sustainability

Article

Decentralized Multi-Agent Reinforcement Learning with Visible Light Communication for Robust Urban Traffic Signal Control

Manuel Augusto Vieira [1,2], Gonçalo Galvão [1], Manuela Vieira [1,2,3,*], Mário Véstias [1,4], Paula Louro [1,2] and Pedro Vieira [1,5]

1 DEETC-ISEL/IPL, R. Conselheiro Emídio Navarro, 1949-014 Lisboa, Portugal; manuela.vieira@isel.pt (M.A.V.); goncalo.galvao@isel.pt (G.G.); mario.vestias@isel.pt (M.V.); paula.louro@isel.pt (P.L.); pedro.vieira@isel.pt (P.V.)
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, R. Alves Redol, 9, 1000-029 Lisboa, Portugal
5 Instituto de Telecomunicações, Instituto Superior Técnico, 1049-001 Lisboa, Portugal
* Correspondence: mv@isel.ipl.pt

Abstract

The rapid growth of urban vehicle and pedestrian flows has intensified congestion, delays, and safety concerns, underscoring the need for sustainable and intelligent traffic management in modern cities. Traditional centralized traffic signal control systems often face challenges of scalability, heterogeneity of traffic patterns, and limited real-time adaptability. To address these limitations, this study proposes a decentralized Multi-Agent Reinforcement Learning (MARL) framework for adaptive traffic signal control, where Deep Reinforcement Learning (DRL) agents are deployed at each intersection and trained on local conditions to enable real-time decision-making for both vehicles and pedestrians. A key innovation lies in the integration of Visible Light Communication (VLC), which leverages existing LED-based infrastructure in traffic lights, streetlights, and vehicles to provide high-capacity, low-latency, and energy-efficient data exchange, thereby enhancing each agent's situational awareness while promoting infrastructure sustainability. The framework introduces a queue–request–response mechanism that dynamically adjusts signal phases, resolves conflicts between flows, and prioritizes urgent or emergency movements, ensuring equitable and safer mobility for all users. Validation through microscopic simulations in SUMO and preliminary real-world experiments demonstrates reductions in average waiting time, travel time, and queue lengths, along with improvements in pedestrian safety and energy efficiency. These results highlight the potential of MARL–VLC integration as a sustainable, resilient, and human-centered solution for next-generation urban traffic management.

Keywords: sustainable urban mobility; intelligent traffic management; multi-agent reinforcement learning (MARL); deep reinforcement learning (DRL); visible light communication (VLC); energy efficiency; pedestrian safety; smart cities

check for updates

Academic Editors: Evangelos Paschalidis and Socrates Basbas

Received: 11 October 2025
Revised: 1 November 2025
Accepted: 5 November 2025
Published: 11 November 2025

Citation: Vieira, M.A.; Galvão, G.; Vieira, M.; Véstias, M.; Louro, P.; Vieira, P. Decentralized Multi-Agent Reinforcement Learning with Visible Light Communication for Robust Urban Traffic Signal Control. *Sustainability* **2025**, *17*, 10056. https://doi.org/10.3390/su172210056

1. Introduction

The rapid growth of urban populations has intensified vehicle and pedestrian flows, exacerbating congestion, delays, and safety risks in cities. Effective traffic management is thus critical for sustainable urban mobility, aiming to balance efficiency, safety, and equity across diverse mobility demands. Conventional centralized traffic signal control systems, while widely used, face scalability challenges and struggle to adapt to dynamic traffic patterns, limiting their effectiveness in large, heterogeneous networks [1,2].

Urban traffic congestion poses significant challenges, including increased travel times, energy consumption, emissions, and pedestrian safety risks. Traditional traffic signal control methods are often inflexible and fail to adapt to dynamic, multi-modal traffic patterns. While Multi-Agent Reinforcement Learning (MARL) and Visible Light Communication (VLC) offer promising adaptive solutions, current approaches do not fully integrate MARL with real-time, high-fidelity communication to optimize both traffic efficiency and urban sustainability. This study addresses this gap by proposing a decentralized MARL–VLC framework that enables intersections to make autonomous, context-aware decisions, improving traffic flow, pedestrian safety, and energy efficiency.

Decentralized Multi-Agent Reinforcement Learning (MARL) approaches have been explored to overcome these limitations. Partially cooperative MARL agents share limited information with neighboring intersections [3], incorporating local and adjacent states into partially shared Q-value functions, enabling improved adaptability, reduced congestion spillover, and multi-objective control for efficiency and fairness. This introduces a level of cooperation without requiring global coordination, feedback-based timing optimization [4] and RL for multi-objective decentralized control. It improves adaptability compared to independent MARL, reduces local congestion spillover effect and enables multi-objective control (e.g., efficiency and fairness) [5]. However, it does not fully account for simultaneous actions of all agents.

However, most existing approaches do not fully capture simultaneous interactions among all agents, and pedestrian flows are often neglected.

Although recent works have begun to explore Multi-Agent Reinforcement Learning (MARL) for decentralized traffic signal control [6] and comprehensive surveys on MARL-based traffic signal control exist [7], the communication layer in these studies is often assumed to rely on generic wireless links (e.g., DSRC, C-V2X) without detailed exploration of optical communication technologies. On the other hand, Visible Light Communication (VLC) has been used in Intelligent Transportation Systems (ITS) to support vehicle-to-infrastructure communication [8] and recent reviews discuss hybrid VLC/RF systems and their challenges [9], but these works do not address integration with MARL. Our work differentiates itself by proposing a native integration of VLC with MARL, where distributed agents utilize VLC for low-latency, high-reliability local information exchange, overcoming limitations related to latency, interference, and dependence on RF networks. Furthermore, unlike VLC relaying approaches for ITS [10], our method incorporates adaptive control mechanisms and prioritization between vehicular and pedestrian flows within a decentralized MARL context.

Urban traffic congestion poses significant challenges, including increased travel times, energy consumption, emissions, and pedestrian safety risks. Traditional traffic signal control methods are often inflexible and fail to adapt to dynamic, multi-modal traffic patterns. While Multi-Agent Reinforcement Learning (MARL) and Visible Light Communication (VLC) offer promising adaptive solutions, current approaches do not fully integrate MARL with real-time, high-fidelity communication to optimize both traffic efficiency and urban sustainability. To address this gap, this study proposes a decentralized MARL–VLC framework that enables intersections to make autonomous, context-aware decisions.

The key contributions of this study are:

- Integration of MARL with VLC: enabling real-time, decentralized traffic signal control.
- Enhanced traffic performance: improving vehicle flow, pedestrian safety, and energy efficiency.

- Scalability and validation: demonstrated through realistic simulations, highlighting applicability to modern urban networks.

2. Related Work

This paper positions itself at the intersection of these two research directions. By combining MARL with VLC, we aim to advance the state of the art in intelligent traffic management, addressing open challenges related to scalability, robustness, and pedestrian–vehicle coexistence. The next subsection reviews related work in adaptive traffic control, multi-agent learning, and vehicular communication technologies, highlighting the gaps that motivate our approach.

2.1. Adaptive Traffic Signal Control

Adaptive traffic signal control has been extensively studied as a means to improve urban mobility by dynamically adjusting signal timings in response to traffic demand. In [11,12], its purpose was to explore the possibilities of applying Smart Traffic Lights to control traffic in large-scale road networks providing a speed harmonization and traffic prioritization between vehicles and pedestrians. Traditional systems rely on rule-based or optimization-based approaches, such as fixed-time plans, actuated control, and model-predictive control. While these methods can reduce delays under certain conditions, their reliance on predefined rules and global coordination often limits adaptability in highly dynamic or heterogeneous traffic environments. Furthermore, scalability becomes problematic when extending such systems to large urban networks with diverse mobility patterns.

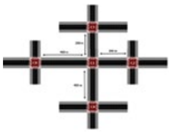
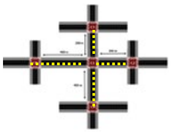
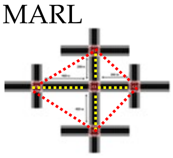
2.2. Reinforcement Learning Approaches: MARL-Based Traffic Signal Control (TSC)

Reinforcement Learning (RL) has emerged as a promising paradigm for traffic signal control, enabling agents to learn optimal policies through interaction with the environment. Early studies focused on single-agent RL, demonstrating improvements in local intersection performance. However, centralized RL approaches typically struggle to scale to city-wide networks due to computational complexity and communication overhead.

MARL has therefore gained traction, allowing distributed agents to coordinate locally while addressing scalability challenges. MARL extends single-intersection reinforcement learning control to a stochastic game environment, where multiple agents (intersections) interact simultaneously across arterial or regional traffic networks [13]. Communication plays a critical role in decentralized MARL, as agents must learn to exchange information using messages to better understand the system and achieve effective coordination. Deep MARL has been used to enable inter-agent communication by learning communication protocols in a differentiable manner [14]. The objective is to achieve a global equilibrium strategy that optimizes overall traffic flow. However, because agents learn and adapt their strategies at the same time, each one faces the challenge of pursuing a moving target [15]. This means that the optimal local strategy of one intersection is constantly influenced by the evolving strategies of its neighbors. Recent works have reported reductions in delays and queue lengths, yet many frameworks ne incomplete or noisy data, which can compromise robustness in real-world deployments.

In Table 1 a comparison of MARL approaches (description, advantages, disadvantages and representative works) for TSC is shown.

Table 1. Comparison of MARL Approaches for Traffic Signal Control.

Type of MARL	Description	Advantages	Disadvantages	Representative Works
Independent MARL 	Each intersection acts as an independent agent, making decisions based only on local states and actions, without coordination.	- Simple to implement - Low communication overhead. - Scales easily in deployment.	- No coordination among agents. - Suboptimal local policies. - Cannot manage network-level congestion.	Thorpe & Anderson (1996) [16]; Wiering (2000) [17]; Abdulhai et al. (2003); Jin & Ma (2015) [18]
Partially Cooperative MARL 	Agents share limited state information with adjacent intersections; partially shared Q-value functions guide decisions.	- More adaptive than independent MARL - Reduces local congestion spillover. - Enables multi-objective control.	- Limited to local cooperation - May lead to unstable/non-equilibrium strategies. - Not fully system-optimal.	Richter et al. (2007) [3]; As et al. (2008) [4]; Aleko et al. (2020) [19]
Joint Action MARL 	Intersections coordinate via joint states and joint actions, seeking a unique system-wide equilibrium.	- Strong inter-agent coordination - Better suited for global optimization - Handles complex multi-intersection dynamics	- Exponential growth of state-action space - High computational and communication costs - Poor scalability in large networks	Zhu et al. (2015) [20]; Medina & Benekohal (2014) [21]; Ma et al. (2024) [22]; Xu et al. (2024) [23]

Several challenges remain in the application of MARL to traffic signal control. Independent MARL approaches are simple to implement but ignore coordination among intersections, which often limits their effectiveness at the network level. Partially cooperative MARL introduces some degree of local cooperation, yet it struggles to achieve system-wide equilibrium. Joint action MARL, on the other hand, enables strong coordination but faces severe scalability and computational challenges, making it less practical for large urban networks. Moreover, most MARL-based studies on traffic signal control focus predominantly on efficiency metrics, such as travel time and waiting time, while critical aspects such as safety and multi-objective fairness remain underexplored. To overcome these limitations, recent research has begun to explore hybrid or hierarchical MARL models. In these frameworks, a central agent is responsible for learning global policies, while local agents execute optimized actions based on multiple reward functions that account for both efficiency and safety. Such designs offer a promising balance, improving scalability, coordination, and robustness in complex urban environments.

2.3. Vehicular Communication Technologies

Communication technologies are fundamental enablers of Intelligent Transportation Systems (ITS), providing the backbone for data exchange among vehicles, infrastructure, and control centers. Traditional approaches such as Dedicated Short-Range Communication (DSRC) and cellular-based Vehicle-to-Everything (C-V2X) have been widely studied, supporting Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications. Despite their maturity, these technologies face persistent challenges, including latency, interference, spectrum congestion, and high deployment costs, which may limit large-scale adoption.

VLC has recently emerged as a promising complementary solution. Leveraging existing LED-based infrastructure such as traffic lights, streetlights, and vehicle headlights, VLC enables dual functionality by combining illumination with high-speed data transmission [24,25]. This approach offers several inherent advantages, including high bandwidth, low latency, security, and cost-effectiveness. While VLC has been explored for vehicular communication and localization, its integration into adaptive traffic signal control frameworks remains underexplored. In this context, Vehicular VLC (V-VLC) systems have been proposed, combining mesh and cellular architectures to support efficient message relaying and edge computing.

The integration of connected vehicle (CV) technologies is revolutionizing urban mobility by enabling real-time communication between vehicles and infrastructure, thus enhancing traffic flow, reducing congestion, and improving safety. Emerging solutions, such as ATSC [26] connected and autonomous vehicles (CAVs) [27], and RL-based control, provide dynamic and intelligent traffic management. ATSC leverages sensor data to optimize signal timings, while CAVs can incorporate functionalities such as speed guidance and cooperative maneuvers to minimize queues and delays. However, these systems also present challenges, particularly the need for seamless interoperability, significant infrastructure investment, and scalability to large urban networks.

Advancements in wireless communication and V2V/V2I systems create new opportunities to couple traffic signal control with driving behaviors, leading to more holistic management of urban networks [28]. Building on this potential, this paper proposes a novel framework that integrates VLC-based localization services with RL-driven traffic signal control. The approach is designed to optimize both vehicular and pedestrian flows in multi-intersection scenarios, reducing waiting times while enhancing overall safety [29]. To evaluate the effectiveness of the proposed V-VLC system, agent-based simulations are conducted using the Simulation of Urban MObility (SUMO) [30]. In this framework, RL agents dynamically control traffic lights by maximizing long-term rewards, which are associated with minimizing delays and ensuring safe traffic interactions. Through iterative learning, the agent develops optimal control policies, illustrated by dynamic phase diagrams and state matrices based on accumulated waiting times.

2.4. VLC Impact and Limitations on Intelligent Transportation Systems

Visible Light Communication (VLC) is gaining traction as a complementary technology in the evolution of Intelligent Transportation Systems (ITS), supporting applications such as communication at signalized intersections, collision avoidance, vehicle localization, and platooning through Vehicle-to-Vehicle (V2V), Infrastructure-to-Vehicle (I2V), and broader Vehicle-to-Everything (V2X) communications [31]. While Radio Frequency (RF)-based communication remains the foundation of ITS, it faces notable limitations in dense urban environments. Challenges such as electromagnetic interference from electronic devices, spectrum congestion caused by high demand, and inherent security vulnerabilities can degrade RF performance. Its open nature also makes it susceptible to eavesdropping and cyberattacks [32].

By contrast, VLC transmits data through modulated LED light, offering immunity to electromagnetic interference, operation in an unlicensed and uncongested spectrum, and enhanced data privacy. The ability to reuse existing LED-based infrastructure—traffic lights, streetlights, and vehicle headlights—makes VLC a sustainable, low-cost, and energy-efficient alternative to traditional RF systems [32]. Unlike DSRC or 5G, it requires no additional spectrum allocation or costly hardware deployment, reducing both energy consumption and operational expenses.

When integrated with Multi-Agent Reinforcement Learning (MARL), VLC enables context-aware and decentralized traffic control. The high-precision, line-of-sight nature of VLC enhances situational awareness at the intersection level, reduces communication latency, and allows MARL agents to coordinate locally without reliance on centralized control. This results in a robust framework capable of maintaining performance under partial communication failures while ensuring equitable management of both vehicular and pedestrian flows.

However, VLC also faces inherent challenges in outdoor environments. Adverse weather conditions—such as fog, rain, or snow—and strong ambient light can degrade signal quality. In particular, direct sunlight may cause optical receiver saturation, leading to reduced communication performance. Therefore, VLC should be viewed not as a replacement but as a complement to RF-based communication. The combined use of both technologies can yield more secure, resilient, and sustainable ITS architectures.

In summary, while adaptive control and reinforcement learning have shown strong potential in optimizing traffic management, most existing solutions overlook pedestrian dynamics and robustness under incomplete information. The integration of MARL with VLC addresses these gaps by combining decentralized, learning-based decision-making with sustainable, low-latency, and secure communication—advancing the development of intelligent and human-centered urban mobility systems.

3. System Model and Problem Formulation

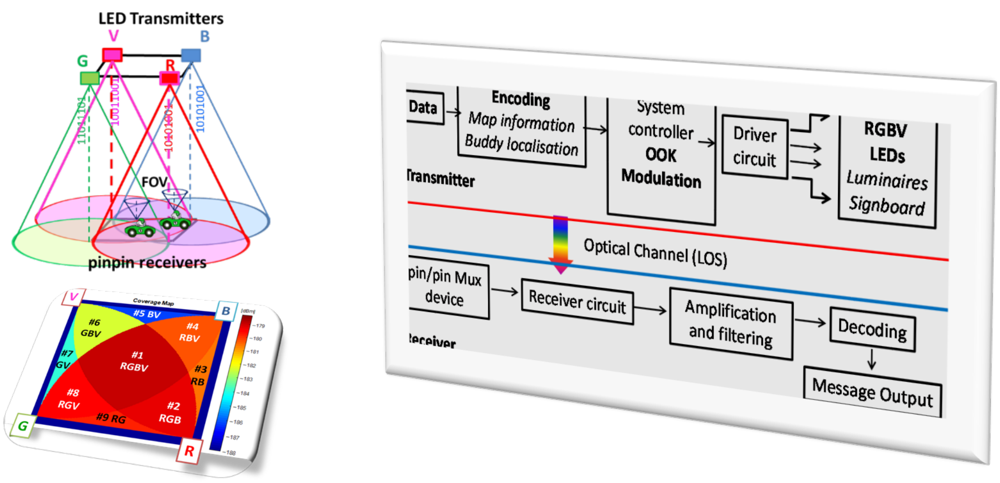
3.1. V-VLC Transmitters and Receivers:

The proposed system is based on LED-driven transmitters and PIN–PIN photodiode receivers that establish wireless visible light communication (VLC) links [25,26]. Each transmitter is composed of tetra-chromatic white LEDs (WLEDs) positioned at the corners of square unit cells, as illustrated in Figure 1a. Figure 1 shows the relative placement of V-VLC emitters and receivers, together with the coverage map and footprint regions of each unit cell (#1–#9).

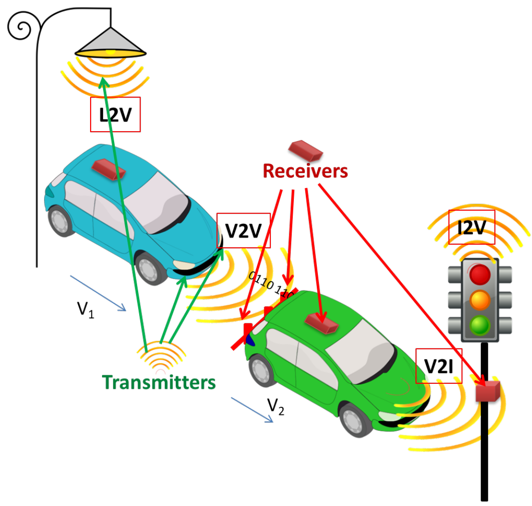
The WLEDs combine red (626 nm), green (530 nm), blue (470 nm), and violet (390 nm) chips to generate white illumination while simultaneously transmitting data. The optical signal is modulated using On–Off Keying (OOK) amplitude modulation, thus enabling a dual-function operation for both lighting and communication [33].

Each VLC transmitter includes four independent emitters capable of producing up to four simultaneous optical excitations, depending on the specific region of the unit cell. This configuration allows multiple signal combinations and distinct photocurrent levels at the receiver, providing fine-grained encoding capability. The receivers are aligned within overlapping transmitter coverage areas, forming multiplexed (MUX) optical signals. This structure operates simultaneously as a positioning system and a data communication channel (Figure 1a).

An orthogonal topology of geo-referenced transmitters is deployed to provide full coverage of the environment. Streetlights (L), spaced approximately 20 m apart, act as geo-transmitters that broadcast Light-to-Vehicle (L2V) messages containing identifiers, synchronization information, and traffic data. Each vehicle or pedestrian is assigned a unique ID. Figure 1b illustrates the spatial relationship between the V-VLC emitters and receivers.



(**a**)



(**b**)

Figure 1. (**a**) V-VLC Emitter and receivers' relative positions and illustration of the coverage map with the footprint regions in the unit cell (#1–#9). (**b**) Emitters' and receivers' locations.

On the receiving side, a dual PIN–PIN photodiode array detects light intensity variations generated by the OOK-modulated signals. Beyond signal detection, the photodiode array performs demultiplexing by separating the superimposed optical components. Through calibrated amplitude and wavelength mapping, it filters, decodes, and accurately reconstructs the original transmitted message, ensuring reliable data recovery [24]. This emitter–receiver configuration enables the VLC system to support high-capacity, real-time data exchange while maintaining the LEDs' primary role as illumination devices.

The integration of VLC facilitates direct monitoring among pedestrians, vehicles, and infrastructure, focusing on critical parameters such as queue formation and pedestrian density at intersections to enhance road safety. Peer-to-Infrastructure-to-Peer (P2I2P) communication enables real-time estimation of travel times, while instantaneous speed and waiting-time data are analyzed using the transmitters' tracking identifiers.

3.2. Connected Vehicles (CV) and Visible Light Communication (VLC):

Connected Vehicle (CV) technology is transforming urban traffic management by enabling real-time V2V and V2I communication. This capability allows continuous data exchange on speed, location, and traffic conditions, improving flow, safety, and congestion management. CV systems are thus essential components of next-generation traffic management frameworks.

In parallel, VLC has emerged as an innovative and complementary solution. By modulating the intensity of LED lights in traffic signals, streetlights, and vehicle headlights, VLC enables dual-purpose functionality: providing illumination while transmitting data. This integration into existing infrastructure offers several advantages, including high bandwidth, low latency, enhanced security, and cost-effectiveness.

The proposed Vehicular VLC (V-VLC) architecture leverages a mesh–cellular hybrid design, with streetlights serving as geo-transmitters (L2P/V) and traffic signals acting as edge-computing nodes (I2P/V), as illustrated in Figure 2. On the vehicular side, Vehicle-to-Infrastructure (V2I) and Infrastructure-to-Vehicle (I2V) communication are enabled through VLC links, allowing the continuous exchange of real-time data such as position, velocity, trajectory, and signal phase information. This communication layer supplies MARL agents with a reliable data stream, which is subsequently used to learn and refine optimal driving policies. In doing so, the system achieves synchronized signal phase control across multiple intersections, reducing congestion and improving network-wide efficiency.

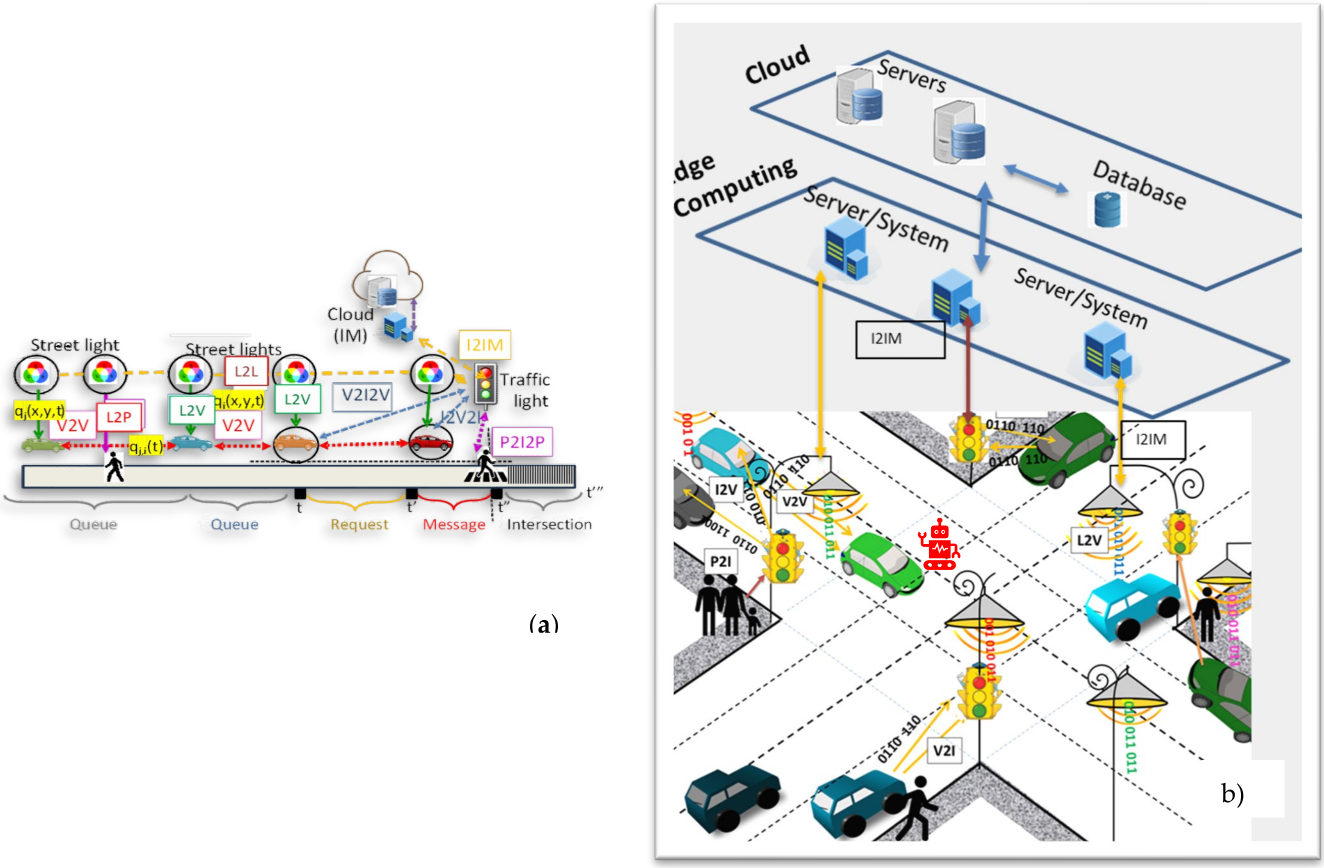


Figure 2. (**a**) Draft of the V-VLC architecture. Overview of visible light communication channels within the intersection area. (**b**) Illustration of the sensing transmission in road sensor networks. The diagram shows the three key components of the cooperative sensing framework: roadside infrastructure, sensing data aggregation, and sensing information delivery to CVs.

For pedestrian interaction, the architecture supports Pedestrian-to-Infrastructure (P2I) and Infrastructure-to-Pedestrian (I2P) communication. Pedestrians transmit crossing requests through VLC-enabled devices, and the infrastructure responds with trajectory assignments and safe crossing phase allocations. These interactions are directly integrated into the MARL decision-making process, allowing for dynamic adjustment of signal timings that not only prioritize traffic efficiency but also enhance pedestrian safety.

To manage intersection flow, a queue–request–response mechanism is employed [34]: approaching vehicles or pedestrians issue crossing requests (V/P2I), and traffic lights respond with acknowledgments (I2V/P), dynamically adjusting phases to avoid conflicts. Vehicle speed is calculated using the transmitter IDs for tracking and the mesh nodes estimate indirect V2V relative poses in scenarios with multiple neighboring vehicles [35]. Requests include positions, directions, and speeds, with leader-follower information aiding in subsequent V2I request confirmation. Delays are determined by observing the number of vehicles queuing in each cell at the beginning and end of green time through V2V2I, as shown in Figure 2 that depicts the various VLC-based communication links considered in the system.

Although VLC still faces challenges such as adverse weather conditions and interference from natural or artificial light sources, recent advances in photodetectors and optical filtering techniques have significantly mitigated these limitations. Moreover, VLC can seamlessly complement radio-frequency (RF) communication, ensuring robustness and continuity of service in scenarios where light-based transmission is partially impaired.

By integrating CV technology with VLC, the system enhances the situational awareness of both vehicles and pedestrians, enabling dynamic signal optimization that reduces delays and improves overall safety across multi-intersection urban environments.

At a broader level, coordinated phase management is achieved through intersectional communication. VLC data streams allow intersections to synchronize and exchange real-time state information, which MARL agents leverage to optimize phase transitions, minimize potential conflict points, and maintain smooth and continuous traffic flow throughout complex urban networks.

Beyond operational efficiency, this coordination framework contributes to sustainability by reducing vehicle idling time, fuel consumption, and greenhouse gas emissions typically associated with stop-and-go traffic patterns. The integration of energy-efficient VLC transmitters into existing lighting infrastructure further supports low-carbon operation, resource optimization, and the long-term sustainability of smart city ecosystems.

Overall, the proposed architecture combines the precision and low latency of VLC with the adaptive and self-learning capabilities of MARL, resulting in a distributed, resilient, and scalable traffic control ecosystem. This synergy enhances road safety, promotes energy-efficient mobility, and fosters the sustainable evolution of intelligent and environmentally responsible urban transport infrastructures.

3.3. Arterial Traffic Signal Control

Arterial traffic signal control involves intersections formed by the crossing of two or more main roads or arterials. Depending on the road network layout and geographic constraints, these intersections often exhibit complex geometries, such as T-shaped, cross-shaped, or skewed configurations [36]. The number of lanes at each intersection varies according to traffic volume and road capacity requirements.

Each approach typically includes multiple lanes to accommodate different turning movements, such as left-turn, right-turn, and through movements, where the schematic diagram of a four-arm junction with coded lanes (L/0–7) and traffic lights (TL/0–15) as illustrated in Figure 3. These intersections operate under standard traffic regulations. Approaching vehicles must yield to oncoming traffic and comply with traffic signals or signage. Movement priorities are generally determined by the installed control devices, including stop signs, yield signs, or traffic signals.

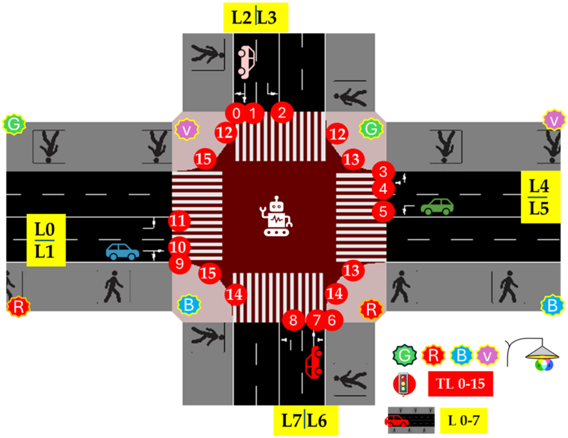


Figure 3. Schematic diagram of a signal-controlled standard intersection with coded lanes (L/0–7) and traffic lights (TL/0–15).

We consider a simplified urban traffic network composed of two intersecting arterial roads: a horizontal road (C0-C1-C2) and a vertical road (C3-C1-C4), which meet at the central junction C1. Each arm has two lanes. One to turn left, one to go straight or turn right optimized for CAVs. Intersection C1 serves as the unique connection point between these "horizontal" and "vertical" arteries. In the setup, C1 has no local sources of traffic—it only receives vehicles from the four adjacent intersections. All incoming flows into C1's lanes are determined by the phase-activation decisions of the neighboring intersection controllers (agents C0, C2, C3, and C4).

The Traffic scenario consisting of 5 homogeneous intersections with 4 arms each is illustrated in Figure 4. In effect, C1's role is to mediate the streams from the two arteries by influencing how and when those neighboring agents release traffic. This configuration makes C1 a central hub whose activity can substantially affect the overall network dynamics.

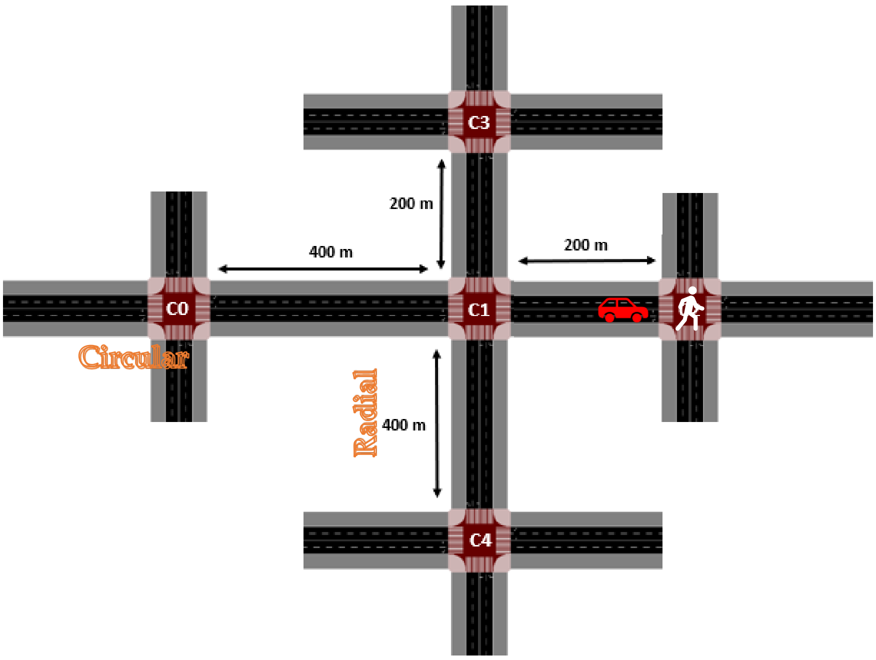


Figure 4. Traffic scenario consisting of 5 homogeneous intersections with 4 arms each.

In this case, C1's coordination might promote balanced traffic dispersal or alleviate congestion, but it could also introduce imbalances if misaligned. We therefore investigate how different priority schemes imposed at C1 affect traffic flow through the whole network. This configuration of five intersections is designated as a cell within the traffic management system.

Traffic signals at these intersections manage vehicle flow to ensure safe and efficient movements. Signal timing is designed considering factors such as traffic demand, intersection geometry, and operational objectives (e.g., minimizing delays and maximizing throughput). Signals are usually configured with multiple phases corresponding to different movements (e.g., green for through traffic and red for conflicting movements). In our scenario, each intersection is assigned four nine-phase traffic signal phases, as shown in Figure 5. The timing of these phases is coordinated to reduce conflicts and enable smooth traffic flow among the various vehicular movements.

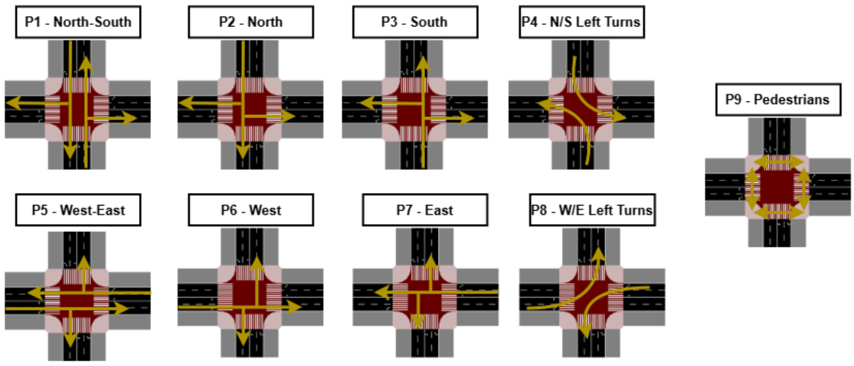


Figure 5. Traffic signal phases considered: eight for vehicles (P1–P8) and one exclusively for pedestrians (P9). The arrows represent the directions of vehicle movements across the eight vehicular phases, while P9 corresponds to the exclusive pedestrian phase at the signalized intersection..

4. Proposed Framework

4.1. Distributed MARL Agents & Deep Q-Network (DQN)

Figure 6 illustrates the architecture of the Intersection Manager (IM), which is composed of a decentralized neural network trained based on the observations and experiences of individual agents.

Each agent is responsible for controlling its own intersection, as depicted in Figure 6. This neural network enables real-time decision-making, dynamically adjusting the active signal phases according to the observed traffic flows on each approach, thereby optimizing traffic movement within the cell.

Each agent performs local observations of its corresponding intersection and makes decisions regarding which signal phases to activate based on the perceived traffic state. The experiences collected by the agents are stored in a centralized replay memory to support the training of the neural network. This neural network, responsible for controlling the cell, is trained under a specific traffic control strategy, allowing it to become effectively adapted to the traffic dynamics characteristic of that strategy. Considering five distinct strategies, a dedicated neural network is trained for each one, resulting in five fully adapted models. These models are subsequently compared to evaluate and analyze their behavioral and performance differences across varying traffic scenarios.

4.2. Multi-Agent Reinforcement Learning (Training Conditions)

In the proposed framework for intelligent traffic signal control, agent-based intersection management plays a central role. Each intersection is managed by a dedicated MARL agent that perceives its local environment through VLC-based data, including vehicle presence and pedestrian crossing requests. This localized perception enables context-aware decision-making at each junction. At the same time, cooperative learning ensures that agents share information with neighboring intersections, promoting coordination for consistent traffic flow across the wider network. Figure 7 displays the MARL Flowchart during simulation and training.

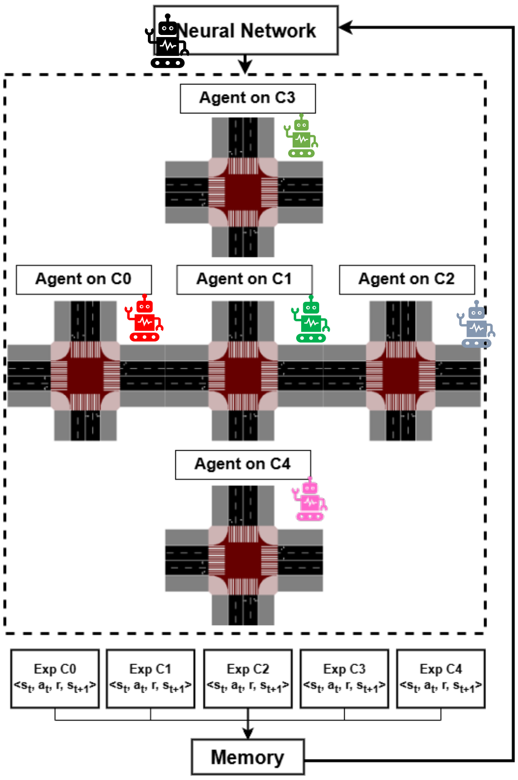


Figure 6. Intersection Manager architecture based on a Deep Neural Network.

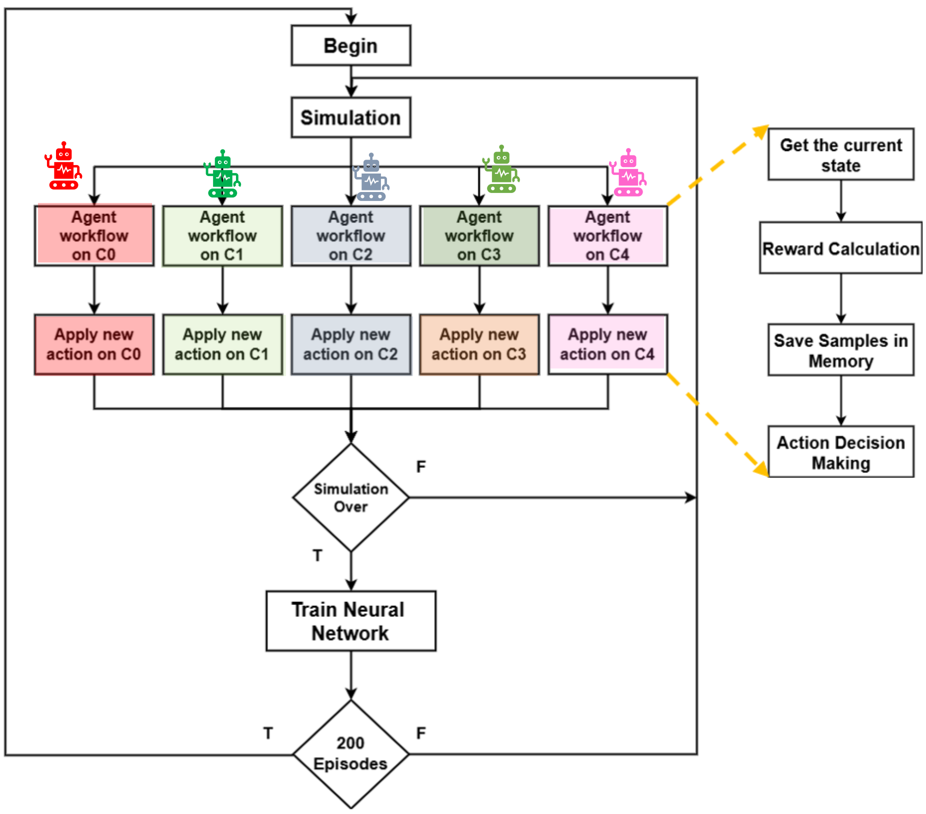


Figure 7. MARL Flowchart during simulation and training.

To optimize signal control, the system integrates a Deep Q-Network (DQN) approach, where all agents contribute to training a unified model. The DQN selects optimal signal phases based on expected cumulative rewards and dynamic traffic states, adapting to evolving conditions. The use of neural networks allows the framework to effectively manage high-dimensional urban traffic data, making the solution scalable and adaptable to large-scale and complex scenarios.

Each intersection is managed by a dedicated MARL agent that perceives its local environment—collecting data on vehicles and pedestrians via VLC-based communication—and cooperates with neighboring agents through shared information. The collected data is used to train a unified Deep Q-Network (DQN), which learns to select the optimal signal phase at each intersection based on expected cumulative rewards. Unlike traditional tabular Q-Learning, which is limited to small state-action spaces, the DQN leverages neural networks to handle the complexity of large-scale urban traffic environments.

The neural network architecture implemented consists of a fully connected layer network (FCLN), and the weights θk of the FCLN are used to approximate its Q-values Q (s, a; θk).

The state representation employed in the proposed architecture encodes the traffic environment through a combination of positional, velocity, and waiting cells. State representation and neural network architecture for traffic signal control is drafted in Figure 8. Specifically, the state vector comprises $4 \times 2 \times 10$ position cells and $4 \times 2 \times 10$ velocity cells, complemented by four waiting cells, resulting in a total of 164 input neurons. This input layer captures the environmental state at each intersection.

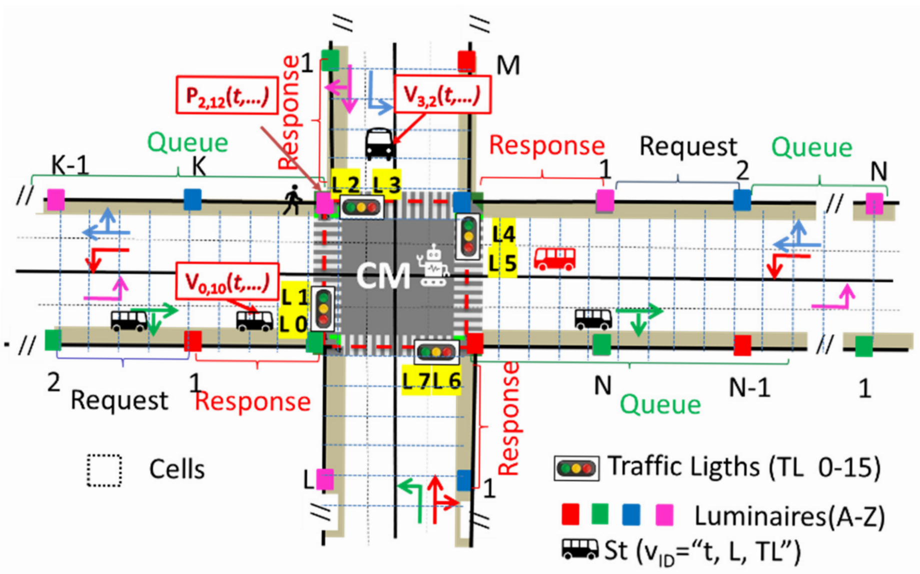


Figure 8. State representation and neural network architecture for traffic signal control.

The network architecture consists of five hidden layers, each comprising 400 neurons, with Rectified Linear Unit (ReLU) activation to introduce non-linearity and support the learning of complex traffic patterns. The output layer contains nine neurons, each corresponding to the Q-value of a possible control action.

Action selection follows a strategy in which the action associated with the maximum Q-value is executed. To account for interdependence among intersections, the Q-value calculation incorporates an additional term that aggregates predicted Q-values from neighboring intersections, thereby enhancing coordinated decision-making. Each neural network is trained with traffic patterns tailored to its specific control strategy, and vehicle generation is biased to reflect target priorities (e.g., East–West or North–South flows). This ensures that each agent adapts its policy to the unique characteristics of the traffic management scenario it controls.

Additional details regarding the network architecture and training configuration are provided in Table 2.

Table 2. Parameters of the neural network architecture and training configuration.

Parameter	Value
Episodes	200
Max Steps	3600
Vehicles	1800
Pedestrians	2000
Width Layers	400
Batch Size	100
Learning Rate	0.001
Loss Function	MSE
Training Epochs	500

In this work, the algorithm employs two neural networks with similar architecture: one responsible for predicting the Q-values, and another, referred to as the Q-target network, which calculates the target Q-values (Equation (1)). The Q-value function incorporates the influence of neighboring intersections by adding a term that aggregates the predicted Q-values from these neighbors. The Q_{target} values are calculated based on Equation (1).

$$Q_{target} = r_t + \Upsilon \cdot \max\left[Q_{pred}\left(s_{t+1}, a'\right) + \beta \cdot \frac{1}{N}\sum_{n}^{N} Q_{pred}\left(n_{t+1}, a'\right)\right] \quad (1)$$

where Q_{pred} is the Q-value predicted by the main network and Q_{target} is acquired using a network similar to the main one but which is not trained, γ is a discount factor applied to the $maxQ_{target}$ value, lowering the importance of the future reward compared to the immediate reward and N denotes the number of neighboring intersections considered, and β is a weighting factor that regulates the influence of these neighbors on the Q-value update. r_t is the reward (Equation (2)).

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

The reward used considers both vehicle and pedestrian accumulated waiting times ($atwt$).

$$atwt_{veh,t} = \sum_{veh=1}^{n} wt_{(veh,t)}$$

$$atwt_{ped,t} = \sum_{ped=1}^{n} wt_{(ped,t)} \quad (3)$$

$wt_{veh,t}/wt_{ped,t}$ 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. n represents the total number of vehicles/pedestrians in the environment in t.

A fair value of β promotes cooperation that benefits not only the individual agent but also its neighbors, fostering a coordinated global traffic control strategy. In this work, β was set to 0.3, as other values were tested and resulted in poorer performance. Higher β values encourage more cooperative decisions, benefiting neighbors but potentially at the expense of the individual agent's own performance, while a lower β favors more independent, locally optimized actions.

The implemented method follows a learning approach where agents share experiences to train a global neural network that selects actions for all intersections. This strategy is feasible due to the homogeneity of the intersections, allowing similar observations across agents to be leveraged collectively. Furthermore, by incorporating neighbor influence into the learning process, the approach promotes coordination between adjacent intersections, enhancing scalability and adaptability within the network.

This setup demonstrates how the integration of MARL and VLC enables scalable, adaptive, and coordinated traffic signal control across a connected urban road network.

Overall, this approach combines local perception and global cooperation to deliver intelligent, data-driven traffic signal decisions. The result is an efficient solution that operates in real time within multi-agent urban environments, addressing both the complexity and dynamism of modern city traffic.

4.3. MARL System with Dynamic Phase Duration: SAPA

The traffic control framework adopts a decentralized Multi-Agent Reinforcement Learning (MARL) system, where each agent manages its own intersection by observing local conditions, selecting active phases, and storing experiences. Since intersections are homogeneous, these experiences can be shared across agents to train a common neural network. This approach enhances adaptability compared to fixed-cycle systems, as it allows signal phases to respond to real-time traffic conditions.

A key improvement is the introduction of dynamic phase duration, supported by data collected through VLC on vehicle queues and lane occupancy. Unlike fixed durations (e.g., 8 or 12 s), phase times are adjusted based on both the number of waiting vehicles and the occupancy of the receiving lanes at neighboring intersections. This micro-level control prevents blockages and optimizes throughput, forming the basis of the Strategic Anti-Blocking Phase Adjustment (SAPA) mechanism.

For example, in a scenario with five intersections, traffic from C0 to C1 (connected by a 400 m link) is managed by monitoring occupancy. If occupancy is below 40%, the green phase is extended proportionally to the queue length; otherwise, only a minimum green is granted, allowing downstream intersections time to clear. Similar rules apply when traffic moves from a 400 m link into a shorter 200 m link, where thresholds are reduced to 35% to avoid overflow.

The SAPA system also incorporates traffic strategies that assign different priorities to radial and circular arteries. A standard strategy balances flows equally (50–50), while others prioritize one artery with up to 65% of total vehicles, reflecting scenarios such as strong outbound or inbound movements. Low-priority phases (e.g., left turns or minor flows) receive reduced weighting (25%). The most critical intersection (C1) dynamically balances both radial and circular flows according to the active strategy.

By combining phase selection via neural networks with adaptive phase duration through SAPA, the system achieves more efficient management of vehicle and pedestrian flows across multiple intersections, ensuring scalability and robustness under diverse traffic scenarios.

4.4. Simulation and Setup Parameters

The SUMO simulations were conducted using a realistic urban scenario based on a selected area of downtown Lisbon, Portugal. Vehicular and pedestrian mobility patterns were derived from validated models described in [37,38], which demonstrated a high correlation between simulated and real-world traffic data in Lisbon's central districts.

The traffic environment includes both vehicle and pedestrian flows, with parameters adjusted to match observed densities and dynamics. Vehicle volumes were set to 1800 veh/h, reflecting typical peak-hour conditions, while pedestrian flows were configured to represent realistic corner densities and crosswalk usage. The β parameter, in

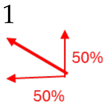
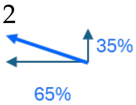
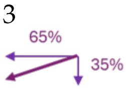
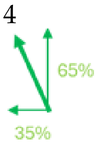
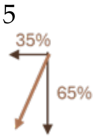
Equation (1), was calibrated through iterative simulation adjustments based on the specific topology and traffic patterns of the selected area, resulting in β = 0.3.

This configuration ensures that the simulation scenario accurately represents realistic urban mobility, allowing for the evaluation of the MARL–VLC framework under conditions that closely mimic actual traffic behavior. This setup ensures that the simulations closely mimic real urban traffic behavior, allowing meaningful evaluation of the MARL–VLC framework. Moreover, by optimizing traffic phase transitions and reducing vehicle idling, the framework supports energy-efficient traffic flow, decreases fuel consumption, and lowers greenhouse gas emissions, contributing to more sustainable urban mobility.

4.5. Traffic Control Strategies

Five distinct traffic control strategies were designed and implemented, each represented by a separate neural network (agent) (Table 3). These strategies differ in how they bias the allocation of green phases across intersections, reflecting different priorities between the circular (horizontal) and radial (vertical) arteries, as well as between inbound and outbound flows relative to the central intersection (C1). The objective is to compare a balanced approach with schemes that emphasize one road or traffic direction.

Table 3. Traffic Control Strategies.

Network	Strategy Name	Artery Prioritization	Directional Priority	Description/Expected Effect
1 	Balanced (50–50%)	None	↔ All directions	Equal treatment for circular (blue) and radial (green) arteries; aims for fairness and overall throughput.
2 	Circular + Outbound Radial	Circular	Radial → Outbound (S→N)	Favors east–west traffic on circular road and northbound flow on radial axis.
3 	Circular + Inbound Radial	Circular	Radial ← Inbound (N→S)	Supports inbound vehicles from the north and east–west circular movement.
4 	Radial + Outbound Radial	Radial	Radial → Outbound (N→S)	Enhances north–south circulation and reduces congestion along the radial axis.
5 	Radial + Inbound Radial	Radial	Radial ← Inbound (S→N)	Prioritizes inbound flow toward the city center, combined with east–west movements.

Each control strategy is defined by a specific assumption on urban traffic demand. Strategies may prioritize either the circular road or the radial arteries, and further distinguish between outbound (from the center) and inbound (toward the center) flows. These priorities were implemented in the simulation by adjusting the vehicle generation process. For example, and considering an total traffic demand of 1800 vehicles, in all strategies under consideration, 75% of the vehicles (1350 vehicles) proceed straight ahead or make right turns, while the remaining 25% (450 vehicles) correspond to left-turn movements.

In Network 2 and 3 of the 1350 vehicles that proceed straight or turn right, 65% (878 vehicles) are generated on the circular artery, whereas the remaining 35% (472 vehicles) originate from the radial artery. To represent inbound and outbound city movements, traffic generation on the radial artery is deliberately unbalanced: 75% of these 472 vehicles

(354 vehicles) are produced in the south–north direction (outbound flow) at intersection C4, with the remaining 25% (118 vehicles) generated at intersection C3, or conversely in the case of inbound flows.

Similarly, in Networks 4 and 5, the same distribution logic applies, but with 65% of the 1350 vehicles (878 vehicles) generated on the radial artery. Again, this generation is unbalanced to simulate inbound and outbound flows, such that 75% (659 vehicles) are generated at C4 and 25% (219 vehicles) at C3. The circular artery, in turn, accommodates the remaining 35% of the 1350 vehicles (472 vehicles).

The opposite distribution defined Networks 6–9. An equivalent procedure was applied to the radial artery, corresponding to a 90° clockwise rotation of this directional bias. This ensured that each trained network adapted its policy to the intended control strategy. Figure 9 summarizes the set of traffic management strategies implemented in the simulation environment.

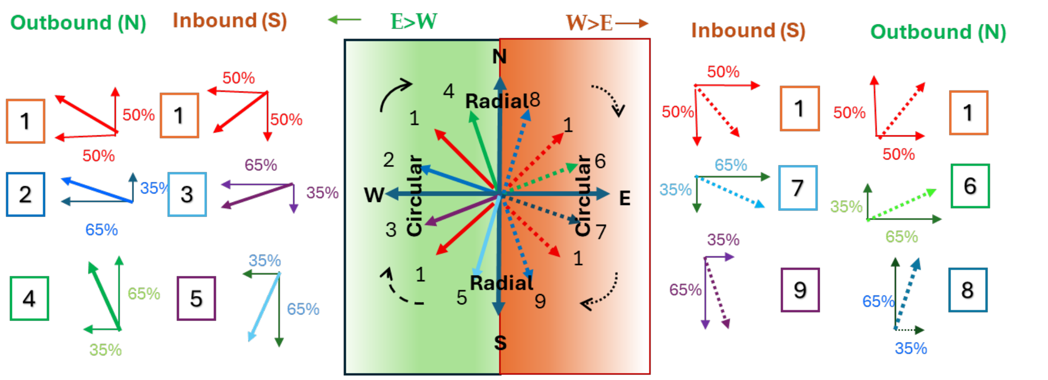


Figure 9. Summary of the simulated traffic control strategies, illustrating prioritization between circular and radial directions.

The experiments considered traffic flows of 1800 vehicles per hour and 2000 pedestrians per hour, simulated over 300 episodes of 3600 s each.

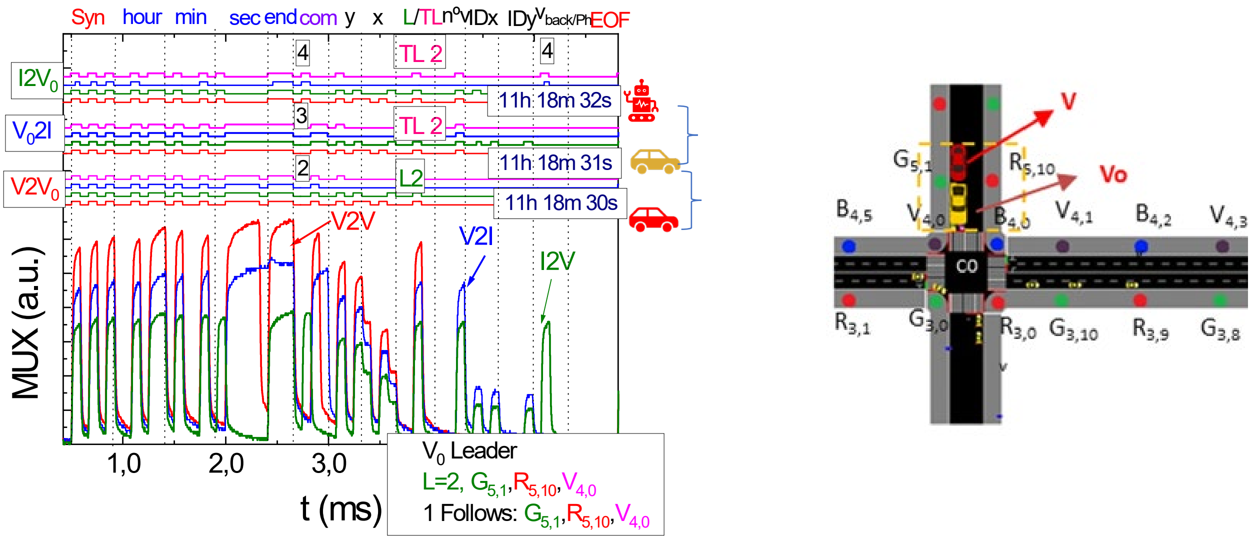
The performance of these strategies will be assessed by analyzing their impact on MARL training and evaluation. In particular, we compare the final cumulative rewards obtained by each network and traffic fluidity indicators such as average waiting time and throughput. This comparison will reveal which prioritization schemes enhance or hinder circulation, and whether the central intersection (C1) can act as an effective global coordinator in this arterial traffic scenario.

5. Results and Discussion

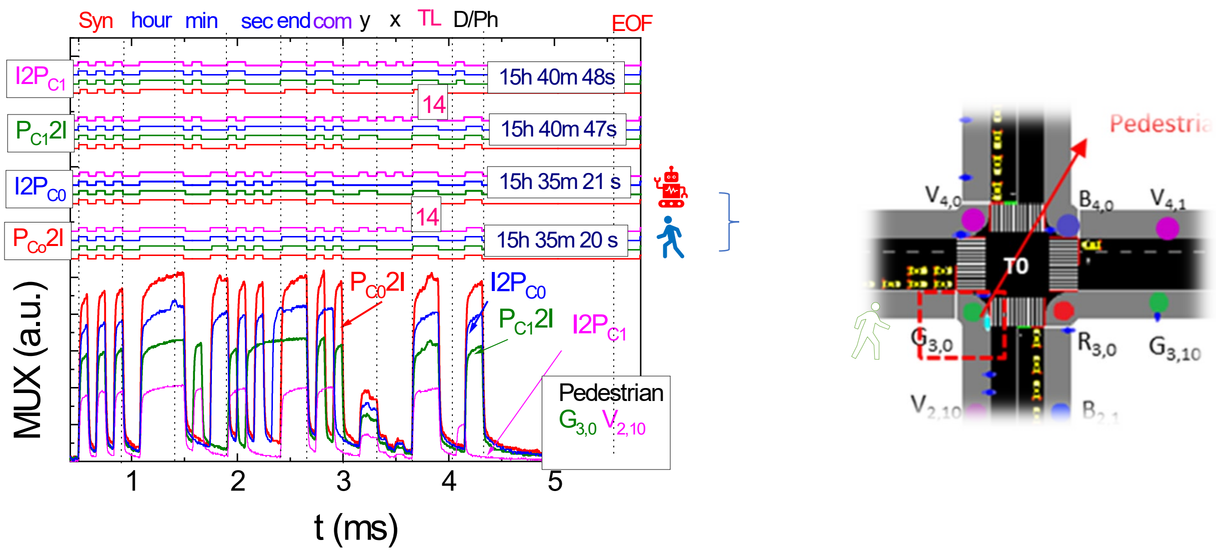
5.1. Adaptive Traffic Control: V-VLC Communication Protocol and Evaluation

The communication protocol defines the rules for information exchange, structured in a frame with synchronization, identification, and payload fields. Each frame begins with a 5-bit synchronization block [10101] marking the Start of Frame (SOF), followed by a 12-bit TIME block encoding hours, minutes, and seconds. A flag [1111] indicates the start of ID blocks, each 4 bits, beginning with the code of the communication type (L, V, P, I). Subsequent fields specify transmitter localization (x, y), lane (0–7), requested traffic lights (0–15), number of following vehicles, assigned ID, cardinal direction, and active phase, depending on whether the message is a request or response. For traffic-related messages, additional payload data includes vehicle identifiers, road conditions, waiting times, and weather. The frame ends with a 4-bit End of Frame [0000], signaling completion. Figure 7 demonstrates the MUX signal and the decoded messages between the vehicles/pedestrians and the traffic lights, respectively [26]. The visualization of MUX signal exchange and the decoded message flow between Vehicles and Traffic Lights (V2I, I2V), as well as between

Pedestrians and Traffic Lights (P2I, I2P), is presented in Figure 10. An inset on the right has been added to better illustrate the decoded signals.



(**a**)



(**b**)

Figure 10. MUX signal exchange visualization and decoded message flow between: (**a**) Vehicles ↔ Traffic Lights (V2I, I2V) and (**b**) Pedestrians ↔ Traffic Lights (P2I, I2P).

Results show that with VLC is possible to details in real time the flow of V2I, V2V, P2I, and I2P communications at various intersections, illustrating a structured communication framework for coordinating traffic and pedestrian movement.

The integration of VLC with MARL establishes a decentralized yet harmonized framework for intelligent traffic management. In this approach, vehicles continuously exchange real-time information with the infrastructure, including position, velocity, and movement dynamics, as well as current traffic signal phases. This bidirectional communication enables MARL agents to infer optimal driving policies while ensuring synchronization of signal phases across multiple intersections.

Simultaneously, pedestrians interact with the infrastructure by transmitting crossing requests through VLC, to which the system responds with trajectory and phase assignments.

MARL incorporates these inputs to dynamically adapt signal timings, thereby enhancing the efficiency of pedestrian crossings and strengthening overall 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.

5.2. MARL Training Approach: Performance Based on Rewards

Figure 11 analyses the cumulative negative rewards over MARL training episodes at C0–C4, under the three generation outbound strategies.

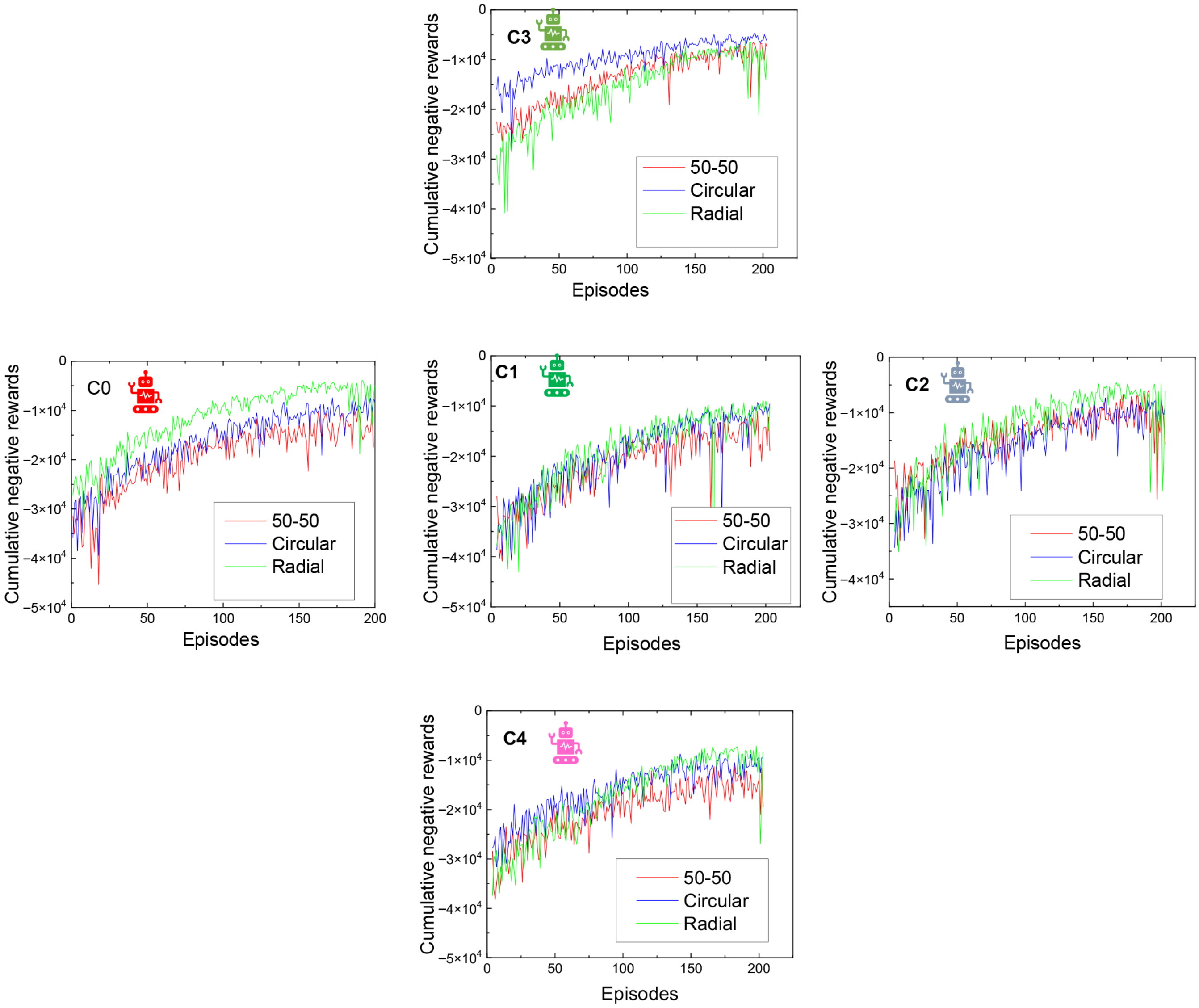


Figure 11. Evolution of negative cumulative rewards during MARL training, with the horizontal axis representing epochs, from intersections C0 to C4. Results are shown for Network 1, representing a balanced topology; Network 2 (Circular + Outbound Radial), and Network 4 (Radial + Outbound Radial).

In training, each of the networks learns to interact with the environment, each dealing with a specific traffic generation, corresponding to the strategy defined. With respect to the cumulative negative rewards, it is evident that each of the trained neural networks successfully converged and adapted to its respective strategy, effectively optimizing the traffic flow within the simulated environment. The evaluation of the MARL training strategies demonstrates how different traffic generation biases influence the overall system performance. In the balanced strategy (Network 1), results consistently show lower rewards across most intersections, suggesting greater difficulty in traffic coordination when no directional priority is established. This outcome indicates that evenly distributed flows tend to amplify congestion effects, as waiting times accumulate uniformly across all approaches.

By contrast, Network 2—which reduces vehicle generation in the radial artery (35% along the west–east axis and 25% along the north–south axis)—achieves higher rewards. The improvement is primarily associated with a lower number of vehicles approaching the intersections, leading to shorter waiting times and more efficient signal utilization.

Among the tested strategies, Network 4 consistently showed the most favorable trends in the simulations. It achieved higher cumulative rewards at circular intersections (C0, C1, C2), and the reduced priority for the west–east direction contributed to lower vehicle accumulation and shorter waiting times. These observations suggest that, under increased traffic demand along the radial artery, Network 4 may provide a more efficient distribution of traffic. While these results are based on simulation trends rather than formal statistical analysis, they offer useful insights for guiding adaptive traffic signal strategies in similar urban networks.

These findings confirm that MARL agents adapt effectively to the imposed traffic conditions, and that prioritization strategies, rather than balanced ones, produce superior performance in terms of reward optimization. This demonstrates the importance of directional bias in enhancing the learning process and overall system efficiency.

5.3. MARL Testing Approach: Vehicle and Pedestrian Halting Times

In training, each of the networks learns to interact with the environment, each dealing with a specific traffic generation, corresponding to the strategy defined. During the network evaluations, traffic metrics such as halted vehicles and pedestrian flow showed promising results.

In Figure 12 the analysis of C0–C4 testing halting metrics across different outbound networks (Networks 1, 2, 4) are presented.

Each network learned a specific control strategy and, when tested, demonstrated the ability to effectively manage traffic, even during periods of high congestion. The models are capable of selecting the most suitable actions based on real-time traffic conditions at the intersection, thereby optimizing the flow of both vehicles and pedestrians. The analysis reveals that vehicle halting values remain relatively consistent across intersections C0, C2, C3, and C4, regardless of the adopted strategy. This uniformity suggests a stable performance of the control approaches throughout the majority of the network. In contrast, intersection C1 emerges as the most critical node, exhibiting the highest variation in halting times. This finding underscores the strategic role of C1 in the overall efficiency of traffic coordination within the network.

Starting in the urban core, strategy 4 prioritizes radial flow (e.g., south-to-north). A shift to strategy 1 in the next cell introduces more balance, and finally strategy 2—further from the center—emphasizes circular flow again. This strategic progression enables efficient redistribution of traffic, reducing congestion and improving overall flow.

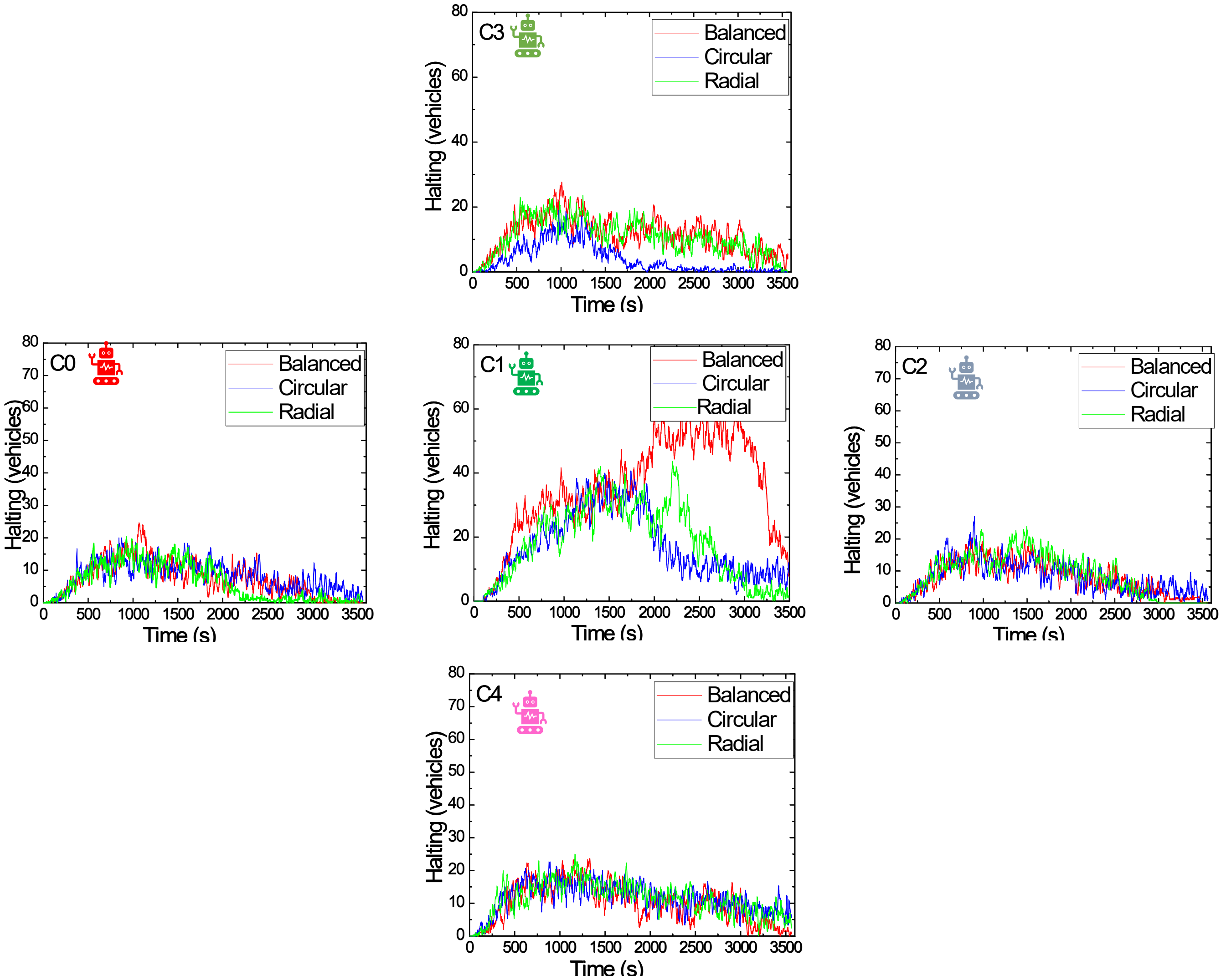


Figure 12. Assessment of halting vehicle behavior over time at intersections C0–C4 across multiple control strategies. Results are shown for Network 1, representing a balanced topology; Network 2 (Circular + Outbound Radial); and Network 4 (Radial + Outbound Radial).

For Network 1, which applies a balanced control strategy, halting times at C1 are substantially higher compared to the other strategies. The absence of prioritization mechanisms at this intersection increases the complexity of traffic management, leading to longer queues and reduced operational performance at this critical point. Nevertheless, Network 1 demonstrates the capacity to sustain balanced halting values across the remaining intersections, indicating that its performance limitations are highly localized to C1 rather than systemic across the network.

Figure 13 display a comparative analysis of pedestrian halting metrics at intersections C0–C4 across the different outbound network strategies.

The analysis of pedestrian halting behavior indicates that the size of pedestrian halting peaks provides a meaningful proxy for the stress level experienced at crossings and reflects pedestrians' response to the interaction with connected vehicles. Overall, halting values remain balanced across the five networks, suggesting that all strategies are capable of integrating pedestrian flows without causing excessive delays. Some temporal fluctuations are observed, primarily associated with the activation of pedestrian phases, yet these do not significantly compromise performance.

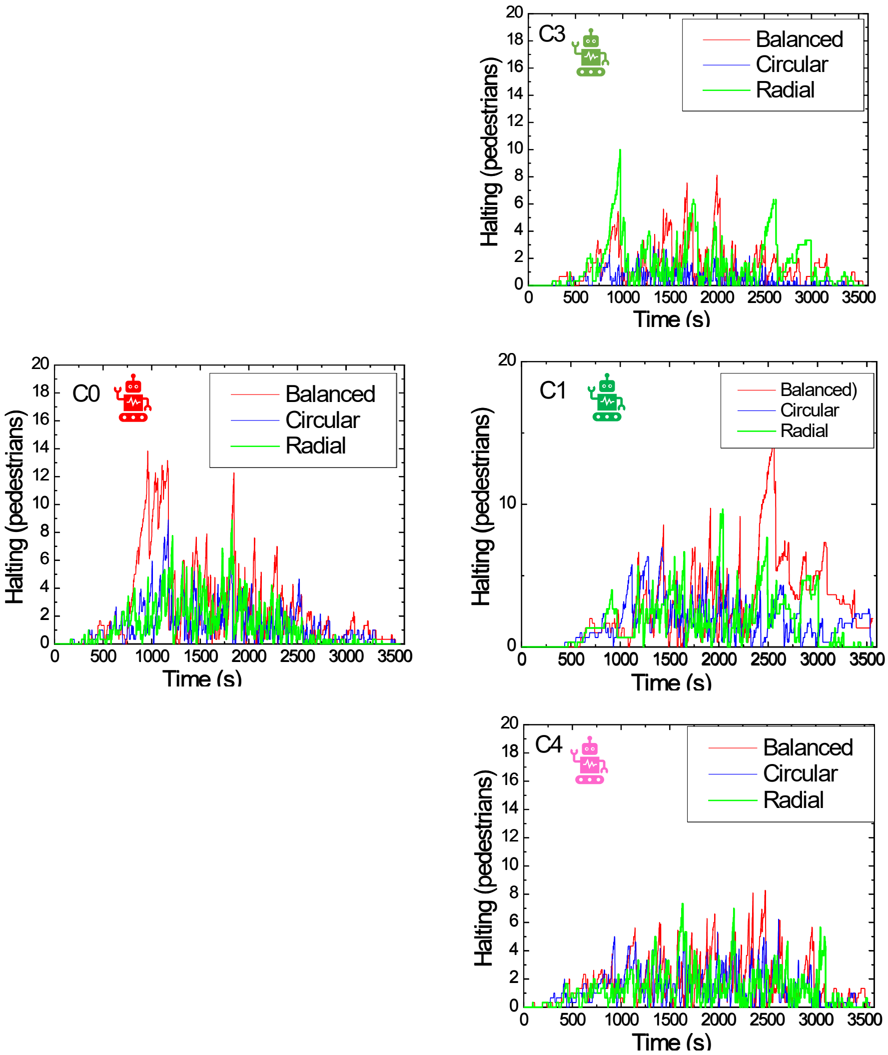


Figure 13. Comparative analysis of pedestrian halting metrics at intersections C0–C4 across different network strategies. Results are shown for Network 1, representing a balanced topology; Network 2 (Circular + Outbound Radial); and Network 4 (Radial + Outbound Radial).

In Network 1, which employs a balanced strategy, pedestrian waiting times are marginally higher than in the other networks. This outcome may be attributed to the absence of prioritization mechanisms, which results in increased competition for the green phase. Nonetheless, pedestrians generally experience short waiting periods before their phase is activated, underscoring the effectiveness of the underlying decision-making process.

Importantly, the number of waiting pedestrians remains consistently low across all strategies, which reflects the efficient incorporation of pedestrian phases into the overall control logic. These findings suggest that the management of pedestrian flows is handled effectively, with minimal disruption to both pedestrian and vehicular movements, thereby reinforcing the robustness of the proposed network control strategies. During the network evaluations, traffic metrics such as halted vehicles and pedestrian flow showed promising results. Each network learned a specific control strategy and, when tested, demonstrated the ability to effectively manage traffic, even during periods of high congestion. The models are capable of selecting the most suitable actions based on real-time traffic conditions at the intersection, thereby optimizing the flow of both vehicles and pedestrians. In Figure 14 it is presented a comparative analysis of vehicles (a) and pedestrian (b) halting metrics at intersections C1 across different inbound network strategies.

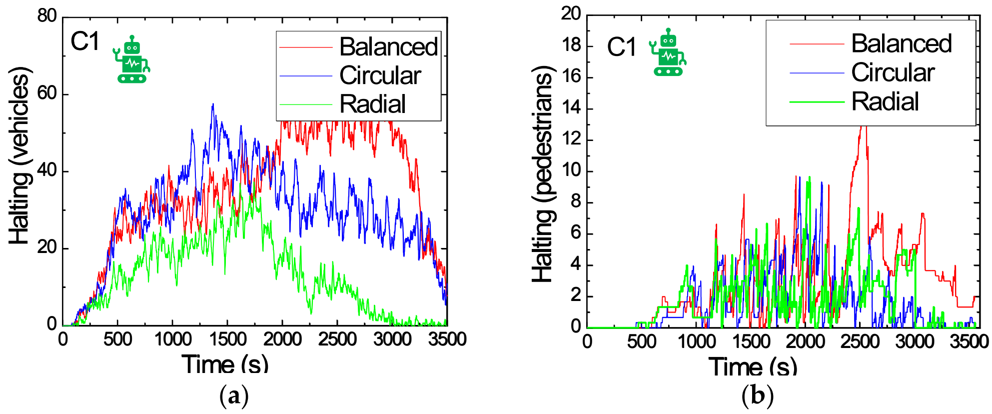


Figure 14. Comparative analysis of vehicles (**a**) and pedestrian (**b**) halting metrics at intersections C1 across different inbound network strategies. Results are shown for Network 1, representing a Balanced topology; Network 3 (Circular + Inbound Radial); and Network 5 (Radial + Inbound Radial).

For both vehicle and pedestrian metrics, it is observed that the strategy maintaining a balanced 50/50 allocation between the two main arteries results in the highest number of vehicles and pedestrians waiting. Since this strategy does not prioritize any specific direction, it activates signal phases in a uniformly distributed manner, which leads to increased vehicle accumulation—particularly at the critical intersection, C1. However, this does not imply that the strategy is ineffective; it serves a specific function within the overall structure of the traffic cell network.

In scenarios where a cell is located farther from the city center, it may be more appropriate to adopt strategy 3, which prioritizes circular traffic flow. As traffic approaches the urban core, control must tighten to prevent congestion. A gradual transition through strategy 1 and then strategy 5 allows for progressively increasing priority on radial flows, ensuring safer and smoother city entry.

5.4. MARL Testing Approach Vehicle and Pedestrian Phasing Diagrams

Figure 15 presents the percentage distribution of the most frequently activated phases over time for each network. The figure illustrates how different control strategies distribute green time among exit phases, revealing the adaptive behavior of the MARL–VLC framework in optimizing outbound traffic flow and reducing congestion at key intersections.

The analysis of phase activation across the examined networks reveals distinct strategies in managing traffic flows. Network 1, implementing a balanced strategy, exhibits a relatively uniform distribution across the main phases. Phases P1 (N–S), P5 (W–E), and P9 (pedestrian) are activated at comparable levels, reflecting the absence of prioritization and an approach aimed at serving all flows equally.

In contrast, Network 2, which follows a Circular + Outbound Radial strategy, places greater emphasis on W–E traffic (P5). The activation of phase P5 reaches approximately 35% at intersections C0, C1, and C2, while P1 is activated at lower levels (10–20%), indicating reduced priority for the radial artery. Pedestrian phase P9 varies between 18 and 35% depending on the intersection, with C1 consistently exhibiting the lowest activation (<20%) due to its critical traffic role and the need to maintain vehicle throughput.

Network 4, applying a Radial + Outbound strategy, prioritizes N–S traffic (P1), with increased activation (~25%) at radial intersections. Conversely, P5 activation decreases to around 25%, showing diminished priority for the circular artery. Pedestrian phase P9 demonstrates wider variability (18–43%) depending on the intersection. Despite P1 and P5 appearing equally activated (~25%), true balance is not achieved, as additional phases P2 and P3 are frequently triggered (P2: 13–20%, P3: 6–15%). These supplementary activations support continued radial flow, revealing a hidden prioritization not apparent when considering only the main phases.

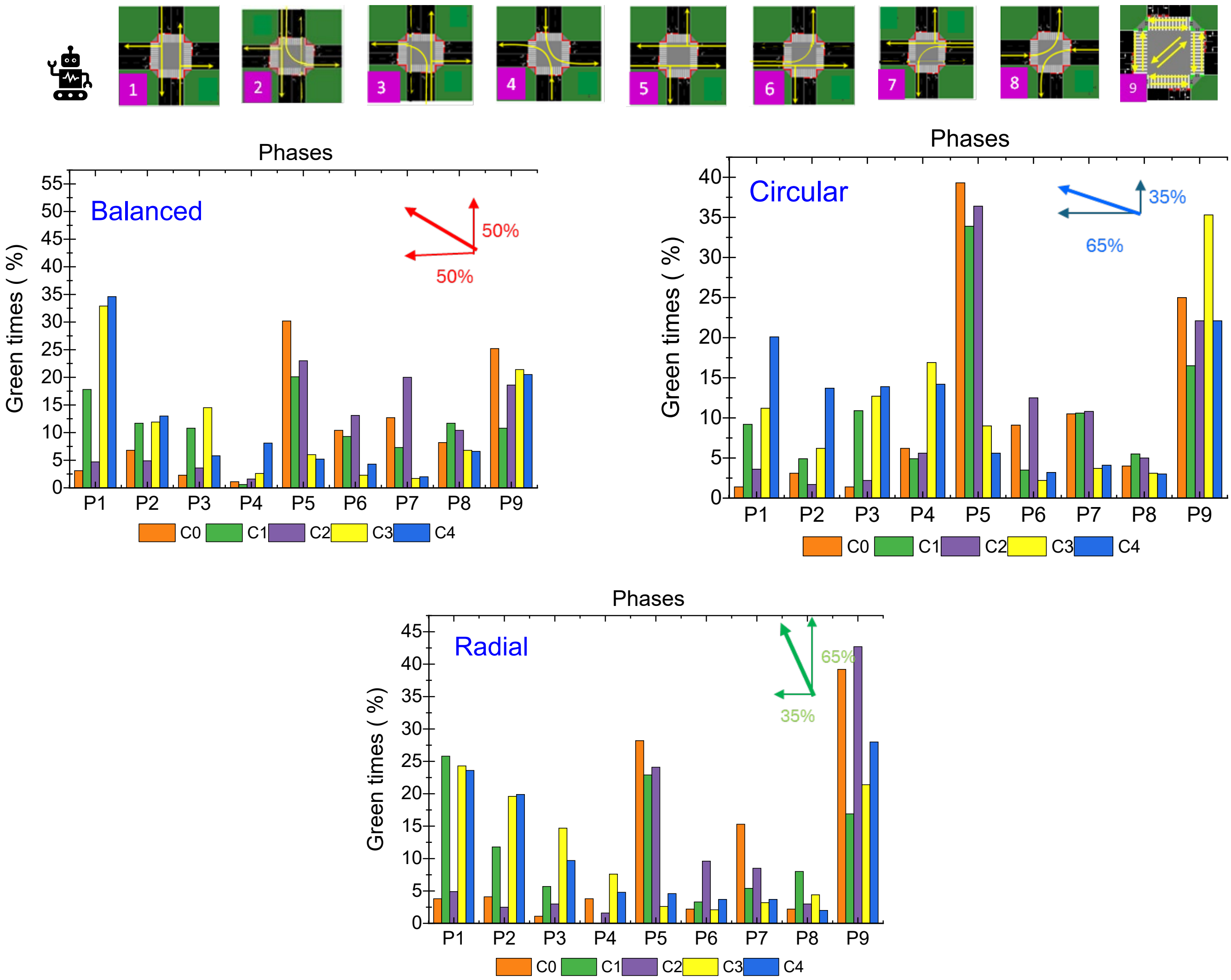


Figure 15. Analysis of the percentage of green time allocated to outbound traffic strategies at intersections C0–C4. Results are shown for Network 1, representing a balanced topology; Network 2 (Circular + Outbound Radial); and Network 4 (Radial + Outbound Radial).

Figure 16 presents the strategies marked by inbound traffic into the city. The figures highlights how different control strategies manage green time distribution for inbound movements, demonstrating the adaptability of the MARL–VLC framework in optimizing incoming traffic flow and reducing intersection delays. In Network 3, which prioritizes circular flow with entry via the radial artery, there is a significant increase in the activation of phase P5 (W–E), ranging from 30% to 40%. Phase P1 (N–S) also sees regular activation, between 18% and 39%, with intersection C3 showing the highest P1 activation—expected due to its role as a key traffic generator on the radial route.

In Network 5, which explicitly prioritizes the radial artery for inbound movement, phase P1 becomes dominant, with activation rates between 33% and 39%, surpassing those in Network 3. This reflects the system's adaptation to traffic needs, ensuring efficient vehicle entry and sustained N–S flow. This also helps reduce queuing at C1, where increased N–S phase activation alleviates local and downstream congestion. Meanwhile, phase P5 decreases to 20–30%, consistent with its reduced priority in this configuration.

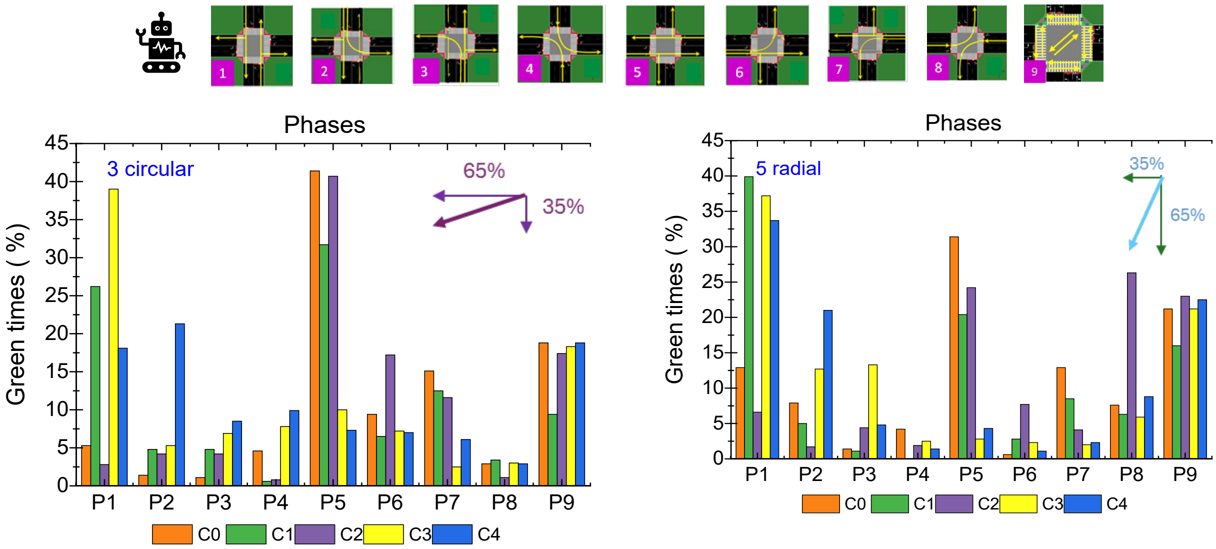


Figure 16. Analysis of the percentage of green time allocated to inbound traffic strategies at intersections C0–C4. Results are shown for Network 1, representing a balanced topology; Network 3 (Circular + Inbound Radial); and Network 5 (Radial + Inbound Radial).

5.5. MARL Training and Testing Approach: Influence of Neighboring Intersections

To evaluate the impact of the neighbor influence factor β on traffic performance, a preliminary analysis was conducted using a range of values: β = 0, 0.1, 0.2, 0.3, and 0.4 in Equation (1). For each value, key performance metrics, average queue lengths, vehicle and pedestrian halting, and average vehicle speeds were monitored across all tested networks. The results indicated that β = 0.3 consistently provided the best balance, achieving the largest reduction in queue lengths while maintaining high intersection throughput. Lower values of β led to insufficient coordination between neighboring intersections, whereas higher values caused overreaction and instability in phase adaptation. Based on these findings, β = 0.3 was selected for the main experiments, providing a transparent and quantitatively justified choice for the neighbor influence factor.

Considering Equation (1), Figure 17 shows the cumulative negative rewards obtained during the training phase for β = 0 and β = 0.3, allowing a direct comparison of learning efficiency. The figure illustrates the impact of the β parameter on learning stability and convergence speed within the MARL–VLC framework. In both cases, the reward curves exhibit an overall upward trend across the episodes, indicating consistent improvements in agent performance.

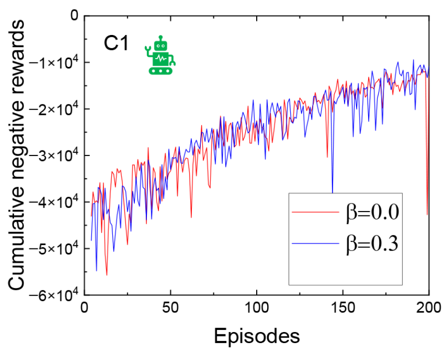


Figure 17. Evolution of negative cumulative rewards during MARL training, with the horizontal axis representing epochs, for Network 1 at intersection C1 under two configurations: β = 0 and β = 0.3.

Figure 18 compares the temporal evolution of vehicle and pedestrian halting, as well as vehicle speeds at intersections C1 and C2, under the two scenarios (β = 0 and β = 0.3).

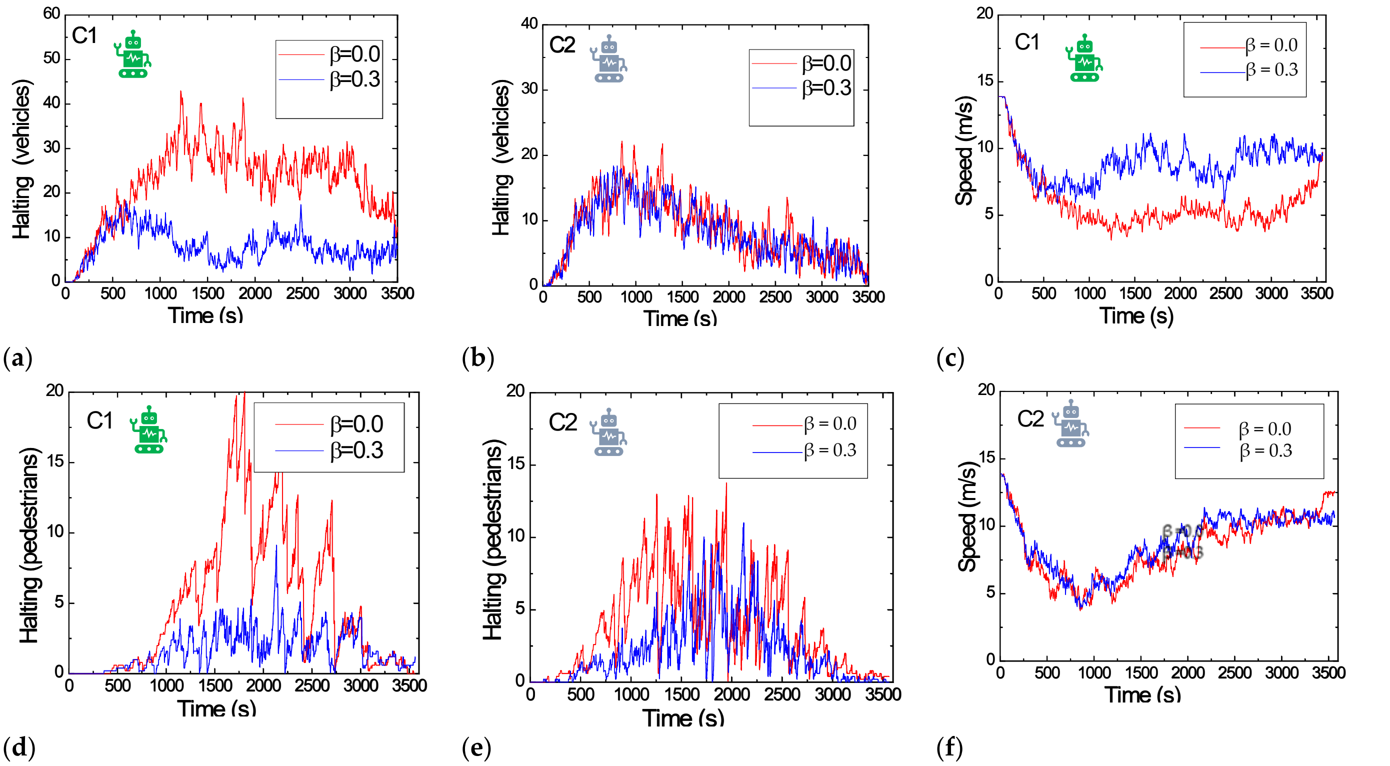


Figure 18. Comparison of vehicle (**a**,**b**) and pedestrian (**d**,**e**) halting trends, and average vehicle speeds (**c**,**f**) at C1 and C2 intersections for β = 0 and β = 0.3 for Network 1 at intersections C1 and C2 under two configurations: β = 0 and β = 0.3.

The results reveal that the β = 0 condition consistently generates a higher number of halted vehicles, reflecting greater congestion levels. The analysis demonstrates that incorporating neighboring intersection influence (β = 0.3) significantly enhances traffic flow management. For vehicles, congestion and halting times at the central intersection (C1) are markedly reduced, with only minor impacts on the surrounding intersections. Pedestrian waiting times also decrease across all intersections, with the most pronounced improvement again observed at C1.

Furthermore, the overall number of vehicles and pedestrians remaining in the environment during the testing phase is lower when β = 0.3 is applied, serving as strong evidence that accounting for neighbor influence improves traffic throughput and reduces delays.

5.6. MARL Testing Approach: System with Dynamic Phase Duration: SAPA

Another contribution of this research is the introduction of a dynamic phase duration mechanism, supported by VLC-derived data on vehicle queues and lane occupancy. Unlike conventional fixed timings, phase durations are adaptively adjusted according to real-time traffic demand and downstream lane conditions. This approach underpins the SAPA framework, designed to prevent link saturation and enhance intersection throughput.

In the case study involving five intersections, C0 and C1, connected by a 400 m link, illustrate the principle. For vehicle movements toward this segment (e.g., phases 5 and 6), the system assesses both the link occupancy (ρ) and the queue length (Q) at C0. When $\rho < 40\%$, the phase duration T is extended proportionally to Q, as expressed by:

$$T = \begin{cases} T_{base} + T_{base}.\, Q.\alpha_{circ}, & \rho_{C0-C1} < 40\% \\ T_{base} \quad , & \rho_{C0-C1} \geq 40\% \end{cases} \tag{4}$$

where $T_{base} = 8$ s and α_{circ} represents the weight assigned to the dominant circular artery. Minor phases receive only a reduced increment ($\alpha_{low} = 25\%$). Depending on local flow patterns, intersections prioritize circular (α_{circ}) or radial (α_{rad}) arteries—C0/C2 favoring circular movements, C3/C4 radial ones, and C1 balancing both according to strategy.

When traffic transitions from a 400 m to a 200 m link, the occupancy threshold decreases to 35%. If this limit is exceeded, the phase is constrained to T_{base}, allowing downstream clearance. The integration of neural-network-based phase selection with SAPA-driven adaptive timing enables efficient, scalable, and congestion-resilient traffic control across interconnected intersections.

During training, each neural network learns to interact with the environment under a specific traffic generation scenario that reflects its assigned control strategy. To validate their performance, the trained models, with and without SAPA module, are tested across different strategies, with the initial analysis focusing on vehicle flow through the intersections of the test environment. This comparison highlights how each network adapts to varying control logics.

A comparison of vehicle halting at each intersection for both networks under Network is exemplify in Figure 19. This strategy establishes a balanced configuration between radial and circular arteries, resulting in uniform activation of N–S and W–E phases. While this equilibrium leads to higher vehicle accumulation—particularly at the critical intersection C1—it serves an important role within the broader coordination of the traffic cell network.

At the critical intersection C1, both networks operate as expected, though the difference in performance is substantial. The SAPA-enabled network maintains an average queue of around 20 vehicles, with brief peaks up to 40, while the fixed-duration network frequently exceeds 40 vehicles and continues to accumulate over time. This improvement stems from SAPA's adaptive phase duration, which allows each activation to clear more vehicles compared to the fixed 8 s phases, reducing total clearance time and congestion. Neighboring intersections (C0, C2, C3, C4) remain stable, but vehicle counts are consistently lower in the SAPA-enabled network, indicating enhanced flow efficiency.

For Strategy 1, both arterial directions should remain balanced, meaning that the active phases in the N–S direction should correspond closely to those in the W–E direction. Figure 20 illustrates the active phases executed by the agents at each intersection over time for both neural networks under study.

The figure illustrates how the MARL–VLC framework dynamically allocates active phases, highlighting differences in phase utilization patterns between the two network topologies throughout the simulation. Overall, it can be observed that the dominant phases are P1 (N–S), P5 (W–E), and P9 (pedestrian phase). A notable difference between the two networks is that the weaker phases in the N–S direction (P2, P3, and P4) and W–E direction (P6, P7, and P8) are less frequently activated in the SAPA network. These weaker phases are largely replaced by pedestrian phase activations, which show an increase of over 20% in the SAPA network compared to the baseline. Nonetheless, both networks maintain an approximately balanced activation between P1 and P5 at intersection C1.

Results has also shown that for outer traffic cells, Strategy 3 (favoring circular flow) is more suitable, whereas proximity to the urban core demands stricter control to prevent congestion. A progressive transition from Strategy 1 to Strategy 5 increases radial prioritization, supporting smoother inbound movement. Conversely, for outbound flows, Strategy 4 emphasizes radial movement (e.g., S→N), followed by Strategy 1 for balance and Strategy 2 farther from the center to reestablish circular dominance. This hierarchical sequencing enables efficient redistribution of traffic and mitigates congestion. After evaluating all five strategies, it can be concluded that the network integrating the SAPA module demonstrated the most effective overall performance. By dynamically extending phase durations, intersection throughput increased, improving mobility and reducing queue lengths. Across all strategies, the critical intersection C1 consistently maintained smooth flow and avoided vehicle congestion, thereby preserving the overall efficiency of the traffic network. These improvements have direct sustainability implications: reduced congestion lowers fuel consumption and CO_2 emissions, contributing to energy efficiency and environmental protection. Additionally, smoother traffic flow enhances safety and reliability for all road users, supporting social equity and more sustainable urban mobility.

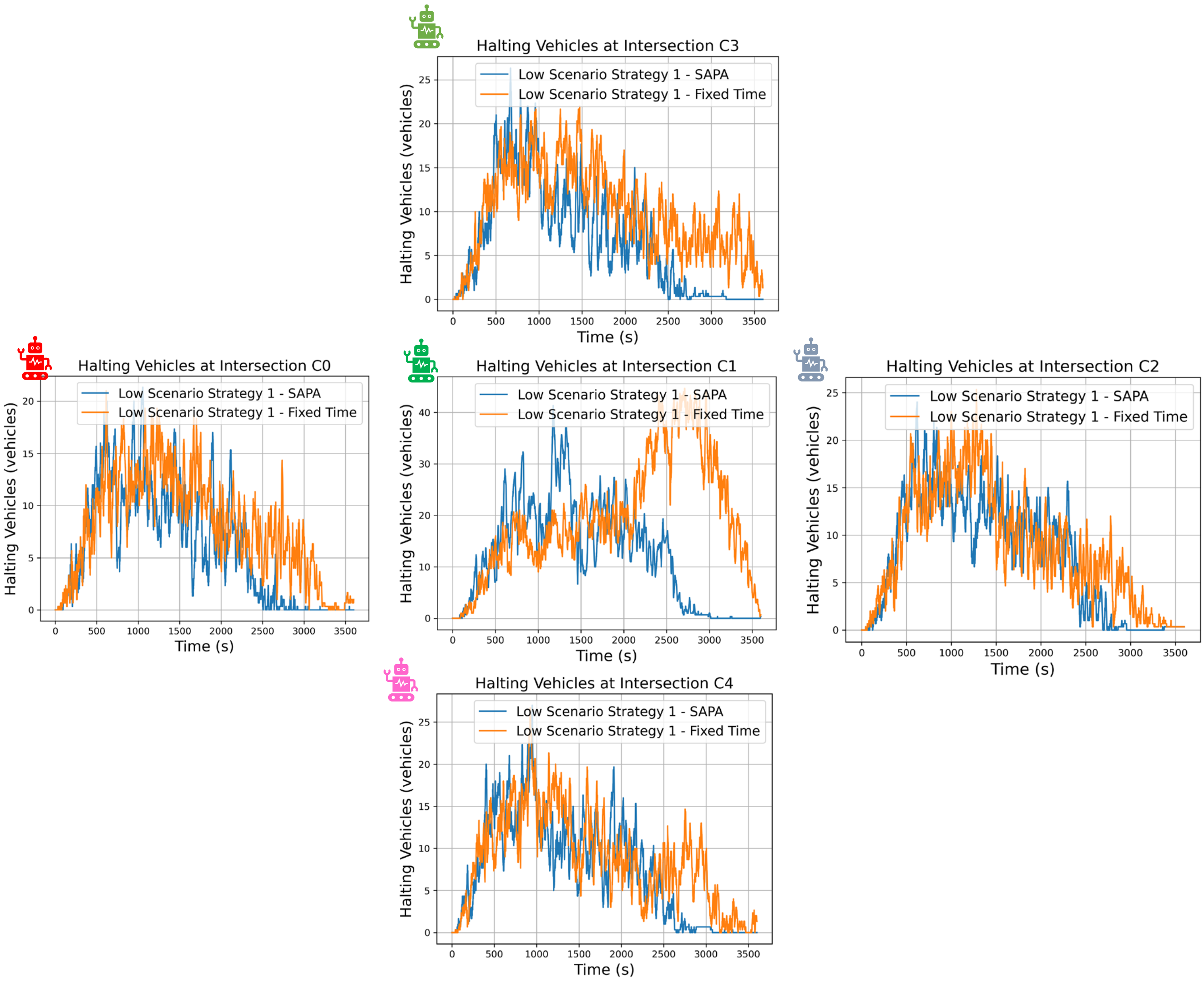


Figure 19. Comparison of vehicle halting over time at each intersection for Network 1, with and without the SAPA module (adaptive MARL–VLC control vs. fixed-time control). The figure illustrates how the SAPA-enabled strategy reduces vehicle stops and improves traffic flow across all intersections compared to the fixed-time baseline.

5.7. Robustness, Sustainability and Scalability of the MARL–VLC Framework

The proposed MARL–VLC framework is designed to remain robust even when DRL agents have only partial visibility of the traffic network. By leveraging Visible Light Communication (VLC), intersections exchange high-fidelity, real-time information, enabling agents to access critical state data from neighboring intersections. This integration allows each agent to make informed, decentralized decisions despite limited local observations.

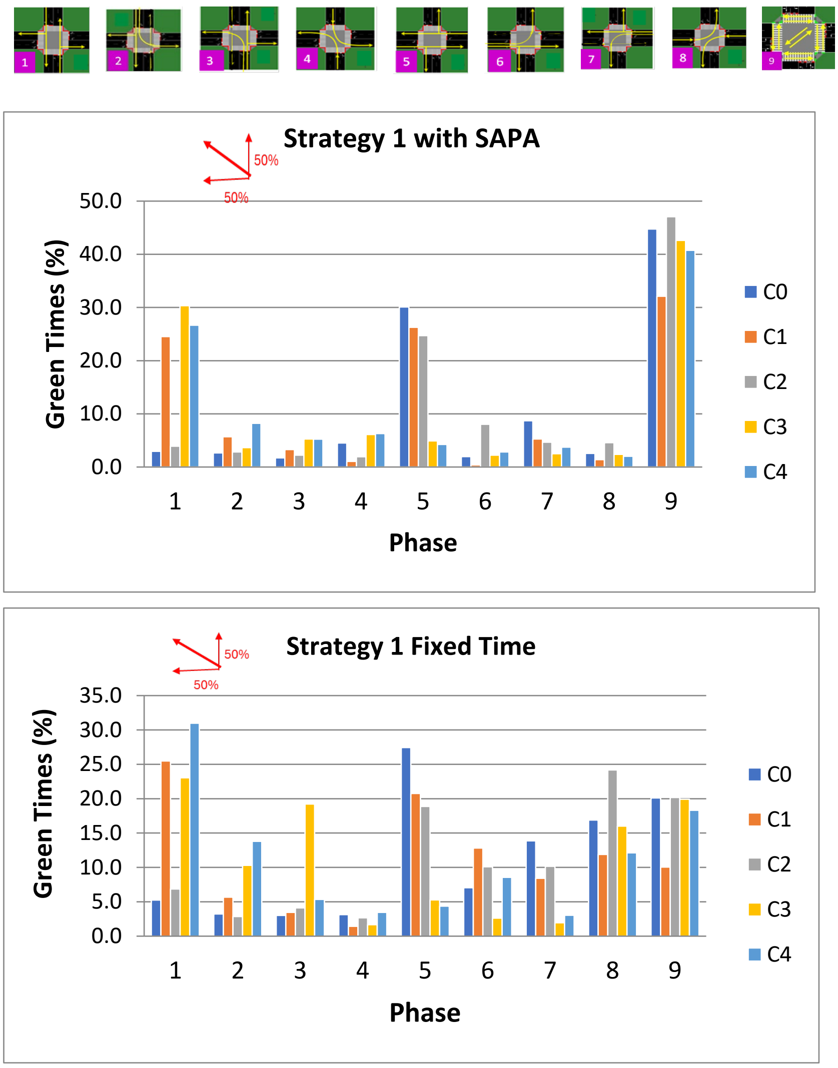


Figure 20. Percentage of active phases over the simulation time at each intersection for both networks under Strategy 1.

The SAPA module further enhances robustness by considering link lengths and congestion levels, dynamically adjusting signal phases and green times to minimize spillback and maintain network fluidity. Simulation results demonstrate that the system effectively manages queues, mitigates congestion, and preserves high intersection throughput under partial, intermittent, or heterogeneous traffic conditions.

The framework's scalability was evaluated under different ramp configurations and varying traffic demands. Circular and radial ramp layouts consistently exhibited reductions in queue lengths and improvements in intersection throughput, confirming that the adaptive phase selection and decentralized decision-making mechanisms can handle diverse and complex network structures.

Importantly, the integration of MARL with VLC and the adaptive capabilities of the SAPA module contribute to urban sustainability. By reducing vehicle stops, queues, and congestion, the framework lowers fuel consumption and CO_2 emissions, improves energy efficiency, and enhances pedestrian safety. These outcomes demonstrate that the proposed system not only optimizes traffic performance but also supports environmentally friendly and socially equitable urban mobility.

6. Conclusions and Future Work

This study proposed a decentralized multi-agent reinforcement learning (MARL) framework for adaptive traffic signal control, integrating deep reinforcement learning (DRL) with visible light communication (VLC) to enable real-time, local decision-making at intersections. Simulation results demonstrated that the SAPA-enabled network with adaptive phase durations significantly outperformed fixed-time control, reducing vehicle queues, improving throughput, and maintaining smoother traffic flow, particularly at critical intersections. The integration of VLC proved crucial for high-fidelity inter-agent communication, supporting dynamic phase adjustments, conflict resolution, and robust performance under partial information. These improvements also promote urban sustainability by reducing energy consumption and CO_2 emissions while enhancing pedestrian safety.

While the framework was primarily evaluated on standard intersections, it is conceptually applicable to more complex layouts, including roundabouts, where the SAPA module effectively minimizes spillback and maintains traffic flow. Some limitations remain, such as the need for scenario-specific tuning in highly non-standard intersections or networks with highly dynamic demand. Moreover, the current evaluation relies on simulation environments, which may not capture all real-world complexities, and assumes ideal communication conditions without fully considering uncertainties in sensors or environmental factors.

Future research will focus on deploying and testing the framework in real urban intersections, extending it to larger, heterogeneous networks with multimodal traffic, and exploring more complex traffic scenarios. Additional directions include investigating advanced reinforcement learning techniques, enhancing robustness and scalability, and integrating further VLC infrastructure and intelligent transportation technologies to enable fully autonomous, self-optimizing, and adaptive traffic systems. Collectively, these efforts aim to consolidate the MARL–VLC framework as a scalable, adaptive, and sustainable solution for modern urban traffic management.

Author Contributions: Conceptualization, M.A.V.; Formal analysis, M.V. (Manuela Vieira), G.G. and M.V. (Mario Vestias); Investigation, G.G., M.A.V. and M.V. (Mario Vestias); Methodology, M.V. (Manuela Vieira), M.A.V. and M.V. (Mario Vestias); Software, G.G. and M.V. (Mario Vestias); Validation, G.G., M.A.V., P.L. and P.V.; Writing—original draft, M.V. (Manuela Vieira); Writing—review and editing, M.V. (Manuela Vieira). All authors have read and agreed to the published version of the manuscript.

Funding: This research received support from FCT—Fundação para a Ciência e a Tecnologia, through the Research Unit CTS—Center of Technology and Systems, with references UIDB/00066 and IPL/IDI&CA2024/INUTRAM_ISEL.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors acknowledge CTS-ISEL and IPL.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Brys, T.; Pham, T.T.; Taylor, M.E. Distributed learning and multi-objectivity in traffic light control. *Connect. Sci.* **2014**, *26*, 65–83. [CrossRef]
2. Zhang, Y.; Zhong, W.; Liu, T. Reinforcement learning for dynamic traffic management: A scalable approach to congestion reduction. In Proceedings of the 2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA), New York City, NY, USA, 7–9 August 2025; pp. 1–6. [CrossRef]
3. Richter, S.; Aberdeen, D.; Yu, J. Natural actor-critic for road traffic optimisation. In Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 4–9 December, 2006; pp. 1169–1176.
4. Cunningham, R.; Garg, A.; Cahill, V. A collaborative reinforcement learning approach to urban traffic control optimization. In Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Sydney, Australia, 9–12 December 2008; pp. 556–560.
5. Aziz, H.M.A.; Feng, Z.; Ukkusuri, S. Reinforcement learning-based signal control using R-Markov average reward technique (RMART) accounting for neighborhood congestion information sharing. In Proceedings of the Transportation Research Board 92nd Annual Meeting, Washington, DC, USA, 13–17 January 2013.
6. Chu, T.; Wang, J.; Codecà, L.; Li, Z. Multi-agent deep reinforcement learning for large-scale traffic signal control. *arXiv* **2019**, arXiv:1903.04527. [CrossRef]
7. Shi, Y.; Wang, X.; Zhang, L.; Li, Q. A survey on traffic signal control problems with MARL. *ACM Comput. Surv.* **2023**, *56*, 1–36. [CrossRef]
8. Vieira, M.A.; Silva, R.; Santos, P. Visible light communication and learning-based control for urban intersections. *Symmetry* **2024**, *16*, 240. [CrossRef]
9. Sikder, P.; Rahman, M.T.; Bakibillah, A.S.M. Advancements and challenges of visible light communication in intelligent transportation systems: A comprehensive review. *Photonics* **2025**, *12*, 225. [CrossRef]
10. Nawaz, T.; Seminara, M.; Caputo, S.; Mucchi, L.; Cataliotti, F.; Catani, J. IEEE 802.15.7-compliant ultra-low latency relaying VLC system for safety-critical ITS. *arXiv* **2019**, arXiv:1906.08773. [CrossRef]
11. Akopov, A.S.; Beklaryan, L.A. Traffic improvement in Manhattan road networks with the use of parallel hybrid biobjective genetic algorithm. *IEEE Access* **2024**, *12*, 19532–19552. [CrossRef]
12. Yang, S.K.; Li, J.C.; Shi, H.B. Mix-attention approximation for homogeneous large-scale multi-agent reinforcement learning. *Neural Comput. Appl.* **2023**, *35*, 3143–3154. [CrossRef]
13. Zhu, C.; Dastani, M.; Wang, S. A survey of multi-agent deep reinforcement learning with communication. *Auton. Agents Multi-Agent Syst.* **2024**, *38*, 4. [CrossRef]
14. Bokade, R.; Jin, X.; Amato, C. Multi-agent reinforcement learning based on representational communication for large-scale traffic signal control. *IEEE Access* **2023**, *11*, 47646–47658. [CrossRef]
15. He, Y.; Wang, Y.H.; Yu, F.R.; Lin, Q.Z.; Li, J.Q.; Leung, V.C.M. Efficient resource allocation for multi-beam satellite-terrestrial vehicular networks: A multi-agent actor critic method with attention mechanism. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 2727–2738. [CrossRef]
16. Thorpe, T.L.; Anderson, C.W. *Traffic Light Control Using SARSA with Three State Representations*; IBM Corporation: New York City, NY, USA, 1996.
17. Wiering, M. Multi-agent reinforcement learning for traffic light control. In Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000), Stanford, CA, USA, 29 June–2 July 2000.
18. Jin, J.; Ma, X. Adaptive group-based signal control using reinforcement learning with eligibility traces. In Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spain, 1–15 September 2015; pp. 2412–2417. [CrossRef]
19. Aleko, D.R.; Djahel, S. An efficient adaptive traffic light control system for urban road traffic congestion reduction in smart cities. *Information* **2020**, *11*, 119. [CrossRef]
20. Zhu, F.; Aziz, H.M.A.; Qian, X.; Ukkusuri, S.V. A junction-tree based learning algorithm to optimize network wide traffic control: A coordinated multi-agent framework. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 487–501. [CrossRef]
21. Medina, J.C.; Benekohal, R. Corridor-based coordination of learning agents for traffic signal control by enhancing Max-Plus algorithm. In Proceedings of the Transportation Research Board 93rd Annual Meeting, Washington, DC, USA, 12–16 January 2014.
22. Ma, T.H.; Peng, K.X.; Rong, H.; Qian, Y.R.; Al-Nabhan, N. Hierarchical coordination multi-agent reinforcement learning with spatio-temporal abstraction. *IEEE Trans. Emerg. Top. Comput. Intell.* **2024**, *8*, 533–547. [CrossRef]
23. Xu, H.L.; Xiao, W.; Cassandras, C.G.; Zhang, Y.; Li, L. A general framework for decentralized safe optimal control of connected and automated vehicles in multi-lane signal-free intersections. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 17382–17396. [CrossRef]
24. Vieira, M.A.; Vieira, M.; Louro, P.; Vieira, P.; Fantoni, A. Vehicular visible light communication for intersection management. *Signals* **2023**, *4*, 457–477. [CrossRef]

25. Yousefpour, A.; Fung, C.; Nguyen, T.; Kadiyala, K.; Jalali, F.; Niakanlahiji, A.; Kong, J.; Jue, P.J. All one needs to know about fog computing and related edge computing paradigms: A complete survey. *J. Syst. Archit.* **2019**, *98*, 289–330. [CrossRef]
26. Bilal, J.M.; Jacob, D. Intelligent traffic control system. In Proceedings of the 2007 IEEE International Conference on Signal Processing and Communications, Dubai, United Arab Emirates, 24–27 November 2007; pp. 496–499. [CrossRef]
27. Yousefi, S.; Altman, E.; El-Azouzi, R.; Fathy, M. Analytical model for connectivity in vehicular ad hoc networks. *IEEE Trans. Veh. Technol.* **2008**, *57*, 3341–3356. [CrossRef]
28. Shen, W.-H.; Tsai, H.-M. Testing vehicle-to-vehicle visible light communications in real-world driving scenarios. In Proceedings of the 2017 IEEE Vehicular Networking Conference (VNC), Torino, Italy, 27–29 November 2017; pp. 187–194.
29. 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]
30. Alvarez Lopez, J.; Behrisch, M.; Bieker-Walz, L.; Erdmann, J.; Flötteröd, Y.P.; Hilbrich, R. Microscopic traffic simulation using SUMO. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 2575–2582.
31. Caputo, S.; Mucchi, L.; Cataliotti, F.; Seminara, M.; Nawaz, T.; Catani, J. Measurement-based VLC channel characterization for I2V communications in a real urban scenario. *Veh. Commun.* **2021**, *28*, 100305. [CrossRef]
32. Miranda, R.F.; Barriquello, C.H.; Reguera, V.A.; Denardin, G.W.; Thomas, D.H.; Loose, F.; Amaral, L.S. A review of cognitive hybrid radio frequency/visible light communication systems for wireless sensor networks. *Sensors* **2023**, *23*, 7815. [CrossRef] [PubMed]
33. Vieira, M.A.; Vieira, M.; Vieira, P.; Fernandes, R.; Louro, P. Dynamic vehicular visible light communication for traffic management. In *Next-Generation Optical Communication: Components, Sub-Systems, and Systems XII.*; Li, G., Nakajima, K., Srivastava, A.K., Eds.; SPIE: Bellingham, WA, USA, 2023. [CrossRef]
34. Vieira, M.; Vieira, M.A.; Galvão, G.; Louro, P.; Véstias, M.; Vieira, P. Enhancing urban intersection efficiency: Utilizing visible light communication and learning-driven control for improved traffic signal performance. *Vehicles* **2024**, *6*, 666–692. [CrossRef]
35. Vieira, M.A.; Vieira, M.; Louro, P.; Vieira, P. Cooperative vehicular communication systems based on visible light communication. *Opt. Eng.* **2018**, *57*, 076101. [CrossRef]
36. Papageorgiou, M. Overview of road traffic control strategies. *IFAC Proc. Vol.* **2004**, *37*, 29–40. [CrossRef]
37. Vieira, P.; Vieira, M.A.; Queluz, M.P.; Rodrigues, A. A novel vehicular mobility model for wireless networks. *Wirel. Pers. Commun.* **2007**, *43*, 1689–1703. [CrossRef]
38. Sousa, I.; Queluz, P.; Rodrigues, A.; Vieira, P. Realistic mobility modeling of pedestrian traffic in wireless networks. In Proceedings of the 2011 IEEE EUROCON—International Conference on Computer as a Tool, Lisbon, Portugal, 27–29 April 2011; pp. 1–4. [CrossRef]