Sensors & Transducers, Vol. 270, Issue 3, November 2025, pp. 69-78

www.sensorsportal.com

Sensors & Transducers

Published by IFSA Publishing, S. L., 2025
http://www.sensorsportal.com

Intelligent Intersection Management through Multi-agent Reinforcement Learning, Self-adaptive Phase Adjustment and Visible Light Communication

[1, 2] Manuel A. VIEIRA, [1] Tomás ANTUNES, [1] Gonçalo GALVÃO, [1, 2, 3] Manuela VIEIRA, [1, 5] Mário VÉSTIAS and [1, 2] Paula LOURO

[1] Electronics Telecommunication and Computer Dept. ISEL, R. Conselheiro Emídio Navarro, 1959-007 Lisboa, Portugal
[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] Instituto de Telecomunicações, Instituto Superior Técnico, 1049-001, Lisboa, Portugal
[5] INESC INOV-Lab*,* Instituto Superior Técnico, Universidade de Lisboa, 1000-029, Lisboa, Portugal
E-mail: manuela.vieira@isel.pt

Received: 31 May 2025 / Revised: 7 Nov. 2025 / Accepted: 12 Nov. 2025 / Published: 28 Nov. 2025

Abstract: Urban traffic management faces growing challenges due to increasing volumes of vehicles and pedestrians, resulting in congestion, delays, and safety concerns. This study introduces an innovative traffic signal control framework that integrates Multi-Agent Reinforcement Learning (MARL) with Visible Light Communication (VLC) and a Self-Adaptive Phase Adjustment (SAPA) module to enhance coordination across urban intersections. In the proposed architecture, each intersection is governed by an independent Deep Reinforcement Learning (DRL) agent capable of making real-time, decentralized decisions for both vehicular and pedestrian flows. Cooperative behavior emerges through the MARL framework, allowing agents to account for the dynamic states of neighboring intersections and improve network-wide efficiency. VLC provides high-resolution, low-latency data exchange between vehicles, pedestrians, and infrastructure, enabling precise sensing of position, speed, queue length, and stop duration. The introduction of the SAPA module further enhances adaptability by dynamically adjusting phase durations based on real-time queue/request/response patterns, resolving conflicts and prioritizing urgent demands. Extensive simulations and field tests demonstrate that the MARL–VLC–SAPA system significantly outperforms centralized and conventional agent-based approaches, reducing waiting and travel times while improving overall safety and responsiveness in complex urban environments.

Keywords: Multi-Agent Reinforcement Learning (MARL), Visible Light Communication (VLC), Intelligent traffic signal control, Real-time decision making, Road safety.

1. Introduction

Urban traffic management poses a persistent challenge for modern cities, particularly during peak hours when high volumes of vehicles and pedestrians converge, resulting in severe congestion, delays, and increased safety risks. Given the limited feasibility of expanding physical infrastructure, intelligent and adaptive traffic signal control at intersections – critical nodes in urban networks – has become essential to ensuring smoother traffic flow and improved safety outcomes.

A promising direction in this domain lies in Multi-Agent Reinforcement Learning (MARL), where multiple agents, each controlling a separate intersection, learn to optimize local traffic signals while coordinating with neighboring intersections. This decentralized learning framework enables the system to dynamically adapt to heterogeneous and evolving traffic patterns across the city [1-3].

To further enhance responsiveness and situational awareness, this study integrates MARL with Visible Light Communication (VLC), a cutting-edge technology that leverages the intensity modulation of existing Light Emitting Diodes (LEDs) in streetlights, traffic signals, and vehicle headlamps. VLC provides real-time, high-bandwidth, and low-latency communication between infrastructure and road users (vehicles and pedestrians), enabling precise, continuous monitoring of traffic conditions such as queue lengths, vehicle speeds, and pedestrian presence [4, 5].

By combining the distributed decision-making capabilities of MARL with the real-time data exchange enabled by VLC, supported by Connected Vehicle (CV) technologies, this approach allows for coordinated and adaptive control of signalized intersections. The central research question addressed in this study is: how can the integrated use of MARL and VLC technologies improve the responsiveness, coordination, and overall efficiency of urban traffic signal control systems, particularly during high-demand periods such as rush hours?

This work investigates the synergistic potential of these technologies to deliver a scalable, intelligent traffic management solution tailored to the needs of modern urban mobility.

It is organized as follows. Section 1, introduces the motivation and context of the study. Section 2, reviews the state of the art, focusing first on reinforcement learning approaches for traffic signal control, particularly MARL-based methods, and then on vehicular communication technologies. Section 3, presents the proposed traffic signal control framework for multi-intersection networks, detailing the role of connected vehicles and Visible Light Communication, the system architecture and communication model, arterial traffic signal control strategies, and the cooperative multi-agent reinforcement learning approach, concluding with the integration of dynamic phase duration through the SAPA mechanism. Section 4, discusses the results, including the performance of VLC-enabled communication within the MARL framework, the training and testing of the network, and the evaluation of the dynamic phase duration system. Finally, Section 5, summarizes the main conclusions and outlines potential directions for future work.

2. Related Work

This study lies at the intersection of Multi-Agent Reinforcement Learning (MARL) and Visible Light Communication (VLC) for intelligent traffic management. By integrating these domains, it aims to address persistent challenges in scalability and pedestrian–vehicle coexistence [6]. The following subsections summarize related work in adaptive traffic control, MARL-based approaches, and vehicular communication technologies, outlining the research gaps that motivate this work.

2.1. Reinforcement Learning Approaches: MARL-based Traffic Signal Control (TSC)

Adaptive Traffic Signal Control (ATSC) dynamically adjusts signal timings based on real-time traffic demand to improve urban mobility. Conventional strategies, such as fixed-time plans, actuated systems, and model-predictive control, have shown effectiveness under specific conditions but are limited by their reliance on predefined rules and centralized coordination. These systems often struggle with scalability and adaptability in heterogeneous or rapidly changing traffic environments.

Reinforcement Learning (RL) has emerged as a key approach for optimizing signal control through data-driven policy learning. While single-agent RL improves local intersection performance, centralized models face scalability and communication bottlenecks when applied to large networks. Multi-Agent Reinforcement Learning (MARL) addresses these issues by distributing control among local agents that coordinate through shared states and actions, extending RL to stochastic game environments across arterial or grid networks.

Recent MARL-based works demonstrate significant reductions in delays and queue lengths [7] [8], but most focus solely on vehicular traffic, neglecting pedestrian dynamics and robustness under partial observability or noisy data. Independent MARL approaches are easy to implement yet lack coordination; partially cooperative models improve local performance but struggle to maintain network-wide equilibrium. Fully joint-action MARL achieves higher coordination but is computationally intensive and difficult to scale [9].

Emerging hybrid and hierarchical MARL frameworks propose a central policy learner supervising decentralized agents, balancing global coordination with local adaptability. These models incorporate multi-objective reward functions, including efficiency, safety, and fairness and represent a promising step toward scalable, resilient, and equitable urban traffic management.

Independent MARL is simple to implement but lacks coordination, reducing network-level effectiveness. Partially cooperative MARL introduces local cooperation but struggles to achieve system-wide equilibrium, while joint-action MARL provides strong coordination at the cost of scalability and computational efficiency. Most MARL studies prioritize efficiency metrics such as travel time and waiting time, often overlooking critical aspects like safety and multi-objective fairness. To address these issues, hybrid and hierarchical MARL frameworks have been proposed, where a central agent learns global policies and local agents execute optimized actions based on multi-objective reward functions. These designs offer a promising balance, enhancing scalability, coordination, and robustness in complex urban networks.

2.2. 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 [10, 11]. 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 [12] connected and autonomous vehicles (CAVs) [13], 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 [14]. Building on this potential, this article 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 [15]. To evaluate the effectiveness of the proposed V-VLC system, agent-based simulations are conducted using the Simulation of Urban MObility (SUMO) [16]. 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.

3. Traffic Signal Control at Multi-intersection Networks

Arterial traffic signal control refers to managing intersections formed by crossing two or more main roads, either radial or circular in design. The layout and spacing between intersections vary depending on traffic volume, road capacity, and network design. Each approach at an intersection comprises multiple lanes to accommodate different vehicle movements such as left-turns, right-turns, and through-traffic. These intersections are governed by standard traffic rules, with priority movements determined by the traffic signals in place.

3.1. Connected Vehicles and Visible Light Communication

CV technology is revolutionizing urban traffic by enabling real-time communication between vehicles (V2V) and infrastructure (V2I), enhancing traffic flow, safety, and reducing accidents. It is a fundamental component of smart systems that optimize road use and reduce congestion.

VLC, using modulated LED lights in streetlights and signals, allows data transmission alongside illumination. This dual-function tech integrates into urban infrastructure, enabling precise, real-time communication to optimize signals and reduce delays. The V-VLC system combines mesh and cellular controllers for efficient message relaying and edge computing [11, 17].

The system includes a transmitter that emits modulated light and a PIN-PIN receiver that detects light variations, both wirelessly connected. Fig. 1 shows the relative positions of the emitter and receivers, the coverage map with footprint regions in the unit cell (#1–#9).

The LED light is modulated using On-Off Keying (OOK). Each square unit cell contains tetra-chromatic white LEDs (WLEDs) at its corners, made from red (R: 626 nm), green (G: 530 nm), blue (B: 470 nm), and violet (V: 390 nm) chips, mixed in the right proportions to produce white light.

Each RGBV signal has a calibrated amplitude tied to its wavelength. With four independent emitters per VLC unit, the received optical signal may include one (#3, #5, #7, #9), two, three (#2, #4, #6, #8), or four (#1) light sources, resulting in 24 optical combinations and 16 possible photocurrent levels at the detector.

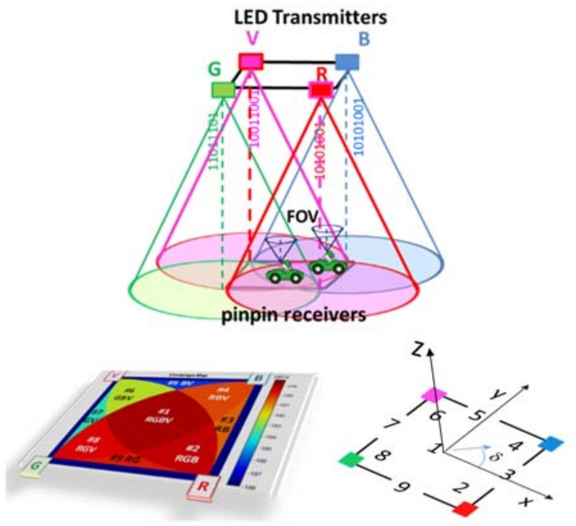

Fig. 1. V-VLC Emitter and receivers' relative positions and illustration of the: coverage map.

A PIN-PIN demultiplexer filters and decodes the combined OOK signals. Using the known amplitude values, it accurately retrieves the original message.

3.2. System Architecture and Communication Model

The system features modulated LED transmitters and multilayers photo receivers detecting light and wavelength variations. Light is modulated via ON-OFF Keying (OOK) [17].

The architecture (Fig. 2) supports Infrastructure-to-Cloud (I2IM) and Vehicle-to-Vehicle (V2V) communication via embedded computing. Streetlights, spaced 20 m apart, act as geo-transmitters, broadcasting L2V messages with IDs, synchronization, and traffic data. Vehicles/Pedestrians receive a unique ID.

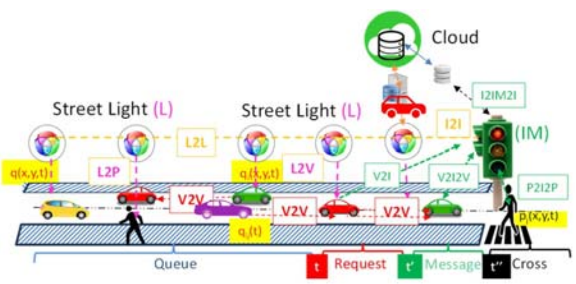

Fig. 2. 2D representation of the V-VLC architecture.

Intersection access follows a request/response protocol: a vehicle/pedestrian (V/P2I) sends a crossing request, and the infrastructure (I2V/P) replies with a trajectory assignment within footprint regions. If collision risk is detected, responses are delayed until safe.

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 [18]. The number of lanes at each intersection varies according to traffic volume and road capacity requirements. Each intersection includes multiple lanes for left-turn, right-turn, and through movements, as illustrated in Fig. 3. Standard traffic rules apply, with vehicles yielding to oncoming traffic and following traffic signals or signage.

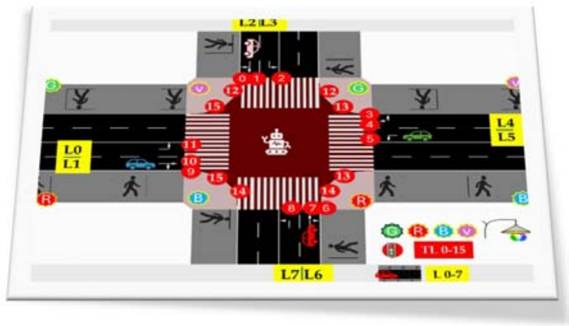

Fig. 3. Schematic diagram of a signal controlled standard intersection with coded lanes (L/0-7) and traffic lights (TL/0-15).

The five-intersection configuration, illustrated in Fig. 4, defines a single management cell within the system.

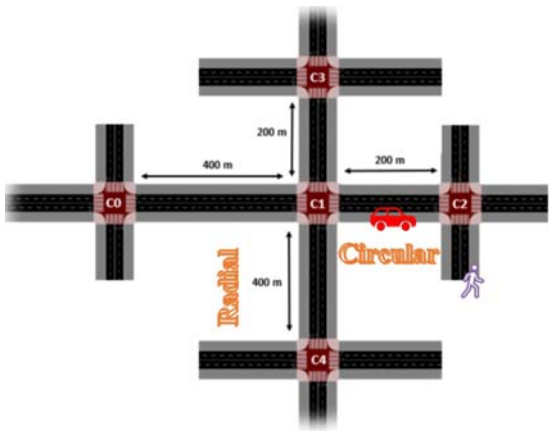

Fig. 4. Traffic scenario consisting of 5 homogeneous intersections with 4 arms each.

The urban network consists of two intersecting arterial roads–a horizontal road (C0–C1–C2) and a vertical road (C3–C1–C4)–meeting at the central junction C1, with each arm having two lanes optimized for CAVs. Junction C1 acts as the hub, receiving traffic solely from the four adjacent intersections. Phase-activation decisions by neighboring agents (C0, C2, C3, C4) control the incoming flows, making C1 critical for balancing traffic and preventing congestion.

Traffic signals at these intersections are designed with multiple phases (Fig. 5) to manage conflicting movements efficiently. Phase timing considers traffic demand, intersection geometry, and operational goals, ensuring smooth flow and safety across all vehicular movements.

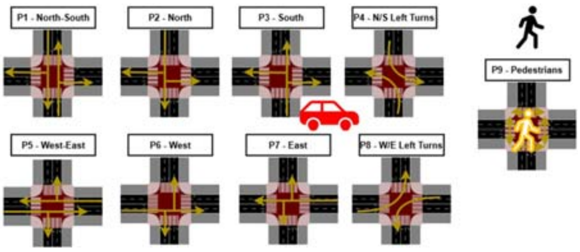

Fig. 5. Phasing diagram. Eight phases for vehicles and one exclusive phase for only pedestrians.

3.4. Cooperative Multi-agent Reinforcement Learning

To effectively manage the traffic data collected via Visible Light Communication (VLC), a real-time, intelligent control system is required to respond dynamically to vehicle and pedestrian movement requests. For this purpose, a Multi-Agent Reinforcement Learning (MARL) approach is adopted [1].

The core of reinforcement learning lies in its continuous interaction and trial-and-error process with the environment, constantly exploring and seeking optimal strategies.

The study models a simplified urban network consisting of five interconnected four-arm intersections, each with two lanes in both directions.

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 first layer of the network is the input layer, formed by an input layer of 164 neurons, representing the state of the environment. Following this, there are 5 hidden layers each one with 400 neurons, each with rectified linear unit (ReLU), an activation function commonly used in deep neural networks, with the ability to introduce non-linearity to the network, allowing the NN to learn complex patterns and representations in the data. Finally comes the output layer, with 9 neurons, which each one will display the Q-Values for each action. The next action that the agent will choose is determined by the maximum Q-Value output.

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 (Eq. 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 the Eq. (1).

$$Q_{target} = r_t + \Upsilon \cdot \max\Big[Q_{pred}(s_{t+1}, a') + +\beta \cdot \frac{1}{N}\sum_n^N Q_{pred}\,(n_{t+1}, a'\,)\Big], \tag{1}$$

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 (Eq. (2)).

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

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

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

$wt_{veh,t}/\ wt_{ped,t}$ 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.

3.5. MARL System with Dynamic Phase Duration: SAPA

The proposed traffic control framework employs a decentralized Multi-Agent Reinforcement Learning (MARL) system, in which each agent independently manages its intersection by observing local traffic conditions, selecting active signal phases, and storing experiential data. Because the intersections share similar structural characteristics, their experiences can be aggregated to train a shared neural network, enhancing overall learning efficiency and generalization. This design significantly improves adaptability compared to fixed-cycle systems, enabling real-time responsiveness to fluctuating traffic conditions.

A key innovation within this framework is the introduction of dynamic phase duration, guided by data gathered through VLC. This data provides continuous updates on vehicle queues and lane occupancy, allowing the system to adjust signal timing dynamically. Unlike conventional fixed durations (e.g., 8 or 12 seconds), green phases are extended or curtailed based on both upstream queue lengths and downstream occupancy levels. This mechanism, termed Strategic Anti-Blocking Phase Adjustment (SAPA), minimizes gridlock and maximizes throughput across interconnected intersections.

For instance, consider a corridor of five intersections, where traffic moves from intersection C0 to C1 via a 400 m link. If downstream lane occupancy remains below 40 %, the system proportionally extends the green phase at C0 according to the queue size; otherwise, it enforces only the minimum green time to allow C1 to discharge vehicles. Similar adaptive thresholds apply to shorter links (e.g., 200 m), where the occupancy limit is reduced to 35 % to prevent spillback.

The SAPA mechanism also supports multiple traffic management strategies, enabling flexible prioritization between radial and circular arteries. A baseline configuration distributes flow evenly (50–50), whereas alternative strategies prioritize one artery with up to 65 % of total vehicle allocation, modeling real-world scenarios such as dominant outbound or inbound movement. Lower-priority flows (e.g., left turns or minor approaches) receive reduced weightings of approximately 25 %. The most critical intersection (C1) dynamically balances radial and circular flows in alignment with the currently active traffic strategy.

By integrating neural network–based phase selection with adaptive duration control through SAPA, the framework achieves efficient, coordinated management of vehicular and pedestrian flows. This hybrid structure enhances scalability, robustness, and responsiveness across varying traffic conditions, offering a promising solution for complex urban networks.

4. Results and Discussion

This section discusses the results, including the performance of VLC-enabled communication within the MARL framework, the training and testing of the network, and the evaluation of the dynamic phase duration system.

4.1. VLC-enabled Communication Integrated with MARL for Traffic Coordination

Fig. 6a and Fig. 6b demonstrate the MUX signal and the decoded messages between the vehicles or the pedestrians and the traffic lights, respectively.

Results show that with VLC is possible to details the flow of V2I, V2V, P2I, and I2P communications at various intersections, illustrating a structured communication framework for coordinating traffic and pedestrian movement. VLC supports detailed and structured interaction between vehicles, infrastructure, and pedestrians across multiple intersections. When combined with MARL, this system enables intelligent, adaptive coordination of traffic and pedestrian flows.

- V2V and V2I Communication: Vehicles exchange real-time data with each other and with infrastructure–such as position, signal phases, and movement status–allowing MARL agents to learn optimal behaviors and synchronize traffic movements efficiently.
- P2I and I2P Communication: Pedestrians interact with the infrastructure via VLC, making crossing requests and receiving confirmations. MARL policies integrate this information to dynamically adjust traffic phases, enhancing safety and responsiveness.
- Coordinated Phase Management: Traffic phases at intersections (e.g., C0, C1, C2) are synchronized based on real-time VLC data. MARL optimizes these phase transitions to minimize conflict points and delays, improving flow in complex urban scenarios.

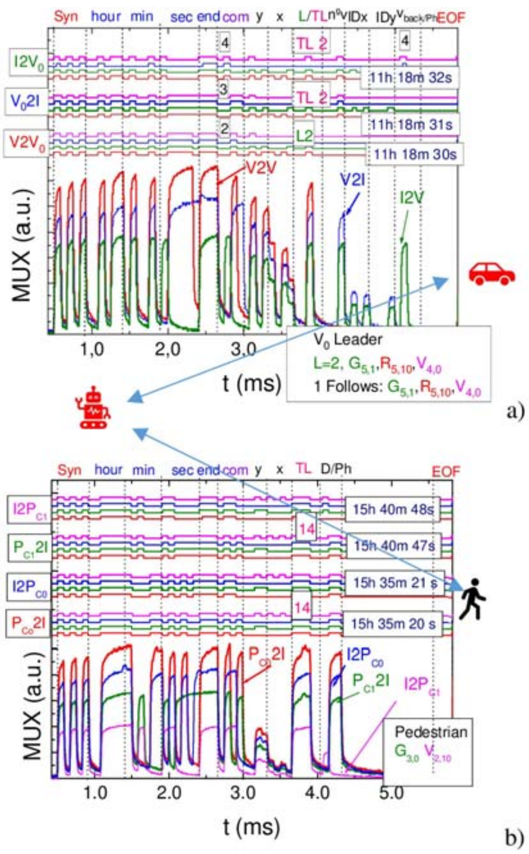

Fig. 6. Normalized MUX signal responses and the corresponding decoded messages, displayed at the top at various frame times, a)V2I, I2V; b) I2P and P2I.

In summary, the integration of VLC with MARL enables a decentralized yet harmonized traffic management strategy. VLC provides the high-resolution, low-latency data needed for MARL agents to learn and adapt to dynamic urban conditions, leading to safer, more efficient, and smarter intersections.

4.2. Network Training and Test

Multiple test scenarios were performed involving a total flow of 2200 vehicles per hour and 2000 pedestrians per hour, over 300 simulation episodes, each lasting 3600 seconds. Of all vehicles, 75 % followed a straight path, while the remaining 25 % turned either left or right. This turning behavior was symmetrically distributed across both main road arteries (see Fig. 3). The straight-moving vehicles were evenly split between the two arteries, ensuring a balanced traffic load and avoiding bias toward any specific direction. To assess the effectiveness of the proposed V-VLC system in multi-intersection environments, simulations were conducted using the Simulation of Urban Mobility (SUMO) [16], an open-source microscopic traffic simulation platform.

The experiments evaluated the traffic signal control system under two conditions:

- *β = 0:* no influence from neighboring intersections;
- *β = 0.3*: neighbor intersection influence included in the learning process.

This setup allowed the assessment of how inter-intersection coordination impacts system performance when using learning-based control strategies.

Fig. 7 illustrates the cumulative negative rewards obtained throughout the training process, enabling the comparison of learning efficiency. Both reward curves show a generally positive slope over the episodes, indicating a steady improvement in agent performance.

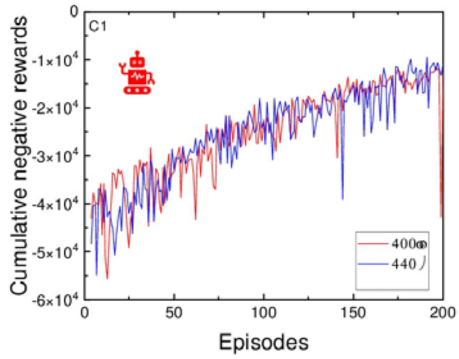

Fig. 7. Cumulative negative reward curves over training episodes for β = 0 and β = 0.3.

Fig. 8 displays the total number of vehicles and pedestrians remaining in the simulation environment during the tests.

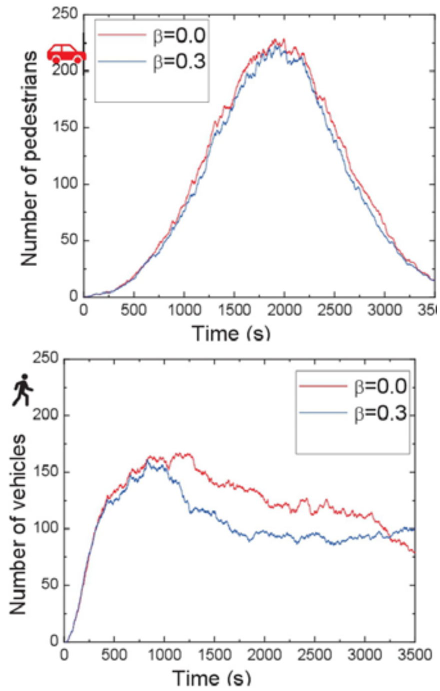

Fig. 8. Total number of vehicles and pedestrians remaining in the simulation environment for β = 0 and β = 0.3.

Fig. 9 presents a comparison of the trends over time for the number of stopped vehicles and pedestrians (Halting) as well as a vehicle speeds at C1 and C2 intersections, comparing scenarios with β = 0 and β = 0.3. The β = 0.0 scenario consistently results in a higher number of halted vehicles and lower speeds, indicating greater congestion.

The results clearly show that incorporating neighbor influence (β = 0.3) leads to significant improvements in traffic flow management. For vehicles, it notably reduces both congestion and halting time at the central intersection (C1), with minimal effects on the performance of other intersections. For pedestrians, waiting times decreased at all intersections, with the most marked improvement also observed at C1.

Additionally, the total number of vehicles and pedestrians remaining in the environment during the tests was lower when neighbor influence was considered, which is a strong indicator of the method's success in improving overall traffic throughput and reducing delays.

4.3. System with Dynamic Phase Duration: SAPA

This research introduces a dynamic phase duration mechanism supported by VLC-derived data on vehicle queues and lane occupancy, forming the basis of the SAPA framework.

In this formulation, T_{base} represents the base phase duration, set to 8 seconds. Q denotes the number of vehicles waiting when phases 5 or 6 are activated. The coefficients α_{rad} and α_{circ} correspond to the applied control strategy, which may either prioritize one of the arteries (65 % or 35 %) or maintain a balanced allocation (50/50 %) between them, depending on prevailing traffic conditions. In this case, only α_{circ} is considered, since at intersection C0 the dominant movements are associated with the circular artery rather than the radial one. The same reasoning applies to intersection C2, where circular flows also prevail.

$$T = \begin{cases} T_{base} + T_{base} \cdot Q \cdot \alpha_{circ/rad/low}, & \text{if } \rho_{C0-C1} < 40\ \% \\ T_{base}, & \text{if } \rho_{C0-C1} \geq 40\ \% \end{cases}$$

However, these are not the only percentages considered. A lower weighting factor, $\alpha_{low} = 25\ \%$, is also introduced for phases that play a secondary role at a given intersection. For example, at intersections C0 and C2, only the α_{circ} factor is applied to the circular artery phases, as the N–S phases are relatively marginal and therefore receive only the α_{low} increment in green time. Conversely, at intersections C3 and C4, the dominant phases correspond to the N–S movements, giving priority to the radial artery. In these cases, radial phases are extended using the α_{rad} factor, while circular phases receive only the α_{low} increment.

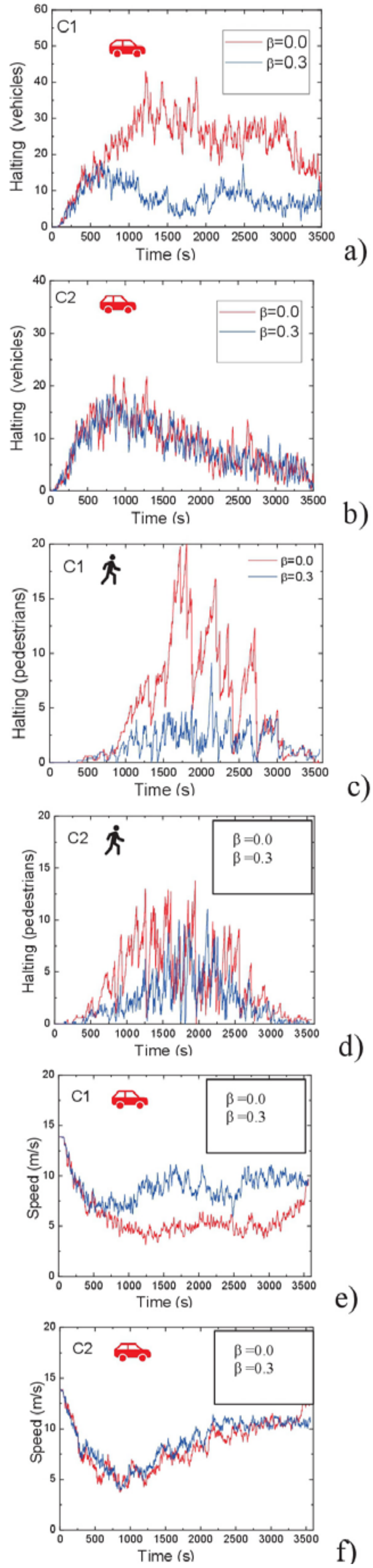

Fig. 9. Comparison of vehicle (a) and pedestrian (b) halting trends over time and vehicle speeds at C1 (e) and C2 (f) intersections for β = 0 and β = = 0.3

At intersection C1, both coefficients $\alpha_{rad}$and $\alpha_{circ}$are considered, since this is the critical intersection that must handle flows from both arteries according to the allocation defined by the selected strategy. Here, $\alpha_{low}$is applied only to the left-turn phases of both the N–S and W–E movements.

In scenarios where traffic moves from a 400-meter link to a 200-meter link, the occupancy threshold ρ must be adjusted downward. As the 400-meter link can store a larger number of vehicles, if the 200-meter link's occupancy is below 35 %, the corresponding phase receives its maximum extension. Otherwise, the phase duration is limited to the base time T_{base} = 8seconds, allowing the shorter link to clear before additional inflow occurs.

A comparison of vehicle halting at C0, C1 and C3 for both networks for the balanced strategy is exemplify in Fig. 10. During training, each neural network learns to interact with its environment under a specific traffic generation scenario reflecting its assigned control strategy. In testing, the models are evaluated across different strategies, with vehicle halting used as a primary performance metric.

Under Strategy 1, which maintains balanced allocation between radial and circular arteries, vehicle queues at the critical intersection C1 are notably higher in the fixed-duration network compared to the SAPA-enabled network. With SAPA, average queues remain around 20 vehicles, peaking briefly at 40, whereas the fixed-duration system often exceeds 40 vehicles and accumulates further over time. This improvement results from SAPA's adaptive phase extensions, which clear more vehicles per activation than the fixed 8-second phases, reducing overall clearance time and congestion. Neighboring intersections (C0, C2, C3, C4) remain stable, but vehicle counts are consistently lower under SAPA, highlighting its capacity to enhance flow efficiency.

Fig. 11 illustrates the active phases executed by the agents at each intersection over time for both neural networks under study.

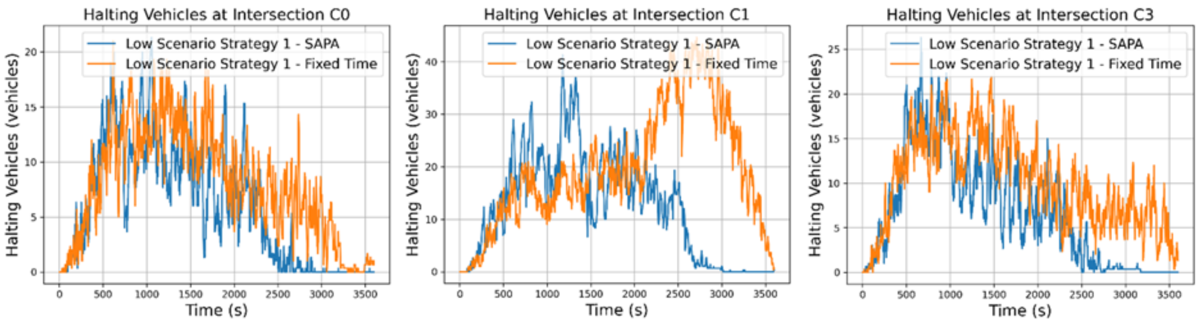

Fig. 10. Comparison of vehicle halting at each intersection for both networks under Network 1.

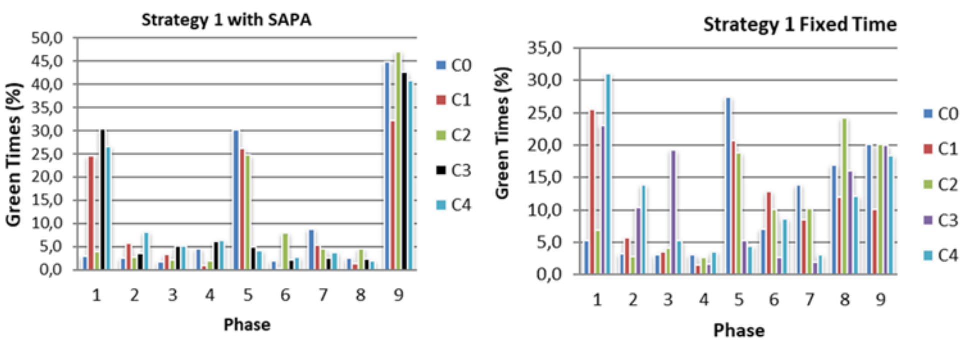

Fig. 11. Percentage of active phases over the simulation time for each intersection in both networks for Strategy 1.

Analysis of phase activations further illustrates the benefits of SAPA. Across intersections, dominant phases are P1 (N–S), P5 (W–E), and P9 (pedestrian). Weaker phases in both N–S (P2–P4) and W–E (P6–P8) directions are less frequently activated under SAPA, being largely replaced by pedestrian phases, which increase by over 20 % compared to the baseline. Despite this shift, balanced activation between P1 and P5 is maintained at the critical intersection C1, supporting effective vehicular flow while accommodating pedestrian movements.

For outer traffic cells, Strategy 3, which prioritizes circular flow, is most effective, whereas proximity to the urban core requires tighter control to prevent congestion. A progressive transition from Strategy 1 to Strategy 5 increases radial prioritization for inbound flows, ensuring smoother entry into the city. Conversely, outbound flows follow a sequence of Strategy 4 (radial emphasis), Strategy 1 (balanced), and Strategy 2 (circular dominance), enabling efficient redistribution of traffic and mitigating congestion. Overall, the network integrating the SAPA module demonstrates superior performance across all strategies, achieving higher intersection throughput, shorter queue lengths, and sustained smooth flow at the critical intersection C1, thereby preserving the efficiency and robustness of the entire traffic network.

5. Conclusions

This study proposes a traffic signal control system that combines VLC with MARL framework to adapt to real-time urban dynamics. The system relies on decentralized DRL agents that learn from local data and interact with neighboring agents through VLC-enabled communication, enabling collaborative and context-aware decision-making.

Results demonstrate that incorporating neighbor interactions (β = 0.3) leads to significant improvements in key traffic indicators, such as queue length, speed, and flow volume, by promoting coordinated learning across intersections. This cooperative strategy enhances responsiveness and overall system performance. To ensure smooth traffic flow, maintaining fluid segments within five-intersection clusters (C0–C2 and C3–C4) is advisable, particularly in less responsive network areas.

Overall, the MARL-based approach with agent cooperation reduces delays, optimizes signal phasing, and improves safety, confirming its scalability and effectiveness for large-scale, complex urban environments.

Acknowledgments

This work was sponsored by FCT within the Research Unit Center of Technology and Systems CTS/UNINOVA/FCT/NOVA CTS/00066 and IPL/IDI&CA2024/INUTRAM_ISEL.

References

[1]. H. Ge, Y. Song, C. Wu, J. Ren, et al., Cooperative deep Q-learning with Q-value transfer for multi-intersection signal control, *IEEE Access*, Vol. 7, 2019, pp. 40797-40809.

[2]. A. Vidali, L. Crociani, G. Vizzari, S. Bandini, A deep reinforcement learning approach to adaptive traffic lights management, in *Proceedings of the Workshop "From Objects to Agents" (WOA'19)*, 2019, pp. 42-50.

[3]. A. Oroojlooy, D. Hajinezhad, A review of cooperative multi-agent deep reinforcement learning, *Applied Intelligence*, Vol. 53, 2023, pp. 13677-13722.

[4]. M. Vieira, M. A. Vieira, G. Galvão, P. Louro, et al., Enhancing urban intersection efficiency: utilizing visible light communication and learning-driven control for improved traffic signal performance, *Vehicles*, Vol. 6, 2024, pp. 666-692.

[5]. M. Vieira, M. A. Vieira, P. Louro, M. Véstias, et al., Enhancing airport traffic flow: intelligent system based on VLC, rerouting techniques, and adaptive reward learning, *Sensors*, Vol. 25, 2025, 2842.

[6]. T. Antunes, G. Galvão, M. A. Vieira, M. Vieira, et al., Intelligent intersection management through multi-agent reinforcement learning and visible light communication integration, in *Proceedings of the 11th International Conference on Sensors and Electronic Instrumentation Advances (SEIA' 2025)*, 24-26 September 2025, Ponta Delgada, São Miguel (Azores Islands), Portugal 2025, pp. 80-85.

[7]. S. K. Yang, J. C. Li, H. B. Shi, Mix-attention approximation for homogeneous large-scale multi-agent reinforcement learning, *Neural Computing & Applications*, Vol. 35, Issue 4, 2023, pp. 3143-3154.

[8]. C. Zhu, M. Dastani, S. Wang, A survey of multi-agent deep reinforcement learning with communication, *Autonomous Agents and Multi-Agent Systems*, Vol. 38, 2024, 4.

[9]. Y. He, Y. H. Wang, F. R. Yu, Q. Z. Lin, et al., Efficient resource allocation for multi-beam satellite-terrestrial vehicular networks: a multi-agent actor critic method with attention mechanism, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, Issue 3, 2022, pp. 2727-2738.

[10]. M. A. Vieira, M. Vieira, P. Louro, P. Vieira, et al., Vehicular visible light communication for intersection management, *Signals*, Vol. 4, 2023, 457.

[11]. A. Yousefpour, C. Fung, T. Nguyen, K. Kadiyala, et al., All one needs to know about fog computing and related edge computing paradigms: a complete survey, *Journal of Systems Architecture*, Vol. 98, 2019, pp. 289-330.

[12]. J. M. Bilal, D. Jacob, Intelligent traffic control system, in *Proceedings of the IEEE International Conference on Signal Processing and Communications*, 2007, pp. 496-499.

[13]. S. Yousefi, E. Altman, R. El-Azouzi, M. Fathy, Analytical model for connectivity in vehicular ad hoc networks, *IEEE Transactions on Vehicular Technology*, Vol. 57, 2008, pp. 3341-3356.

[14]. W. -H. Shen, H. -M. Tsai, Testing vehicle-to-vehicle visible light communications in real-world driving scenarios, in *Proceedings of the IEEE Vehicular Networking Conference (VNC'17)*, 2017, pp. 187-194.

[15]. X. Liang, X. Du, G. Wang, Z. Han, A deep reinforcement learning network for traffic light cycle control, *IEEE Transactions on Vehicular Technology*, Vol. 68, Issue 2, 2019, pp. 1243-1253.

[16]. A. Lopez, J. Blanco, F. Gonzalez, et al., Microscopic traffic simulation using SUMO, in *Proceedings of the 21st IEEE International Conference on Intelligent Transportation Systems (ITSC'18)*, 2018, pp. 2575-2582.

[17]. M. A. Vieira, M. Vieira, P. Vieira, P. Louro, Optical signal processing for a smart vehicle lighting system using a-SiCH technology, *Proceedings of SPIE*, Vol. 10231, 2017, 102311L.

[18]. M. Papageorgiou, Overview of road traffic control strategies, *IFAC Proceedings Volumes*, Vol. 37, Issue 19, 2004, pp. 29-40.