and radial arteries. Because it must manage both flows, its role is to maintain mobility in this area. If the queue lengths begin to rise uncontrollably, congestion will propagate to adjacent intersections, severely disrupting overall traffic conditions. Based on the observed queue lengths across all strategies, however, this intersection consistently remains highly fluid, never reaching levels that would be difficult to manage.

When comparing the balanced strategy with the two radial-priority strategies, the balanced strategy exhibits a more persistent plateau of approximately 20 vehicles. This reflects the increased difficulty and, even if moderate, pressure is faced by the agent when controlling both arteries simultaneously.

Between the strategies with radial priority, it is clear that the inbound approach produces a higher plateau, around 18 vehicles, indicating that this intersection becomes more congested when inbound traffic to the city is prioritized, and more fluid when outbound traffic is prioritized.

Finally, when comparing the two neural network models, one using DQN-SAPA with dynamic phase timings and the other using MAPPO with fixed phase durations, their performance is generally well balanced. Still, SAPA shows a noticeable advantage in the strategies where radial priority is applied. Increasing the phase duration allows more vehicles to clear a given intersection, thereby reducing its queue; however, these vehicles subsequently contribute to increased halting at downstream intersections. This explains why the SAPA-based network exhibits greater oscillations, while MAPPO displays smaller variations due to its fixed timings. MAPPO ultimately performs better under the balanced strategy, where arterial equilibrium is required, as shorter and fixed phase durations appear beneficial for maintaining consistent flow across both arteries.

With respect to the circular artery, Figure 7 shows the halting values for intersections C0 and C2 across all strategies. When comparing the results obtained from the two neural network models, it is evident that MAPPO produces more controlled values, with oscillations of lower amplitude than those observed in the DQN-based network.

However, this difference must be interpreted in light of the phase duration associated with each model. As previously noted, the limited variability in MAPPO is directly linked to its fixed 8-s phase duration, which restricts traffic flow by allowing only a small number of vehicles to clear the queue at any given time.

The values shown in Figure 7 reflect not only the vehicles generated from the West and East directions at intersections C0 and C2, respectively, but also those arriving at C0 from C2. In the case of MAPPO, because the agent at C1 also operates under a fixed phase duration, only a small number of vehicles are released at a time into the C1–C0 link. Since this link has high capacity, the vehicles traveling through it do not significantly burden the agent at C0, as they are quickly discharged following a W–E phase activation.

In contrast, under DQN's dynamic phase durations, larger traffic volumes may flow from C0 to C1, from C1 to C2, and also in the opposite direction. This increased mobility results in a slight rise in the number of queued vehicles for the DQN model. Importantly, this increase does not indicate poorer performance; rather, it reflects higher mobility and, more importantly, a controlled level of fluidity within the circular artery.

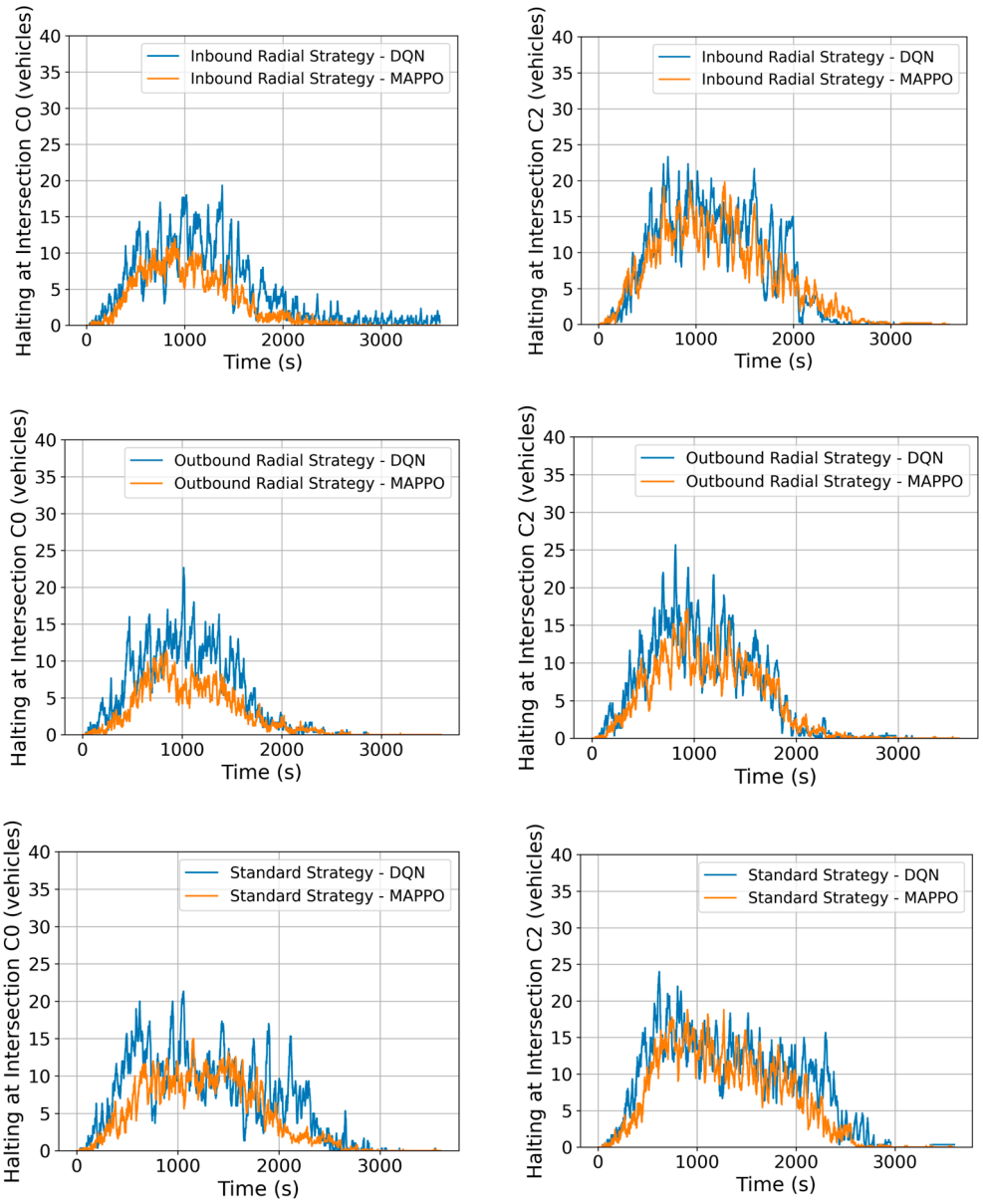


Figure 7. Vehicle halting along the circular artery at each intersection, for the strategies under consideration.

Regarding pedestrian flow in the environment, Figure 8 presents the pedestrian halting values at the critical intersection C1. Because this intersection lies at the convergence of multiple arteries, activating a pedestrian phase can be less advantageous for overall traffic conditions: when this phase is active, no vehicular movement is permitted through the intersection, thereby increasing pedestrian safety. During the 8-s pedestrian phase, followed by a 10 s clearance interval, vehicle queues inevitably grow. This increase must be carefully managed to ensure that traffic flow remains stable and the intersection does not reach critical congestion levels.

Based on the results, the number of pedestrians waiting to cross at the designated crossing area remains well balanced across all strategies. Furthermore, because the scenario assumes a low vehicle-density environment, there is greater opportunity to activate the pedestrian phase, as reduced vehicular pressure provides more flexibility for accommodating pedestrian flows.

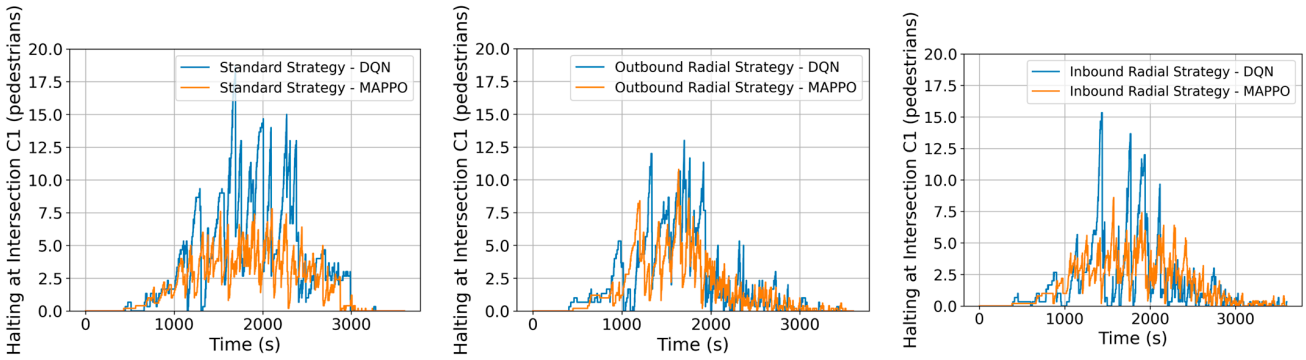


Figure 8. Pedestrian halting at intersection C1 for each of the evaluated strategies.

4.3. Agent Phase Activation

Given the different queueing dynamics and pedestrian behaviors observed across intersections for the various strategies, we now proceed to analyze the percentage of active phases selected by the agents at each intersection over time. This analysis allows for a direct comparison between the two neural network models, highlighting not only differences in phase activation patterns but also the distinct ways in which each network interprets and resolves the underlying traffic control problem.

Figure 9 presents the percentages of active phases over time for the five agents at each intersection under standard strategy, for both neural network models. The results show that the distribution of active phases aligns with the intended behavior of the strategy: an overall balance between Phase Ph1 (N–S) and Phase Ph5 (W–E), particularly at intersection C1. Intersections C0 and C2 exhibit a higher percentage of activations for Ph5, as they belong to the circular artery, whereas intersections C3 and C4 show a greater activation of Ph1, consistent with their placement along the radial artery. Pedestrian phases are activated with considerable regularity across all five intersections.

When comparing the two networks, MAPPO appears more selective in its phase activations, focusing predominantly on the stronger phases, Ph1 and Ph5, that accommodate movements in more than one direction. Phases Ph2, Ph3, and Ph8 are largely unused in this strategy. This selectivity enables more frequent activation of the phases that promote increased vehicular flow in their respective directions, consistent with the halting results observed in Figure 6. Even with fixed phase durations, MAPPO is able to match and at times exceed the performance of the DQN model operating with dynamic phase lengths.

In Figure 10, the analysis shifts to active phase patterns under the outbound radial strategy, where the phase activations again reflect the characteristics of the strategy.

This strategy prioritizes the radial artery while giving lower priority to the circular artery, with strong outbound traffic movements from C4 to C3. Consequently, Phase Ph1 is activated more frequently than Ph5. As more vehicles originate from C4, both networks also increase the activation of Phase Ph3 (S-All), thereby enhancing the S to N flow. This increase is particularly pronounced in the MAPPO network, reaching up to 42% activation at intersection C3. By controlling this phase, the flow of vehicles moving from N→S is regulated, forcing them to queue and preventing additional pressure on C1, which is already heavily loaded by traffic from C4. As a result, vehicles arriving from C1 at C3 can be dispatched more efficiently, given that this link is only 200 m long, half the capacity of the 400-m C4–C1 link. At C4, a substantial activation of Ph1, around 57%, is sufficient to manage the high volume of vehicles exiting the city.

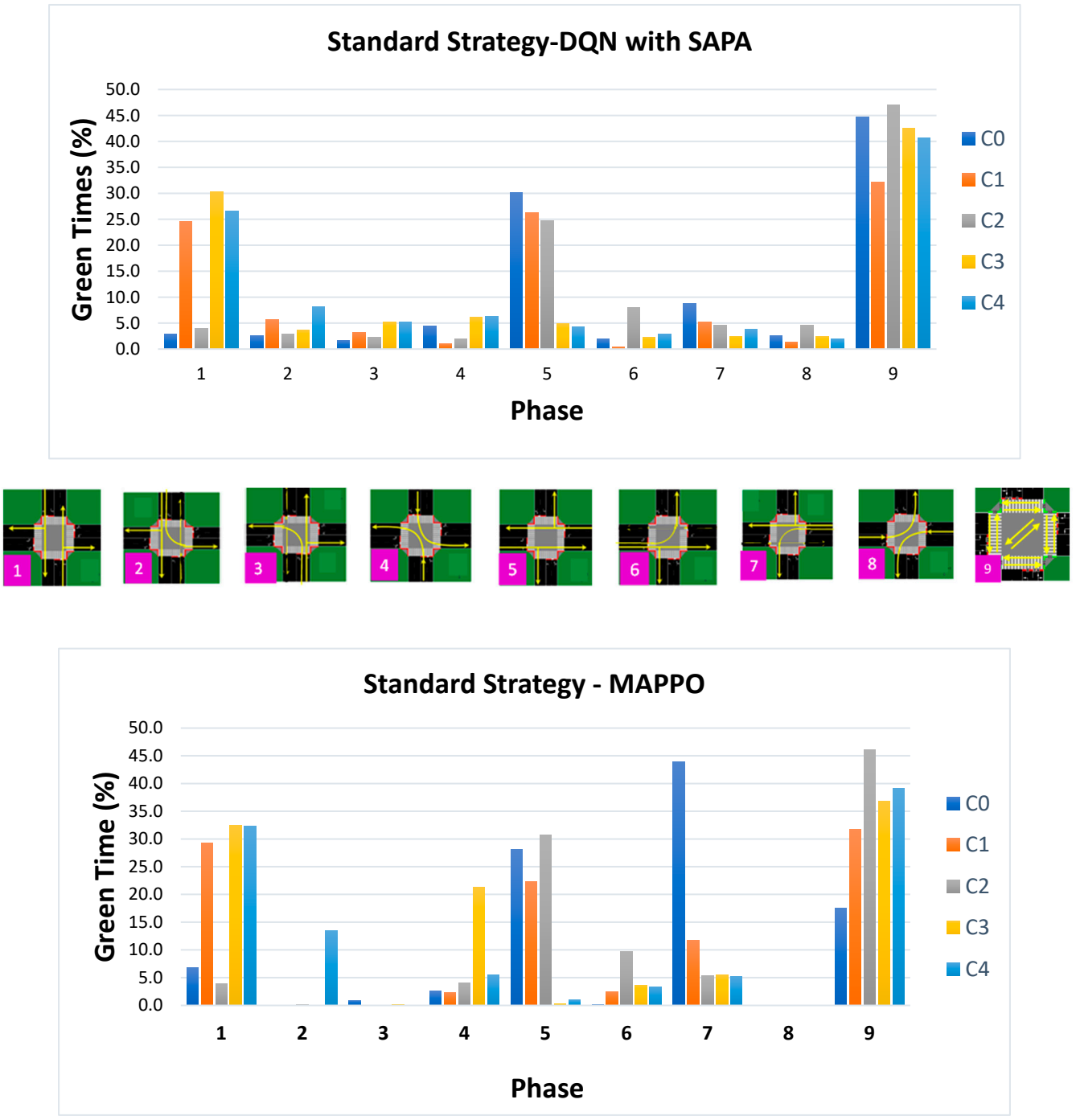


Figure 9. Percentage of active phases over time for the agents at their respective intersections under the standard strategy.

Regarding the pedestrian phase, its activation decreases considerably compared to the vehicular phases. Nonetheless, pedestrian flow remains well maintained, as also observed in Figure 8. Finally in Figure 11, we examine the inbound radial strategy, which, similarly to the previous strategy, exhibits a higher activation of Phase Ph1 (N–S) compared with Phase Ph5. In this case, however, the dominant movements in the radial artery correspond to inbound traffic, specifically from C3 to C4. This leads to an increased activation of Phase Ph2 (N–All) in the DQN network, most notably at intersection C4. The MAPPO network, however, adopts a different phase–activation pattern that not only satisfies vehicular demands but also accommodates the high mobility induced by the dynamic timings of the

DQN-SAPA network. Rather than reinforcing Ph2, MAPPO further increases the activation of Ph1, thereby maintaining stable flow in both the northbound and southbound directions.

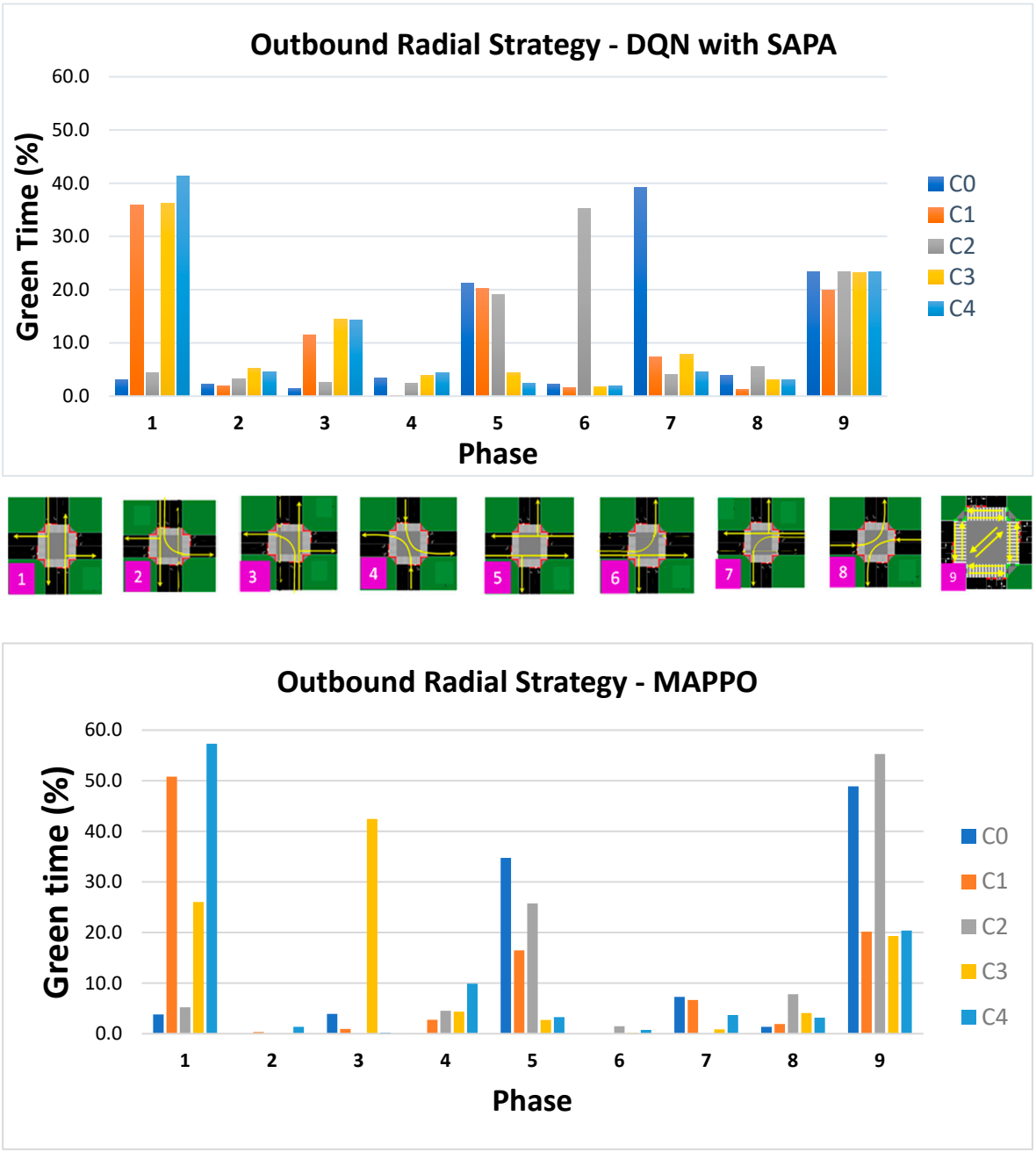


Figure 10. Percentage of active phases over time for the agents at their respective intersections under the outbound radial strategy.

Unlike the previous strategy, where activating Ph3 at intersection C3 was advantageous, the inbound strategy presents a more delicate situation. Because the link between C3 and C1 has a capacity of only 200 m, great care must be taken to prevent excessive congestion. If this link becomes fully saturated, C3 will experience substantial pressure, and congestion will propagate to C1 and C4, ultimately degrading the overall network performance. For this reason, the network selectively activates Phase Ph3 to allow vehicles on this short link to dissipate while simultaneously controlling the inbound flow from C3 to C4, thereby preventing renewed congestion along the 200-m C3–C1 segment, but now in the opposite direction. This reasoning explains why MAPPO does not significantly

increase the activation of Ph2 in inbound scenarios, as doing so would compromise fluidity on the C1–C3 link. It is worth noting that C3 still activates Ph3 approximately 12.4% of the time, ensuring that even with a strong influx of inbound traffic, saturation levels remain controlled.

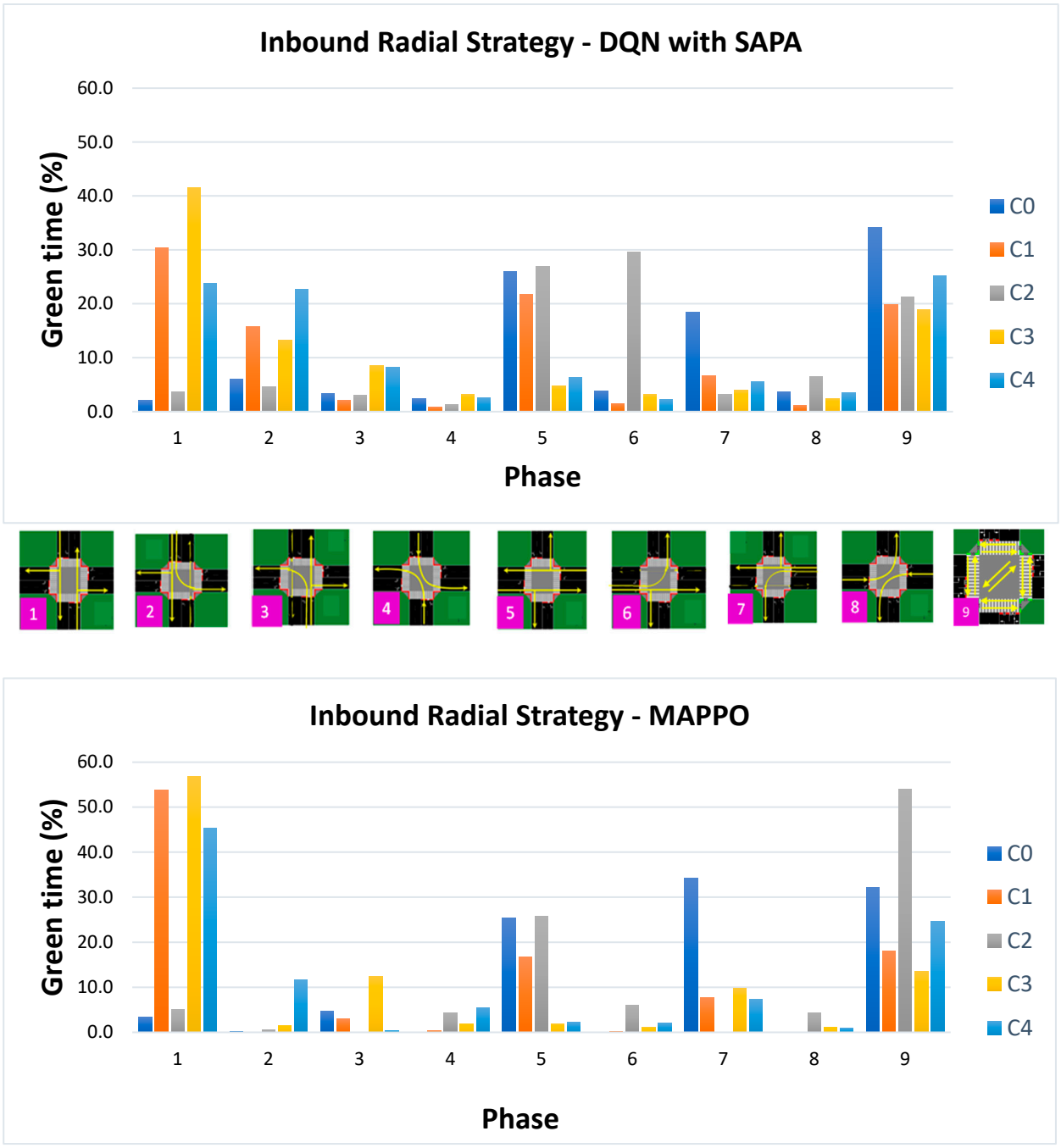


Figure 11. Percentage of active phases over time for the agents at their respective intersections under radial inbound strategy.

4.4. Ablation Analysis of the SAPA Mechanism

This ablation study compares two configurations of the DQN agent: a fixed phase duration of 8 s, representing the state-of-the-art baseline, and the proposed DQN-SAPA method, which introduces adaptive phase durations.

In the fixed-time configuration, each traffic signal phase lasts a minimum of 8 s, though the duration can extend if the agent selects the same phase consecutively. While this allows

frequent control updates, the short base duration limits the number of vehicles that can pass per green phase, leaving downstream capacity underutilized. From an environmental standpoint, the frequent phase switching promotes stop-and-go behavior, increasing fuel consumption and emissions.

In contrast, the DQN-SAPA method dynamically adjusts phase durations based on the number of vehicles waiting in each lane. Longer green phases allow more vehicles to clear the intersection, improving traffic flow, reducing queue lengths and waiting times, and increasing vehicle speeds. Crucially, this extension is applied in a controlled manner: the agent also considers the occupancy of the downstream segment, reducing the number of vehicles allowed to pass if that segment is congested, thereby preventing the propagation of congestion through the network.

As observed across all intersections, in Figures 12 and 13, the network incorporating the SAPA block consistently presents a lower number of waiting vehicles. The difference is particularly significant at intersection C1. At the intersections where vehicles enter the network (C0, C2, C3, and C4), the differences are less noticeable. This occurs because the SAPA mechanism also controls the extension of phase durations in order to regulate the number of vehicles moving toward the critical intersection.

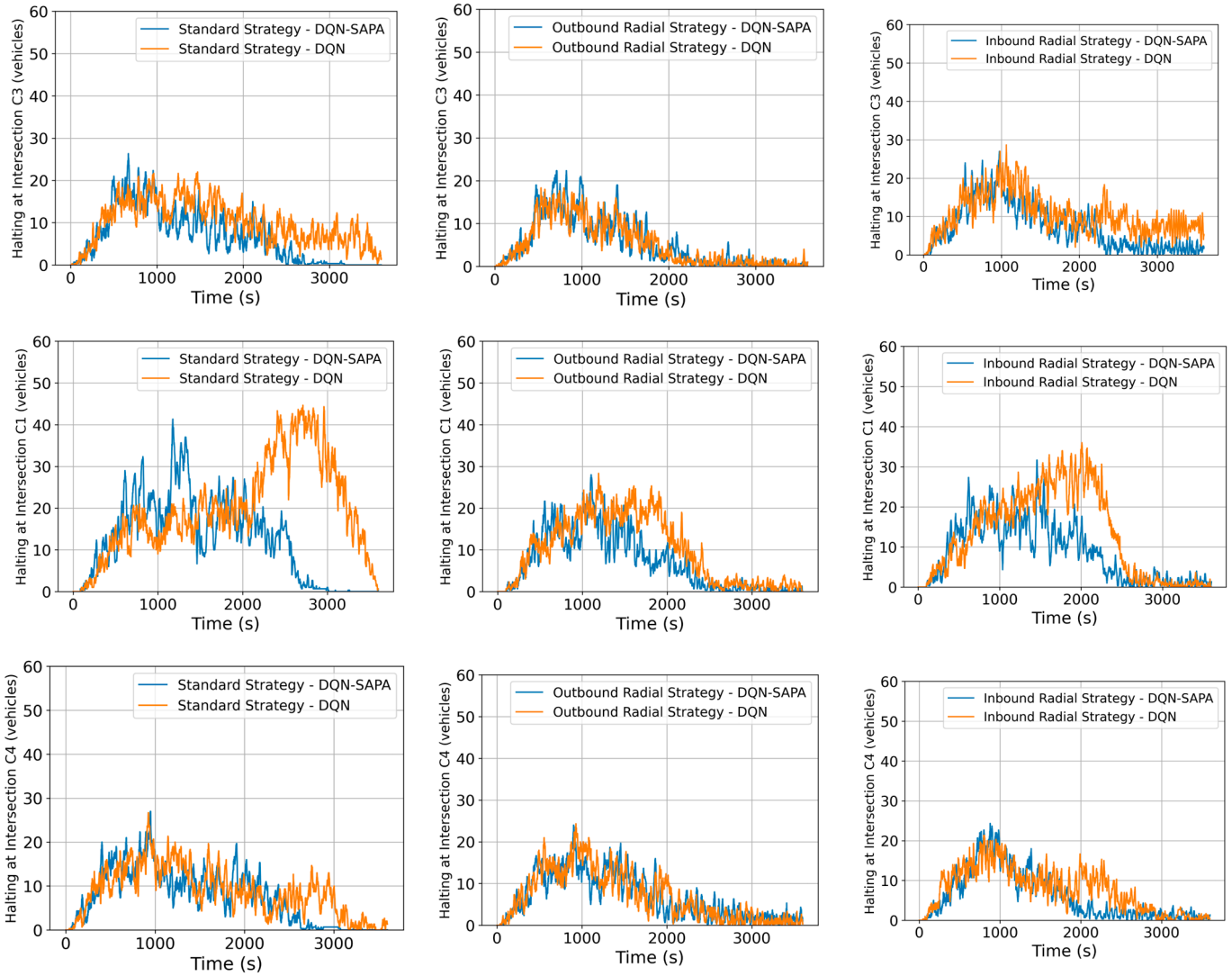


Figure 12. Vehicle halting along the radial artery at each intersection, for the strategies under consideration, comparing the state-of-the-art DQN with DQN-SAPA.

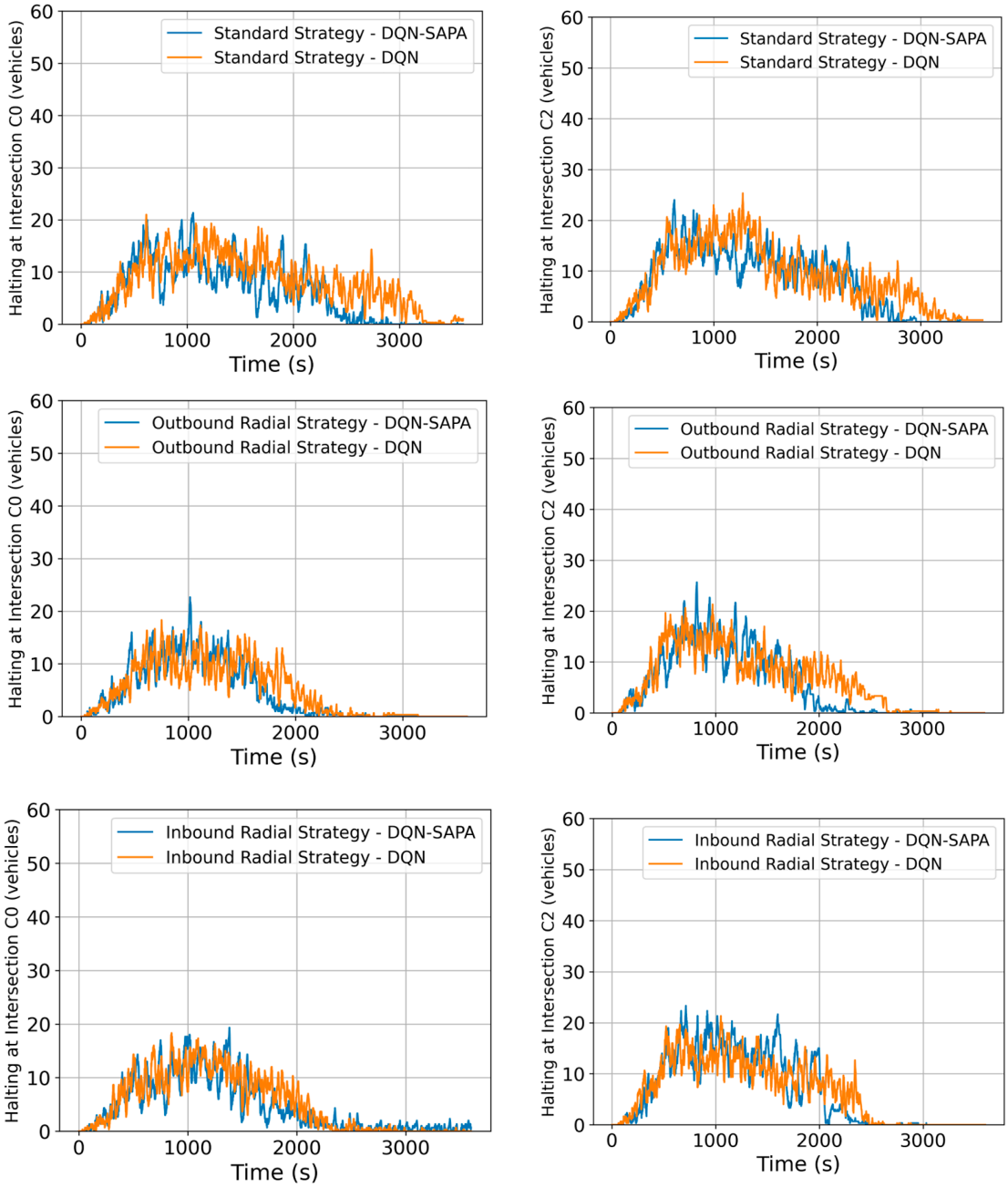


Figure 13. Vehicle halting along the circular artery at each intersection, for the strategies under consideration, comparing the state-of-the-art DQN with DQN-SAPA.

In contrast, at the critical intersection C1 the differences become more evident. For the DQN without phase time extension, the vehicle arrival rate exceeds the departure rate, leading to a gradual increase in the number of vehicles accumulating at the intersection over time. Conversely, with the DQN-SAPA approach, the system is able to dissipate traffic more effectively and in a controlled manner. This results in a higher capacity to process vehicles at the intersection, allowing the traffic load to decrease more rapidly over time.

Table 5 presents comparative metrics for intersection C1, which is the most critical intersection in the environment, for the standard, outbound, and inbound strategies. With respect to the DQN-SAPA model, it exhibits an increase in average speed of approximately 11% compared to the DQN model and an increase of about 6.11% relative to MAPPO, across the considered strategies. Regarding average waiting times, the values are largely similar.

Table 5. Traffic metrics at intersection C1 for the three considered strategies and models.

	Average Queued Vehicles (Vehicles)	Average Halting Pedestrians (Pedestrians)	Average Waiting Time (s)	Average Speed (km/h)
Standard Strategy				
DQN	19.3	7.2	11.5	29.6
DQN-SAPA	12.1	3.3	10.8	34.4
MAPPO	10.8	1.9	10.2	32.2
Outbound Radial Strategy				
DQN	9.7	7.5	7.8	34.6
DQN-SAPA	6.9	2.0	7.1	37.3
MAPPO	9.5	1.5	7.8	34.7
Inbound Radial Strategy				
DQN	12.8	9.6	7.1	32.9
DQN-SAPA	8.4	1.6	7.8	36.0
MAPPO	8.2	1.2	7.9	34.6

However, it is important to highlight that the implementation of the SAPA mechanism leads to a significant improvement in overall traffic flow across the environment. This can be observed through the analysis of the average number of vehicles in waiting queues, where the SAPA-enabled model achieves, on average across all scenarios, a reduction of approximately 33% in queue length compared to DQN. Similarly, pedestrian performance is also enhanced, as individuals experience shorter waiting times before their crossing phase is activated. In particular, the results indicate a reduction of approximately 70% in the average number of pedestrians waiting. Regarding the MAPPO approach, it can be observed that it achieves balanced performance levels comparable to the DQN-SAPA model, both in terms of the average number of vehicles in waiting queues and the number of pedestrians awaiting service. This indicates that the incorporation of the SAPA mechanism into the DQN framework significantly enhances its performance.

In particular, the results demonstrate that augmenting a value-based method such as DQN with the proposed SAPA module allows it to approach the performance of a more advanced policy-based multi-agent method like MAPPO. This finding highlights the effectiveness of SAPA as a lightweight yet impactful enhancement, capable of improving system efficiency without requiring the full complexity of more sophisticated multi-agent learning algorithms.

5. Conclusions

The proposed solution assumes a fully connected V2X environment enabled by VLC, in which vehicles communicate both with one another and with the surrounding infrastructure. Traffic information collected from this environment feeds a MARL system, which controls a traffic cell within the city and adapts signal phases to current traffic demands. Several traffic generation strategies were designed to reflect realistic urban mobility patterns, including arterial, circular, and radial flows, capturing both inbound and outbound movements across the city. Moreover, the SAPA module was introduced to dynamically adjust phase duration based on both the number of vehicles in the queue and the storage capacity of adjacent intersections.

To obtain the experimental results, two learning algorithms were compared: DQL (value-based) and MAPPO (policy-based). When comparing the two networks, we found

that MAPPO spends considerably less time activating weaker phases and, in some cases, avoids them entirely. Instead, it reallocates that time to stronger phases that are capable of maintaining a stable vehicle flow, while still matching the performance achieved by the DQL-SAPA network. However, MAPPO is computationally heavier due to its network architecture, which may limit the scalability of the system when controlling a larger number of intersections across the city. A promising direction for future work is to explore how the two networks could operate jointly, for instance, applying DQL to intersections with lower traffic relevance and deploying MAPPO in more critical locations, such as intersection C1.

These findings support the potential of the proposed intelligent control system as a scalable and adaptive solution for modern urban traffic management. However, the present work has certain limitations regarding real-world applicability. The results assume that all traffic information is reliably collected via VLC, without communication failures, interference, or data loss due to environmental factors. It is also assumed that all vehicles are equipped with VLC-enabled devices, which may not reflect actual penetration rates in real urban settings.

Consequently, the proposed approach would require additional mechanisms to handle heterogeneous communication capabilities and potential disruptions in data acquisition to ensure robust performance in practical deployments. In particular, robustness analyses under heterogeneous communication conditions and partial VLC penetration are important aspects that will be addressed in future work. These investigations are expected to provide further insights into the practical deployment of VLC-enabled multi-agent traffic management systems and to support the development of hybrid VLC/RF communication strategies for real-world scenarios.

Future work will also involve studying interactions between multiple traffic cells, where distinct strategies may be active in each one based on various traffic metrics such as time of day, queue lengths, and waiting times.

Author Contributions: Conceptualization, G.G., M.V. (Manuela Vieira) and M.A.V.; Formal Analysis, G.G., M.V. (Manuela Vieira), M.A.V. and M.V. (Mário Véstias); Investigation, M.A.V., M.V. (Mário Véstias) and P.L.; Methodology, M.V. (Mário Véstias); Software, G.G. and M.V. (Mário Véstias); Validation, M.V. (Manuela Vieira), M.A.V. and P.L.; Writing—Original Draft, G.G.; Writing—Review and Editing, M.V. (Manuela Vieira). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through the Research Unit CTS—Center of Technology and Systems (UID/00066). Additional support was provided by Fundação Santander Portugal and Instituto Superior de Engenharia de Lisboa (ISEL) under the project IPEX2025/INUV-LIGHT_ISEL. This work was also funded by national funds through FCT—Fundação para a Ciência e a Tecnologia, I.P., under the projects UID/06486/2025 (https://doi.org/10.54499/UID/06486/2025), UID/PRR/06486/2025 (https://doi.org/10.54499/UID/PRR/06486/2025), and UID/PRR2/06486/2025 (https://doi.org/10.54499/UID/PRR2/06486/2025).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Acknowledgments: The authors acknowledge CTS-ISEL and IPL.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ravish, R.; Swamy, S.R. Intelligent Traffic Management: A Review of Challenges, Solutions, and Future Perspectives. *Transp. Telecommun.* **2021**, *22*, 163–182. [CrossRef]
2. Gul, F. A review of control algorithm for autonomous guided vehicle. *Indones. J. Electr. Eng. Comput. Sci.* **2020**, *20*, 552–562. [CrossRef]

3. Ait Ouallane, A.; Bakali, A.; Bahnasse, A.; Broumi, S.; Talea, M. Fusion of engineering insights and emerging trends: Intelligent urban traffic management system. *Inf. Fusion* **2022**, *88*, 218–248. [CrossRef]
4. Zhang, J.; Wang, F.Y.; Wang, K.; Lin, W.H.; Xu, X.; Chen, C. Data-driven intelligent transportation systems: A survey. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 1624–1639. [CrossRef]
5. Soto, I.; Calderon, M.; Amador, O.; Urueña, M. A survey on road safety and traffic efficiency vehicular applications based on C-V2X technologies. *Veh. Commun.* **2022**, *33*, 100428. [CrossRef]
6. Michailidis, P.; Michailidis, I.; Lazaridis, C.R.; Kosmatopoulos, E. Traffic Signal Control via Reinforcement Learning: A Review on Applications and Innovations. *Infrastructures* **2025**, *10*, 114. [CrossRef]
7. Papageorgiou, M. Overview of Road Traffic Control Strategies. *IFAC Proc. Vol.* **2004**, *37*, 29–40. [CrossRef]
8. Memedi, A.; Dressler, F. Vehicular Visible Light Communications: A Survey. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 161–181. [CrossRef]
9. Shaaban, K.; Shamim, M.H.M.; Abdur-Rouf, K. Visible Light Communication for Intelligent Transportation Systems: A Review of the Latest Technologies. *J. Traffic Transp. Eng.* **2021**, *8*, 483–492. [CrossRef]
10. Yang, X.; Shi, Y.; Xing, J.; Liu, Z. Autonomous driving under V2X environment: State-of-the-art survey and challenges. *Intell. Transp. Infrastruct.* **2022**, *1*, liac020. [CrossRef]
11. Lowe, R.; Wu, Y.I.; Tamar, A.; Harb, J.; Pieter Abbeel, O.; Mordatch, I. Multi-agent actor-critic for mixed cooperative-competitive environments. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–12.
12. Yang, W.; Liu, H.; Cheng, G. Modeling and Performance Study of Vehicle-to-Infrastructure Visible Light Communication System for Mountain Roads. *Sensors* **2024**, *24*, 5541. [CrossRef]
13. Ramzi, S.R.; Hameed, S.M.; Sabri, A.A. VLC Tunnels. *Opt. Contin.* **2024**, *3*, 1990–2005. [CrossRef]
14. Al Hasnawi, R.; Marghescu, I.; Rusu-Casandra, A. Reliability and Capacity Evaluation for Vehicle-to-Vehicle VLC. In *Proceedings of the 2024 International Conference on Communications (COMM), Bucharest, Romania, 3–4 October 2024*; IEEE: New York, NY, USA, 2024; pp. 1–6. [CrossRef]
15. Liang, X.; Du, X.; Wang, G.; Han, Z. A Deep Reinforcement Learning Network for Traffic Light Cycle Control. *IEEE Trans. Veh. Technol.* **2019**, *68*, 1243–1253. [CrossRef]
16. Vieira, M.A.; Galvão, G.; Vieira, M.; Louro, P.; Vestias, M.; Vieira, P. Enhancing Urban Intersection Efficiency: Visible Light Communication and Learning-Based Control for Traffic Signal Optimization and Vehicle Management. *Symmetry* **2024**, *16*, 240. [CrossRef]
17. Chen, H.-C.; Li, S.-A.; Chang, T.-H.; Feng, H.-M.; Chen, Y.-C. Hybrid Centralized Training and Decentralized Execution Reinforcement Learning in Multi-Agent Path-Finding Simulations. *Appl. Sci.* **2024**, *14*, 3960. [CrossRef]
18. Ge, H.; Song, Y.; Wu, C.; Ren, J.; Tan, G. Cooperative Deep Q-Learning with Q-Value Transfer for Multi-Intersection Signal Control. *IEEE Access* **2019**, *7*, 40797–40809. [CrossRef]
19. Genders, W.; Razavi, S. Evaluating Reinforcement Learning State Representations for Adaptive Traffic Signal Control. *Procedia Comput. Sci.* **2018**, *130*, 26–33. [CrossRef]
20. Haydari, A.; Yilmaz, Y. Deep Reinforcement Learning for Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 11–32. [CrossRef]
21. Rasheed, F.; Yau, K.L.A.; Noor, R.M.; Wu, C.; Low, Y.C. Deep Reinforcement Learning for Traffic Signal Control: A Review. *IEEE Access* **2020**, *8*, 208016–208044. [CrossRef]
22. Oroojlooy, A.; Hajinezhad, D. A Review of Cooperative Multi-Agent Deep Reinforcement Learning. *Appl. Intell.* **2023**, *53*, 13677–13722. [CrossRef]
23. Luong, N.C.; Hoang, D.T.; Gong, S.; Niyato, D.; Wang, P.; Liang, Y.C.; Kim, D.I. Applications of Deep Reinforcement Learning in Communications and Networking: A Survey. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3133–3174. [CrossRef]
24. Lopez, P.A.; Behrisch, M.; Bieker-Walz, L.; Erdmann, J.; Flötteröd, Y.P.; Hilbrich, R.; Lücken, L.; Rummel, J.; Wagner, P.; Wießner, E. Microscopic Traffic Simulation Using SUMO. In *Proceedings of the 2018 International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018*; IEEE: New York, NY, USA, 2018; pp. 2575–2582. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.