Article

Enhancing Urban Intersection Efficiency: Utilizing Visible Light Communication and Learning-Driven Control for Improved Traffic Signal Performance

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

1   DEETC-ISEL/IPL, R. Conselheiro Emídio Navarro, 1949-014 Lisboa, Portugal; mv@isel.pt (M.A.V.); a45903@alunos.isel.pt (G.G.); plouro@deetc.isel.ipl.pt (P.L.); mario.vestias@isel.pt (M.V.); pedro.vieira@isel.pt (P.V.)
2   UNINOVA-CTS and LASI, Quinta da Torre, Monte de Caparica, 2829-516 Caparica, Portugal
3   NOVA School of Science and Technology, Quinta da Torre, Monte de Caparica, 2829-516 Caparica, Portugal
4   Instituto de Telecomunicações, Instituto Superior Técnico, 1049-001 Lisboa, Portugal
5   INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, 1000-029 Lisboa, Portugal
*   Correspondence: mv@isel.ipl.pt

Abstract: This paper introduces an approach to enhance the efficiency of urban intersections by integrating Visible Light Communication (VLC) into a multi-intersection traffic control system. The main objectives include the reduction in waiting times for vehicles and pedestrians, the improvement of overall traffic safety, and the accommodation of diverse traffic movements during multiple signal phases. The proposed system utilizes VLC to facilitate communication among interconnected vehicles and infrastructure. This is achieved by utilizing streetlights, headlamps, and traffic signals for transmitting information. By integrating VLC localization services with learning-driven traffic signal control, the multi-intersection traffic management system is established. A reinforcement learning scheme, based on VLC queuing/request/response behaviors, is utilized to schedule traffic signals effectively. Agents placed at each intersection control traffic lights by incorporating information from VLC-ready cars, including their positions, destinations, and intended routes. The agents devise optimal strategies to improve traffic flow and engage in communication to optimize the collective traffic performance. An assessment of the multi-intersection scenario through the SUMO urban mobility simulator reveals considerable benefits. The system successfully reduces both waiting and travel times. The reinforcement learning approach effectively schedules traffic signals, and the results highlight the decentralized and scalable nature of the proposed method, especially in multi-intersection scenarios. The discussion emphasizes the possibility of applying reinforcement learning in everyday traffic scenarios, showcasing the potential for the dynamic identification of control actions and improved traffic management.

Keywords: traffic management; intersection controlled by light; queue length; transmitters using white LEDs; silicon carbide light detectors; on–off keying (OOK) modulation method; density of pedestrians; model based on reinforcement learning

1. Introduction

Visible Light Communication (VLC) represents a cutting-edge technological paradigm, revolutionizing data communication through the innovative modulation of the intensity of the light produced by Light-Emitting Diodes (LEDs) [1,2]. This dynamic technology has a considerable impact on various applications, thanks to its straightforward design, operational efficiency, and wide geographic coverage. In the field of vehicular communications, VLC seamlessly integrates into the environment, as vehicles, streetlights, and traffic signals entirely adopt LEDs for illumination and signaling commitments [3]. This

integration extends to the use of exterior automotive and infrastructure lighting, such as streetlamps, traffic signaling, and head and tail lamps, for both communication and illumination purposes [4,5]. While VLC does have directional constraints, it is possible that the use of VLC in combination with other communication technologies will overcome this limitation. For instance, VLC can be employed for high-speed, short-range communication, while other wireless technologies such as Wi-Fi or cellular networks can complement VLC for broader coverage and omnidirectional connectivity.

Traffic lights equipped with VLC transmitters can not only control traffic but also transmit data to vehicles and roadside sensors, maximizing the utility of existing infrastructure. Utilizing VLC technology to optimize traffic signal efficiency represents a novel approach to urban intersection management. VLC offers advantages such as high data rates, security, and interference-free communication, which can revolutionize traditional traffic signal systems. VLC systems can be designed to provide precise localization capabilities, allowing traffic control devices to accurately determine the position and movement of vehicles and pedestrians. This enables more precise control of traffic flow, including adaptive signal timing and dynamic lane control.

The advent of VLC localization paves the way for advancing security, efficiency, and scalability in multi-intersection traffic signal control, particularly within the context of mixed traffic flows [5]. To tackle the hurdles of coordination, scalability, and integration, our solution involves implementing a traffic signal control system based on distributed reinforcement learning, specifically designed for Vehicular Visible Light Communication (V-VLC). The model's concept is inherently versatile and can be applied to any outdoor pedestrian setting, provided there is access to street database and traffic data. A validation of the mobility model was undertaken using Lisbon's city center as a case study, affirming its efficacy [6]. Incorporating learning-based control algorithms introduces adaptability and intelligence into traffic signal optimization. By leveraging machine learning or artificial intelligence techniques, the system can continuously adapt and improve its performance based on actual traffic conditions and historical data, leading to enhanced efficiency and responsiveness.

The main goal of the paper is to help with the progress of Intelligent Transport Systems (ITS) technology, with a focus on optimizing traffic safety and efficiency. This endeavor involves leveraging enhanced situation awareness and reducing accidents through various communication modes, incorporating Vehicle-to-Vehicle (V2V), Vehicle-to-Infrastructure (V2I), and Infrastructure-to-Vehicle/Pedestrian (I2V/P) communication [7–9]. Recognizing the shortcomings of the conventional control of the traffic light cycle, marked by extended delays, our focus shifts towards dynamic adaptations driven by real-time traffic data. The final goal is to enhance safety and traffic flow at intersections by deploying cooperative drive strategies [10,11]. The combination of VLC technology and learning-based control represents a synergistic approach to urban intersection optimization. By integrating these two innovative technologies, the system can achieve greater efficiency, reliability, and adaptability than traditional traffic management systems, ultimately leading to improved traffic flow, reduced congestion, and enhanced safety in urban areas. The proposed approach may also offer scalability and sustainability benefits, as VLC infrastructure can be relatively easy to deploy and maintain, while learning-based control algorithms can adapt to varying traffic patterns and environmental conditions over time, contributing to long-term urban mobility solutions.

The structure of the paper is as follows: Following the introduction, Section 2 provides an in-depth examination of the V-VLC system, outlining its architecture, communication protocol, and coding/decoding techniques. Section 3 presents experimental results, system evaluations, and a Proof of Concept (PoC) through a phasing traffic flow diagram based on V-VLC. In Section 4, we delve into an agent-based dynamic traffic control simulation using SUMO, an urban mobility simulator tool. Finally, Section 5 summarizes the paper's findings and conclusions, highlighting the transformative potential of V-VLC in traffic signal control and intersection management.

2. Traffic Controlled Multi-Intersections

2.1. Multi-Intersection Complexity

The differences between vehicles and pedestrians, including disparities in speed, size, and movement patterns, introduce additional complexities. Interactions between pedestrians and vehicles can impede each other, leading to reduced traffic flow efficiency and potential safety hazards. Striking an optimal balance between these two components of traffic presents a significant challenge that necessitates thoughtful consideration [12,13].

In the context of multi-intersection scenarios, another obstacle arises. While a straightforward solution involves a single entity controlling traffic lights across all intersections, scalability issues hinder this approach. The rapid increase in state and action spaces makes it impractical for real-time control applications. While optimization schemes for single intersections exhibit scalability within their domain, extending their effectiveness to multi-intersection environments requires innovative solutions.

Researchers [14,15] have delved into collaborative mechanisms to tackle this challenge, incorporating elements like queue length in adjacent intersections and modeling the interdependencies among these intersections. These attempts aim to strike a balance between scalability and efficiency in multi-intersection scenarios, acknowledging the necessity for a new approach to optimize traffic control effectively. Our adaptive traffic control strategy aims to adapt to actual traffic demand by modeling current and expected future traffic flow data. In contrast to conventional ground coil detectors used in traffic settings, an adaptive traffic control system operating within a Vehicle-to-Everything (V2X) environment has the capability to collect comprehensive data, including precise vehicle positioning, speed, queue length, and stopping durations. While V2V connections are especially vital for safety features like pre-crash detection, Infrastructure-to-Vehicle/Pedestrian (I2V/P) links offer connected vehicles and pedestrians (Ps) access to a diverse array of information [16,17].

2.2. V-VLC Communication Link

The communication system shown in Figure 1a is designed to make it easy for different parts of the traffic control system to share and process data smoothly. At the heart of this system is a hybrid mesh cellular structure, which includes two types of controllers placed at street and traffic lights. This setup is crucial for improving the system's performance and scalability [18] The mesh controllers are placed at streetlights along roads at strategic intervals, acting as central nodes in the network. Their main role is to relay messages to nearby vehicles efficiently, thereby ensuring the timely distribution of information like geo-distribution, pose ($q(x,y,t)$), and traffic notifications. Positioned at intersections, the mesh/cellular hybrid controllers play a multifaceted role within the system. They serve as border routers facilitating edge computing (V2I), enabling seamless integration between mesh and cellular networks. Additionally, they serve as gateways for data exchange between edge devices and the central cloud infrastructure (I2IM), establishing robust communication pathways to ensure uninterrupted data flow. The system utilizes embedded computing platforms to enhance data processing capabilities at the network edge. These platforms enable tasks like real-time sensor data processing, the precise detection of traffic flow patterns, and geo-distribution. Through local data processing, the system decreases response times and alleviates the load on the central cloud infrastructure.

The V-VLC system consists of a transmitter emitting modulated light and a receiver detecting differences in the received light. Both are connected via a wireless channel. The LED light is modulated using ON–OFF keying amplitude modulation. The environment features a grid of square cells arranged orthogonally, with tetra chromatic white light (WLED) sources at corners. These WLED sources combine Red (R: 626 nm), Green (G: 530 nm), Blue (B: 470 nm), or Violet (V: 390 nm) chips to generate white light, facilitating various data channels along roads and intersections.

The modulation and conversion of information bits from digital to analog are achieved through signal processing techniques. Figure 1b depicts the mapping of the coverage of
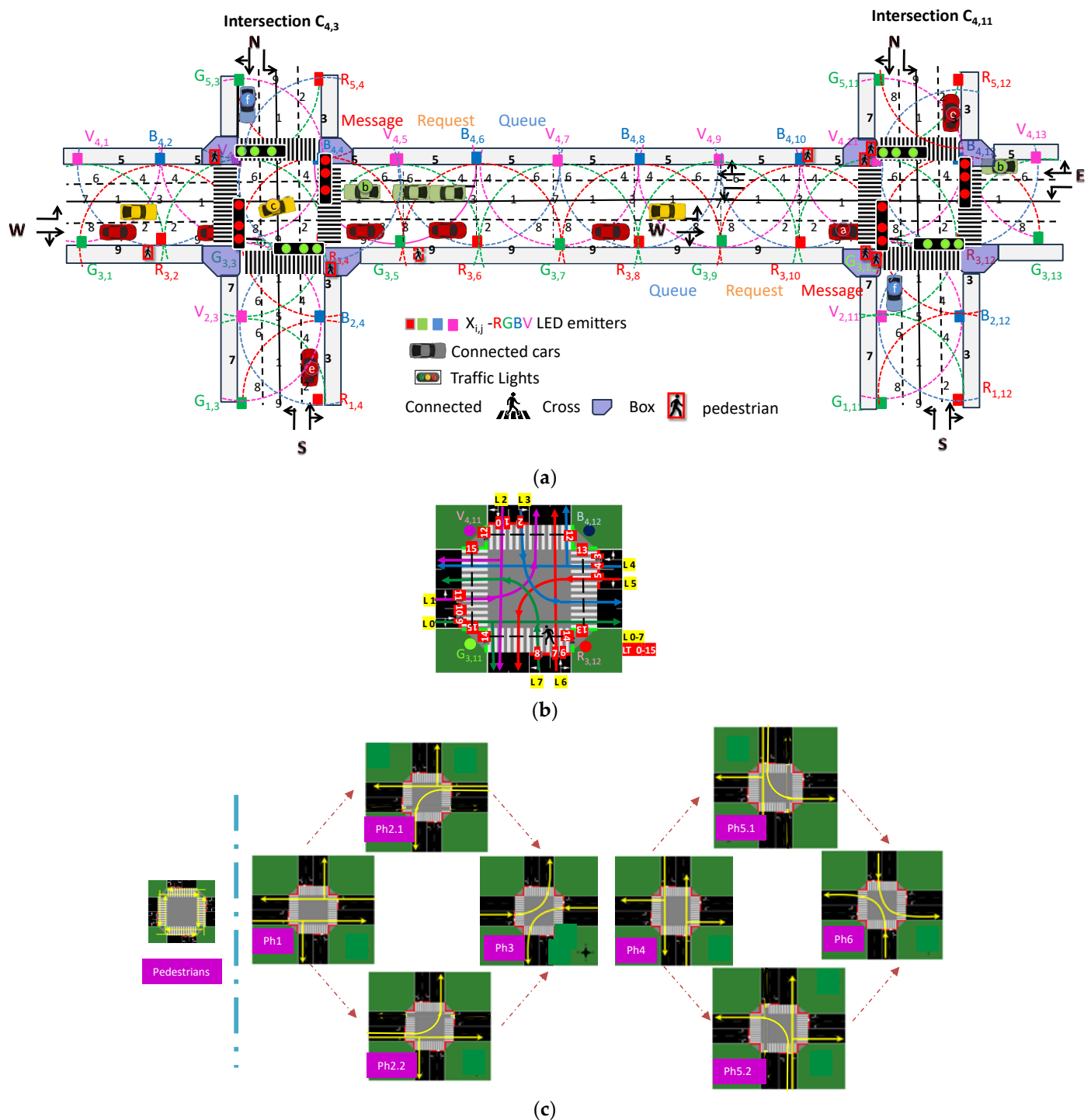
an intersection with four arms, highlighting nine distinct intersections (#1–#9) known as footprint regions, along the cardinal points; δ [19–22].

(**a**)

(**b**)

Figure 1. (**a**) Visual depiction in two dimensions of simultaneous localization relative to node density and transmission range. (**b**) Coverage map. Each region (footprint) is labeled from #1 to #9, and each region has a corresponding steering angle code ranging from 2 to 9 [22].

The system receives encoded signals from sources like road lamps and signal lights. Those indicators are meant for direct communication with identified vehicles (I2V) or indirect communication between vehicles using headlights (V2V). Each transmitter sends a message to vehicles (I2V) containing a unique identifier and traffic information. When a vehicle or pedestrian comes within range of the streetlight, upon receiving the light signal, the receiver reacts by allocating a unique identifier and the traffic message.

To control the flow of vehicles at intersections, methods such as queue/request/response and temporal/space relative pose concepts are used. PIN/PIN photodetectors with light filtering capabilities, integrated into mobile receivers, receive and decode the coded signals. The MUX receiver then combines various optical channels, performs different filtering processes (like amplification and switching), detects multiple signals, determines the centroid of received coordinates, and stores them as points of reference for the position. Nine reference points are identified for every unit cell, enabling the precise localization of pedestrians and vehicles within each cell. (See Figure 1b for illustration) [19].

2.3. Scenario, Environment, and Sequential Phases Used for the Simulation

The simulated scenario depicts a multi-intersection layout, as illustrated in Figure 2a, comprising a pair of four-way intersections (C1 and C2).

Figure 2. (**a**) Simulated scenario: depiction of an intersection with two sets of four arms and its surrounding environment featuring the optical infrastructure (X_{ij}), the resulting footprints (1–9), and the presence of connected cars and crossing pedestrians. (**b**) Identification of traffic lights (TL) and lanes (L), along with the illustration of possible trajectories for vehicles within an intersection. (**c**) Sequential progression of phases within the intersections, illustrating the evolution of operations over time [22].

Each intersection is equipped with two lanes on every arm, which approach from the cardinal points, leading to a configuration featuring two lanes on every arm. Each arm covers 100 m in length, with every lane measuring 3.5 m in width. Within each lane on every arm, specific directions for vehicle movement are delineated: the right lane accommodates

vehicles turning right or proceeding straight ahead, while the left lane only permits left turns. Positioned at the intersection, a traffic system, overseen by the Intersection Manager (referring to the agent) handles the flow of incoming traffic. Emitters (streetlamps) are strategically positioned by the roadside with a spacing of 15 m between them. Each lane is subdivided into three distinct segments, each serving a specific purpose: the first segment is dedicated to accommodating vehicles in motion or queuing along the lane (queue distances); the second segment is reserved for vehicles requesting permission to cross the intersection (request distance); and finally, the third segment, known as the message distance, is where vehicles receive the requested permission to proceed with crossing.

In Figure 2b, a schematic of the intersection is presented, depicting potential trajectories for vehicles and pedestrians, coded lanes, and traffic signals. Meanwhile, Figure 2c provides a visual illustration, offering insight into the sequential evolution of phases within the intersections. This carefully arranged process follows a precisely organized cycle duration, comprising a dedicated pedestrian phase and eight separate vehicular phases arranged into two segments. The sequence of these phases depends on the ever-changing traffic patterns. Each phase is further divided into specific time intervals or states, creating a detailed temporal structure that regulates the intersection's functionality [20,21]. Throughout the pedestrian phase, all vehicular traffic comes to a halt.

The "environment" is based on clusters of unit cells, forming an orthogonal topology as shown in Figure 2a. Each transmitter, labeled as X subscript i,j, has its own color (Red, Green, Blue, or Violet) and horizontal and vertical position (i,j) in the network. In PoC, crossroads were assumed to be located at the intersections of columns 3 and 11 with line 4. Figure 2a illustrates four distinct traffic flows along the cardinal directions. A binary choice (turn left/go straight or turn right) is provided in the road request and response segments.

Each simulated car represents a percentage of the traffic flow. We have assumed a total influx of 2300 cars per hour approaching the intersections, with 80% originating from the east and west directions. Subsequently, 25% of these cars are expected to make either a left or right turn at the intersection, while the remaining 75% will continue straight. The pedestrian influx is approximately 11,200 per hour, generated from both vertical roads and crossing the intersection in all directions, with an average speed of 3 km/h.

To illustrate the diverse traffic flows within a cycle, let us examine the following scenario:

- Twenty-four vehicles, from the west (W), approach the intersection. Among these, twenty vehicles (category a_1) continue forward, depicted by the red flow, while four vehicles (category c_1) exclusively make left turns, represented by the yellow flow.
- Vehicles from the east (E) contribute to the green flow, with thirteen vehicles (category b_1) continuing straight, and two vehicles (category b_2) making left turns.
- The orange flow originates from the south (S) and consists of six vehicles (category e_1). Within these, two vehicles take a left-turn approach (category e_2), while the other four continue straight.
- Lastly, the blue flow comprises thirteen vehicles (category f_1) arriving from the north. Nine of them proceed straight ahead, while the others execute a left turn at the intersection.

This breakdown offers a glimpse into how traffic is distributed across each flow, outlining vehicle movements such as going straight or making left turns at the intersection. The top three requests are assumed to be a_1, b_1, and a_2, pursued, respectively, by b_2, a_3, and c_1 in the fourth, fifth, and sixth positions. In the seventh, eighth, and ninth request positions are, respectively, b_3, e_1, and a_4. The tenth position is taken by c_2, followed by a_5 in the second-to-last request and f_1 in the final one.

2.4. Communication Protocol

To encode information, we utilized an OOK modulation scheme with synchronous transmission employing a 64-bit data frame. Each infrastructure is outfitted with tetra chromatic LEDs (refer to Figure 1b), allowing the concurrent transmission of four signals.

Consequently, the PIN/PIN receiver must possess the capability to actively filter each channel, resulting in a quadruple increase in bandwidth.

The communication protocol, as outlined in Table 1, identifies the structure and regulations regulating information exchange. This protocol includes specifications for synchronization, identification, and payload sections within the transmitted frame.

Table 1. Simplified Protocol For Communication.

		COM	Position		ID (veic)			Time				Payload			
L2V	Sync	1	x	y	0 bits		END	Hour	Min	Sec					EOF
V2V	Sync	2	x	y	Lane (0–7)	Veic. (nr)	END	Hour	Min	Sec	Car IDx	Car IDy	nr behind	EOF
V2I	Sync	3	x	y	TL (0–15)	Veic. (nr).	END	Hour	Min	Sec	Car IDx	Car IDy	nr behind	EOF
I2V	Sync	4	x	y	TL (0–15)	ID Veic.	END	Hour	Min	Sec	Car IDx	Car IDy	nr behind	EOF
P2I	Sync	5	x	y	TL (0–15)	Direct.	END	Hour	Min	Sec				EOF
I2P	Sync	6	x	y	TL (0–15)	Phase	END	Hour	Min	Sec				EOF

Each frame within the communication protocol (designated as 1–6) adheres to a structured format, starting with a synchronization block, followed by identification blocks, and ending with an End-of-Frame (EoF) block. This organized framework ensures a systematic and standardized communication protocol for the Visible Light Communication (VLC) system.

The synchronization block initiates the frame with a 5-bit sequence, represented by the pattern [10101], which synchronizes receivers and marks the start of a new frame. Identification (ID) blocks are crucial as they encode information using binary representation for coded decimal numbers. This information includes the type of communication (1–6), the location of transmitters (x, y coordinates), and timeline details (END, Hour, Min, Sec). The time sub-block, identified by the pattern [111], informs the decoder that the following bit sequence (6 + 6 + 6) pertains to time identification rather than payload. Other ID blocks contain essential data such as the number and temporary identification of vehicles following the leader, details about the occupied lane (Lane 0–7), traffic signal requests (TL 0–15), cardinal direction, or active phase conveyed by the infrastructure in a "response" or "request" message at the intersection.

The traffic message, forming the core of the message, furnishes additional critical information. This encompasses vehicle details, x and y coordinates, and the order of cars behind the leader seeking or receiving permission to cross the intersection (Car IDx, Car IDy, number behind). The traffic information payload includes road conditions, average waiting time, and weather conditions. The frame concludes with a 4-bit EoF block, identified by the pattern [0000], indicating the end of the frame.

2.5. Transmitted and Decoded VLC Signals

Each RGBV signal transmitted carries a specific wavelength-calibrated amplitude, defining its unique characteristics. With four independent emitters in each VLC infrastructure, the optical signal received can have one to four excitations, resulting in 2^4 distinct combinations leading to 16 different photocurrent levels at the photodetector. A filtering operation is obtained through a double PIN/PIN demultiplexer, a critical component in the decoding process. With pre-established knowledge of calibrated amplitudes, the PIN/PIN demultiplexer precisely decodes the transmitted message.

Aiming to clarify both the communication protocol (see Table 1) and the decoding technique using calibrated signals (Figures 1 and 2), Figure 3a provides a visual representation. This illustration showcases the decoded optical signals (depicted at the topmost part of the figures) and the MUX signals received within a V2I (code 3) and a V2V (code 2) VLC scenario. In this scenario, at "10:25:46", the leader, a_o, positioned on lane L0 (direction E) at $R_{3,10}$, $G_{3,11}$, $B_{4,10}$, communicates with the IM (agent) at C2, asking permission to cross and

informs the agent that, behind him, three additional vehicles (V_1, V_2, and V_3) positioned, respectively, at $R_{3,8}$, $G_{3,6}$, and $R_{3,4}$, are following him.
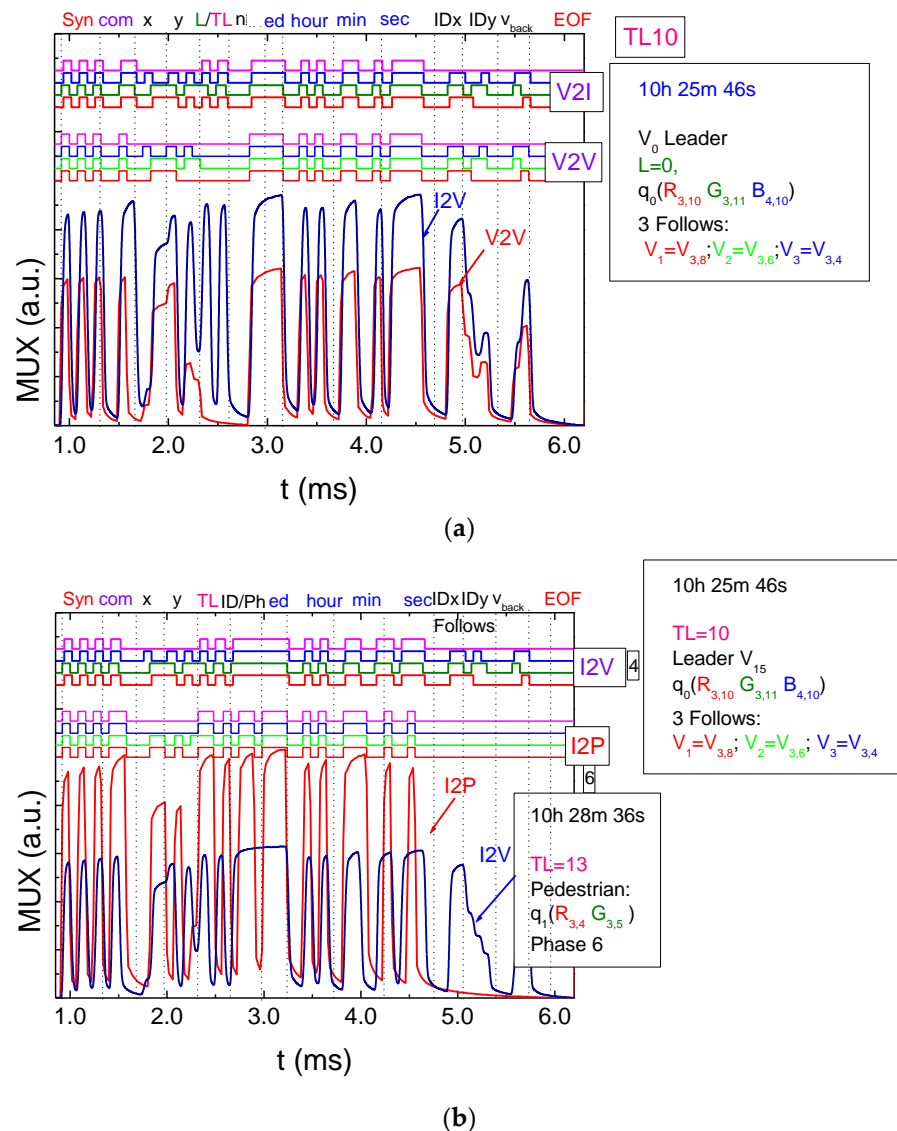

(**a**)

(**b**)

Figure 3. MUX signal requests (**a**) and responses (**b**) are categorized for different types of V-VLC communication. Deciphered messages are presented at the top of the figure [22].

Figure 3b demonstrates the infrastructure response, encompassing both I2V and I2P signals, issued by TL10 and TL13 traffic lights. These responses address the crossing requests initiated by the a_o vehicle and by the q_1 pedestrian positioned at the waiting corner "$R_{3,4}$, $G_{3,5}$ "of C1. The response from TL10 was transmitted at "10:25:46", while the response from TL13 was sent at "10:28:66". To investigate pedestrian behavior, two variables are needed: average pedestrian speed and halting. The former evaluates how the cycle durations of vehicles affect pedestrian speed, while the latter enables the analysis of the number of inactive individuals in waiting corners at all intersections, offering insights into population density in the waiting zone over time.

Figure 4a depicts the MUX signal transmitted to the traffic lights (TLs) by two pedestrians at the corners ($P_{1,2}2I$) to cross C1 and C2, respectively. The top part of the figure exhibits the decoded messages, while the content of the message is outlined on the right-hand side. Similarly, in Figure 4b, the MUX signal sent by the traffic lights ($I2P_{1,2}$) is depicted. The upper section of the figure displays the decoded messages, while the right-hand side offers a

summary of the message. This visual representation aids in understanding the communication between pedestrians waiting at corners and the corresponding traffic lights, shedding light on the signals exchanged for pedestrian crossings at both C1 and C2 intersections.
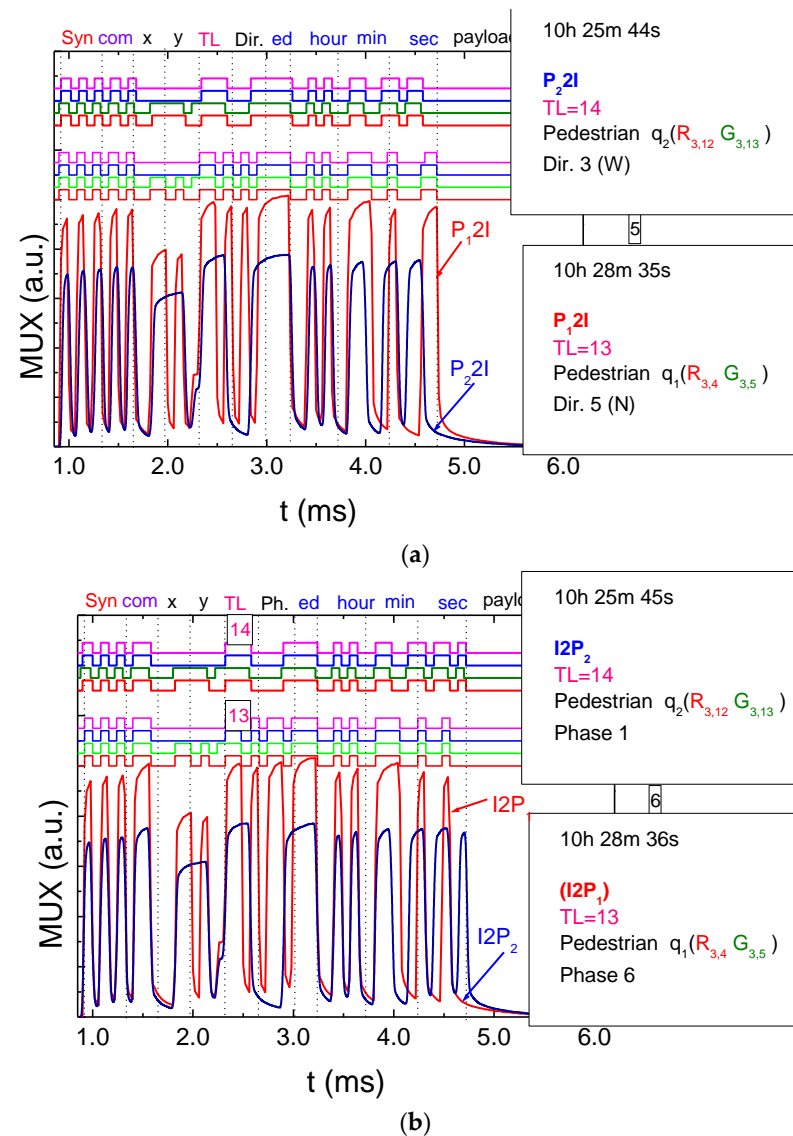

Figure 4. Requests and responses in normalized MUX signals and decoded signals (on the top) (**a**) transmitted by the waiting pedestrians ($P_{1,2}2I$) and (**b**) received by them ($I2P_{1,2}$) over different frame durations [22].

This illustration provides an understanding of the interaction between pedestrians and traffic lights across various intersections. The findings suggest that pedestrians initiate their crossing towards W, intending to traverse through TL14 waiting in positions $R_{3,12}$-$G_{3,13}$ before proceeding. At just "10:25:44", pedestrian 2 ($P_2 2I$), begins the communication with the TL14, and at "10:25:45", a response arrives ($I2P_2$). The pedestrian must wait until the pedestrian phase becomes active. With this information, it becomes evident that the current active phase is N-S (Phase 1) signifying that the pedestrian missed their designated phase (Phase 0). So, the pedestrian is required to wait for about 120 s before having the opportunity to cross.

3. Dynamic Traffic Flow Control: Simulation

This section introduces a dynamic control system model aimed at enhancing the efficient management of vehicular and pedestrian traffic at intersections. The model simulates expected outcomes resulting from implementing VLC technology for both vehicles and pedestrians. It utilizes information from V2V, V2I, and I2V communications to strategically make decisions regarding phase activation. This decision-making process prioritizes lanes with higher traffic, following a predetermined sequence of phases outlined in Figure 2b. Additionally, a comprehensive study analyzes the system's performance during high- and low-traffic cycles to estimate the number of vehicles efficiently managed within a one-hour timeframe.

3.1. SUMO Simulation: State Representation

The SUMO simulation environment, as shown in Figure 2, is constructed based on an existing Lisbon scenario. This scenario considers the impact of roads on traffic flow at two intersections. The traffic dynamics on the W–E arm, designated as the focal or "target" road, have a notable impact on traffic flow, with particular emphasis on this arm. The past influence on the target road of traffic conditions from other roads is restricted to a specific timeframe. The transmission of traffic flow and traffic waves quantifies the duration during which the traffic state of other roads affects the target road within the same timeframe. As vehicles continuously enter the system, the composition of traffic flow on the target road undergoes gradual changes over time, thereby influencing the cycle length at both intersections.

In order to improve traffic flow, adjustments were implemented to the originally suggested phases, as shown in Figure 2. These changes require a direct shift from the pedestrian phase (Ph0) to the N>S phase (Ph4), with subsequent phases proceeding as planned in both intersections. By reordering the phases and refining the traffic light control strategy based on simulation findings, enhancements in traffic flow, the alleviation of congestion, and overall intersection efficiency can be realized.

Regarding vehicle circulation, all vehicles are assumed to have an average speed of 10 m/s and a length of 4.5 m. However, as vehicles approach the traffic light at the beginning of the cycle, particularly during pedestrian evacuation, their speed is reduced to 5 m/s. Considering this adjusted speed, it is estimated that each vehicle requires approximately three seconds of green light to pass through the traffic signal. This represents the time needed for a vehicle traveling at 5 m/s to traverse a 15-m-long intersection. Therefore, considering the length of the cars, a minimum interval of 5 m between them is required to prevent collisions at this velocity. By incorporating this information into the incentive system, the agent is motivated to make decisions that optimize traffic flow, minimize delays, and ensure the efficient use of green light time, thereby enhancing overall intersection efficiency.

To accommodate pedestrians within the dynamic system, two scenarios were examined: the high- and low-traffic scenarios. In the high-traffic scenario, which lasts for 120 s, 76 cars are sent out, amounting to 2300 cars per hour. The low-traffic scenario, with a duration of 88 s, sends off 44 cars, equivalent to 1800 vehicles per hour. Each intersection experiences a pedestrian flow of 7200 at C1 and 4000 at C2. Pedestrians are introduced exclusively on the N and S roads, in both directions, at various distances from the intersection, mirroring real-life conditions where pedestrians originate from diverse starting points. All pedestrians are integrated into the SUMO simulator at a speed of approximately 1 m/s, which is equivalent to 3 km/h, a value closely resembling reality.

The IM, acting as the agent, strategically controls traffic signals to facilitate efficient and safe movement within the intersection. To achieve effective traffic optimization through learning, the state representation encompasses information about the environment, vehicle distribution obtained from V-VLC-received messages (refer to Table 1 and Figure 4b), and the proposed phasing diagram guiding agent actions (Figure 2b). The primary goal is to

minimize the total accumulated waiting time in each intersection arm, a metric calculated based on vehicle speed and queue alerts [21].

The reward function evaluates the difference in accumulated waiting time between the current and previous steps in all lanes, with negative rewards indicating higher waiting times. The agent learns to optimize traffic by taking actions (dynamic phases; Figure 2b) based on the current state, with training involving stored data samples to enhance decision-making. These decisions are then communicated to drivers and pedestrians through VLC response messages (Figure 3b), where the vehicle ID is assigned.

The agent's state, denoted as s_t, serves as a representation of the environment's situation at a specific agent step t. Its effectiveness in facilitating the agent's learning to optimize traffic is contingent upon furnishing ample information about the car distribution on each road. Figure 5 illustrates the state representation of the target road at the intersections throughout a simulated timeframe [22]. This representation incorporates discrete sub-cells designated for "response," "request," and "queue" zones, enabling the detection of vehicle entry into incoming lanes. Preceding the stop line of the intersection, each lane is divided into five cells: 0 for messages, 1 for requests, and 2 to 5 for queues. Each lane is equipped with its own dedicated traffic light, resulting in a total of 40 state cells during simulation, with lanes denoted as L/0–7 and traffic lights as TL/0–15. The simulation monitors the physical positions of waiting vehicles across lanes (L; 0–7) at C1 and C2. Each lane is segmented into small cells from the intersection, with each cell capable of accommodating a single vehicle. Sub-states provide detailed information regarding the nearest cell's position to the intersection and the maximum queue length.
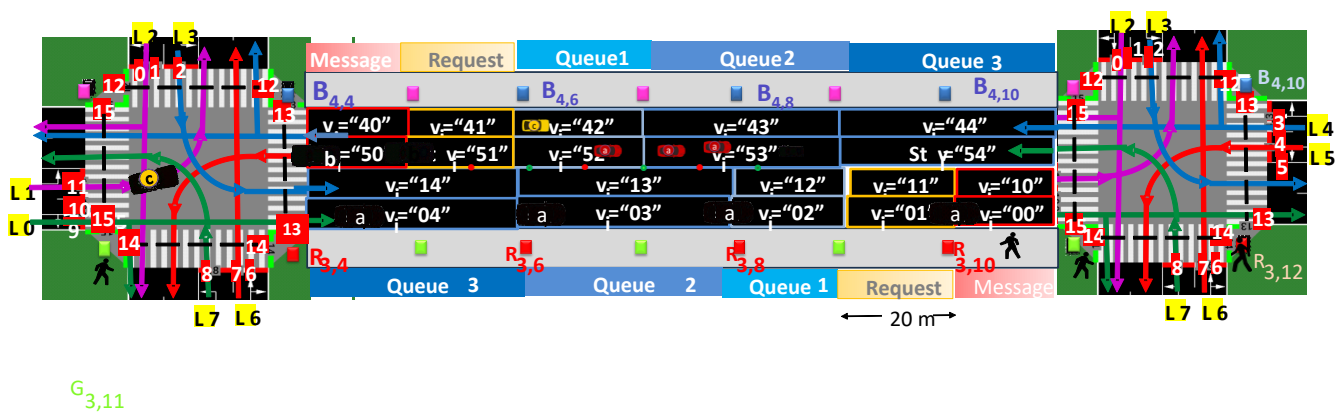

Figure 5. State representation (v_i) for the "target road" encompasses data on traffic lights (TL 0–15) and lanes (L 0–7), along with the visualization of vehicle and pedestrian trajectories [22].

The "complete state" refers to all the factors that contribute to the decision-making process. This could include various elements or aspects that are relevant to the environment or context in which an agent operates. Within this complete state, there are sub-states. These sub-states represent different facets or perspectives of the situation in a specific time step (t) as perceived by the agent. Each sub-state provides a unique representation of the environment at that particular moment. These representations help the agent understand and respond to the dynamic conditions of the environment at each step in its decision-making process. In the position state system at the intersection, a vehicle is referred to as "v_i", where "i" is the order of the request to cross, stated as a two-character sequence. The first character identifies the lane where the vehicle is located, while the second indicates its precise location within that lane. Referring to Figure 5, the states of the leader a_0 and subsequent vehicles are v_{15} = "00", v_{16} = "02", v_{17} = "03", and v_{18} = "04".

Each cell has the capability to measure the speed of a single vehicle. Vehicle speed is monitored during the simulation, representing the movement of vehicles among lanes (L; 0–7) segmented into small cells. Sub-states capture speeds ranging from "{0, 0.1, 0.2, . . ., 0.9, 1}". A speed of "=1" denotes the maximum legal speed, such as 90 km/h, while "=0"

indicates 0 km/h. As a result, the IM receives requests (V2I: illustrated in Figure 4a) from all leader vehicles and pedestrians seeking access to the intersection at different moments. The V2I data provide the IM with the precise location and speed details of all leading vehicles, as well as their followers' corresponding data, conveyed through V2V communication (Figure 3). Armed with this information, the IM can forecast the initial arrival times and speeds of vehicles at the different sections of the intersection.

In the queuing length system, the "queue length" denotes the count of stationary vehicles in a lane at the intersections. It fluctuates in response to incoming traffic and is influenced by departure rates. Vehicles at rest in the queue have a 0 km/h speed. The system's state is represented by the highest queue length across lanes (L; 0–7), and the number of possible states corresponds to this maximum among all lanes. For example, if the maximum queue length is 5, then the possible states could be "=0", "=1", "=2", "=3", "=4", and "=5". If there are three waiting vehicles in lane L5 at C 1, the queue states are indicated as"=1" for the three waiting vehicles and "=0" for the vehicles in motion. The queue length changes dynamically as vehicles arrive (increase queue length) and depart (decrease queue length). This representation allows for modeling and analyzing the traffic dynamics at the intersection based on the number of waiting vehicles in each lane. The goal is likely to optimize traffic flow and minimize congestion by understanding and managing the queuing system.

The traffic light state at each intersection changes between two states. When the signal is "Red Traffic Light (TL 5)," denoted as "=1", it indicates a red-light scenario. This state resets to "=0" when the light changes to green or yellow. Conversely, when the signal is "Green Traffic Light (TL 0)," represented as "=1", it signifies a green light situation. This state resets to "=0" when the light switches to red or yellow.

The traffic light phase state reflects the current traffic flow configuration at any given time "t". The simulation represents the current traffic phase at the intersection. For example, if "C 1 = (1, 0, 0, 0, 0, 0, 0, 0)", it signifies that only traffic phase 1 is currently activated (Figure 2c).

The simulation considers the speed of pedestrians at pedestrian traffic light corners (TL; 12–15). The average pedestrian speed reflects the movement of pedestrians during the simulation. The term "halting pedestrian" refers to the count of pedestrians waiting at a corner of intersections C1 or C2. This count fluctuates due to pedestrian arrivals and is influenced by cross rates. Pedestrians at a standstill have a 0 km/h speed. The system's state is characterized by the maximum number of halting pedestrians across pedestrian traffic light corners (TL; 12–15), and the number of possible states corresponds to this maximum count. For example, if there is a certain number of waiting pedestrians at corner TL14 of C1, the states are expressed as "=n" for the pedestrians in waiting and "=0" for those in motion.

3.2. SUMO Simulation: Cycle and Phases Durations

The SUMO Application Programming Interface (API) allows for seamless interaction with external programs, enabling smooth integration with the simulation environment. SUMO offers an extensive array of statistics pertaining to overall traffic flow. Additionally, it produces a range of results, such as diagrams that visualize the duration of individual states or the traffic light colors observed throughout the simulation.

Utilizing the scenario illustrated in Figures 2 and 5, we constructed a state diagram for the peak traffic scenario, integrating both vehicles and pedestrians through the SUMO simulation. Figure 6a,c showcase the phase diagrams for the interconnected intersections, C1 and C2, spanning two cycles lasting 120 s each. Meanwhile, Figure 6b provides the SUMO environment characterized by high pedestrian and moderate vehicle traffic flows.
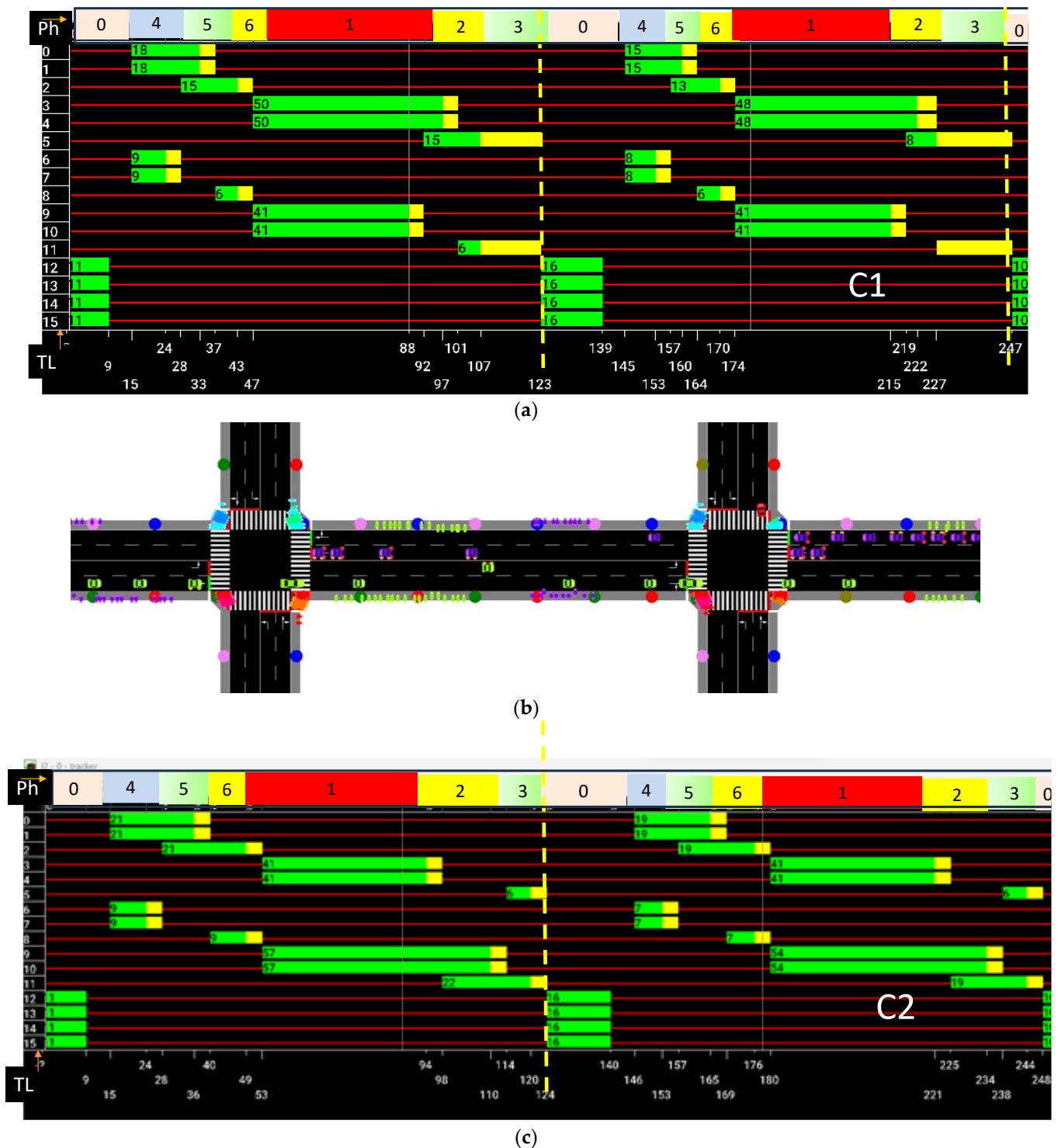
(a)

(b)

(c)

Figure 6. State phasing diagrams for two synchronized intersections are presented as follows: (**a**) Intersection C1; (**b**) the surrounding environment; and (**c**) Intersection C2. Phase numbers along the cycles are provided at the top of the state phase diagrams.

In Figure 6, we can discern the various cycles occurring during the simulation. It consistently kicks off with a pedestrian phase, allowing some individuals to cross the crosswalk, with the signal turning red for pedestrians after 11 s. Subsequently, phases dedicated to vehicles unfold until their conclusion at 123 s. Following this, the second cycle begins, marked by the reactivation of the pedestrian phase. This cycle repeats until 247 s,

marking the conclusion of the second cycle and the commencement of a third cycle. These diagrams correlate with the analysis conducted for pedestrians that ensues.

3.3. Dynamic vs. Intelligent Traffic Management: Leveraging VLC and DRL

Dynamic traffic management systems involve real-time adjustments to signal timings and phases based on the actual traffic conditions. These systems rely on ground sensors, cameras, and other data sources to monitor traffic patterns continuously. Adjustments are made reactively in response to changes in traffic flow, aiming to optimize traffic flow and reduce congestion. While dynamic systems are effective in managing immediate traffic issues, they may lack foresight in anticipating future congestion or optimizing long-term traffic management strategies. The integration of VLC into dynamic traffic control systems has represented a novel approach to improving urban intersections [22].

Intelligent traffic management systems utilize advanced algorithms and artificial intelligence to optimize traffic management strategies proactively. These systems analyze large datasets from various sources, including VLC-enabled infrastructure, vehicles, and pedestrians, to predict traffic patterns and optimize traffic flow. By leveraging predictive modeling, machine learning, and optimization algorithms, intelligent traffic management systems can anticipate congestion before it occurs and implement preemptive measures to mitigate its impact. They continuously improve over time, adapting to changing traffic conditions and optimizing long-term traffic management strategies.

Some advantages of using VLC and DRL can be summarized as follows: VLC technology enables the collection of real-time data from various sources, providing valuable insights into traffic patterns and behavior. By combining VLC data with DRL algorithms, predictive modeling can anticipate traffic congestion and optimize traffic management strategies accordingly. Leveraging VLC and DRL allows traffic management systems to take a proactive approach by anticipating congestion before it occurs and implementing preemptive measures to alleviate traffic congestion and enhance traffic flow. Intelligent traffic management systems using VLC and DRL continuously learn from past experiences and adapt their strategies accordingly. This iterative learning process optimizes long-term traffic management strategies, resulting in improved traffic efficiency and reduced congestion over time. Integrating VLC and DRL enables efficient resource allocation, allowing traffic resource allocation systems such as traffic light durations and phases more effectively. This ensures optimal traffic flow while minimizing delays and congestion at intersections.

So, while dynamic traffic management systems focus on real-time adjustments to traffic conditions, intelligent traffic management systems using VLC and DRL take a proactive and data-driven approach. By leveraging advanced algorithms and predictive analytics, these systems can optimize traffic management strategies, anticipate congestion, and improve overall traffic efficiency.

4. Intelligent Traffic Flow Control Simulation

In traffic control problems, RL-based approaches consider traffic flow states at intersections as observable states (Figure 5). Signal timing plan changes are actions, with feedback on control performance being crucial. This section details building an urban traffic control system using reinforcement learning [23–25].

4.1. Reinforcement Learning and Deep Q-Learning

Reinforcement learning (RL) [26] represents a category within the machine learning (ML) framework, wherein an agent undergoes a learning process by actively engaging with an environment [27]. The RL algorithm is very suitable for automatic control [28] and, therefore, a promising approach to intelligent traffic light control. The primary objective for these agents is to attain a goal within an environment characterized by uncertainty and potential complexity. Feedback, in the form of rewards or punishments, serves as the guiding mechanism for the agent's learning process. The underlying concept involves the agent acquiring optimal behaviors or strategies through a series of trial-and-error

experiences. The reward function assesses the disparity in accumulated waiting time across all lanes between the current and previous steps, with negative rewards denoting increased waiting times.

The reward function uses the accumulated total waiting time, $atwt_t$, as a metric which is defined in the following equation:

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

where $wt_{(veh,t)}$ denotes the duration in seconds during which a vehicle *veh* maintains a speed of less than 0.1 m/s at *agent step t* since its introduction into the environment, and *n* represents the total number of vehicles in the environment at agent *step t*. This metric ensures that when a vehicle exits without crossing the intersection, the $atwt_t$ value does not reset. The reward function, r_t, at agent *step t* is defined as follows:

$$r_t = atwt_{t-1} - atwt_t$$

with $atwt_t$ and $atwt_{t-1}$ denoting the accumulated total waiting time of all the vehicles in the intersection attained, respectively, at *agentstep t* and *agentstep t* − 1.

The agent optimizes traffic by taking actions (dynamic phases, as shown in Figure 2b) based on the current state, utilizing stored data samples during training to improve decision-making. These decisions are conveyed to drivers and pedestrians via VLC response messages (as depicted in Figure 3b), which include assigned vehicle IDs. At each discrete time step $t \in T$, the agent perceives its Markovian (or memoryless) decision-making factors (or state s_t) and obtains a state input based on the observed state of the environment and selects and performs an action (a_t) that transforms the observed state into a subsequent state (s_{t+1}). The reward (r_t) is then computed based on this action. Positive environmental rewards reinforce the likelihood of the agent reproducing the corresponding behavior, while negative rewards have the opposite effect. Following this action, the agent observes the subsequent state s_{t+1} and receives an immediate reward (or cost) $r_{t+1}(s_{t+1})$ which depends on the next state s_{t+1} for the state-action pair (s_t, a_t). The overarching objective is to maximize the cumulative discounted reward. Throughout this learning process, experiences in the form of (s_t, a_t, r_t, s_{t+1}) are stored in memory at each time step. Figure 7 provides a visual representation of the schematic for Deep Reinforcement Learning.

The replay memory comprises a dataset of an agent's experiences $D_t = (e_1, e_2, e_3...)$, accumulated as the agent interacts with the environment as time over time ($t = 1, 2, 3, ...$). In training, a batch of random samples is chosen to train the agent. This random selection of samples breaks the temporal correlation between consecutive samples. If the network learned only from consecutive samples of experiences as they occurred sequentially in the environment, the samples would be highly correlated and would therefore lead to inefficient learning. The neuronal network consists of a layered network, and the weight θ_k of the network is used to approximate its Q-values Q($s, a; \theta_k$) at iteration k.
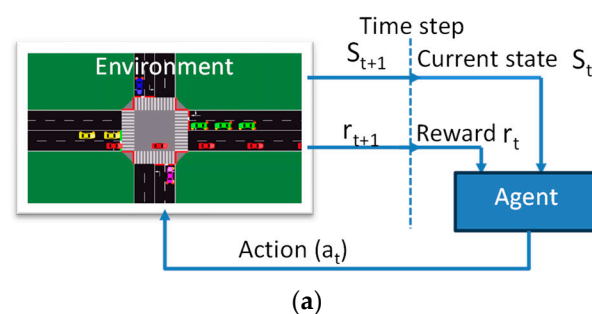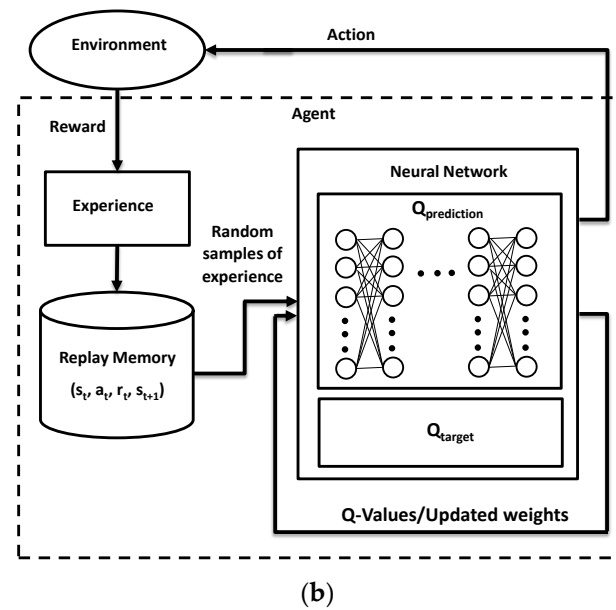

(a)

Figure 7. *Cont.*

(**b**)

Figure 7. (**a**) The Deep Reinforcement Learning schematic. (**b**) Scheme of the deep neural network used.

To train the agent, the deep Q-Learning technique is employed, leveraging the Q-Learning algorithm. The Q-value (quality value) represents the expected cumulative reward of taking a particular action in a particular state and following the optimal policy thereafter. This algorithm introduces the Q-Function, an action-value function that estimates the value of selecting action a_t at state s_t. The Q-Function predicts the expected cumulative and discounted future reward. In traditional Q-Learning, the algorithm maintains a look-up table storing the Q-value coupled with each state-action pair, earning it the name "tabular Q-Learning". This method guarantees convergence to the optimal value with infinite visits to state-action pairs.

However, this tabular approach is effective only for problems with small-scale state and action spaces. Real-world challenges with continuous and large-scale state and action spaces led to the adoption of deep Q-Learning networks. In this approach, a neural network predicts Q-values, taking the state as input and outputting Q-values for each possible action. This contrasts with estimating Q-values for each state-action pair separately.

Each traffic lane approaching an intersection is represented by 10 discrete cells, each of which represents the presence of a vehicle, resulting in a representation of the state of the environment of 80 cells per intersection. The input layer of the neural network is then composed of 80 neurons representing the state of the environment. Following this, there are five hidden layers, each containing 400 neurons with rectified linear units (ReLUs). The network concludes with an output layer featuring eight neurons, displaying the Q-values for each potential action. To enhance Q-value predictions, a Mean Squared Error (MSE) function is employed. MSE quantifies the disparity between predicted Q-values and target Q-values, contributing to the refinement of the learning process.

$$\text{MSE}_{\text{Loss}} = \frac{1}{N} \sum_{i=1}^{N} \left(Q_{target} - Q_{pred} \right)^2$$

N is the number of samples stored in memory, and the target and predicted value, Q_{target} and Q_{pred}, respectively. After each episode of training, the target Q-values for action-state pairs are calculated based on the following equation:

$$Q_{target} = r_t + \gamma . max Q_{pred} \left(s_{t+1}, a' \right)$$

where r_t is the reward obtained and γ is a discount factor applied to the $maxQ_{pred}$ value, lowering the importance of the future reward compared to the immediate reward.

The MSE$_{Loss}$ function calculates the squared difference between each predicted and target value. During training, the objective is to minimize this loss, indicating that the model strives to make predictions as close as possible to the true target values. The process involves iteratively adjusting the weights, θ_k, of the neurons in the neural network to decrease the difference between the initial prediction and the target, influenced by the learning rate.

Through repeated updating iterations, the neural network refines its approximation of the Q-value, bringing it closer to the target Q-value. As the loss decreases, the quality of the prediction improves. Consequently, the agent becomes more adept at making decisions regarding actions based on the observed environment. The iterative adjustment of weights enables the model to learn and adapt, enhancing its ability to navigate the environment and make informed choices over time.

4.2. RL-Based Traffic Control Model with VLC Integration

In reinforcement learning scenarios, we operate under the assumption that an agent, such as traffic lights, engages with its environment across a series of discrete time steps with the aim of maximizing rewards [29,30].

The agent's state, *st*, captures a representation of the environment's condition at a specific time step *t*. In the RL framework, the objective is to optimize traffic lights at two intersections (Figure 2), each comprising four arms of different lengths ranging from 160 to 400 m. It is important to notice that through multi-V2V communication among follower vehicles and V2I communication from the leader to the infrastructure, we ensure uninterrupted transmission within lanes ranging from 160 to 400 m in length.

The state representation integrates data on vehicle distribution and velocities across each road. PIN/PIN sensors, deployed at traffic lights, monitor vehicles within request and response distances through V2I, and indirectly at queue distances via V2V. The state space is structured with 32 cells per intersection, delineating lanes (L/0–7) and traffic lights (TL/0–15), discretizing the continuous environment (as depicted in Figure 5). This design incorporates spatial information on vehicle presence, speed, and discretized cells. Figure 8 showcases the grid layout of the agent's state space (indicated by dotted lines), underscoring its pivotal role in enabling the RL agent to learn and optimize traffic control policies based on observed conditions.
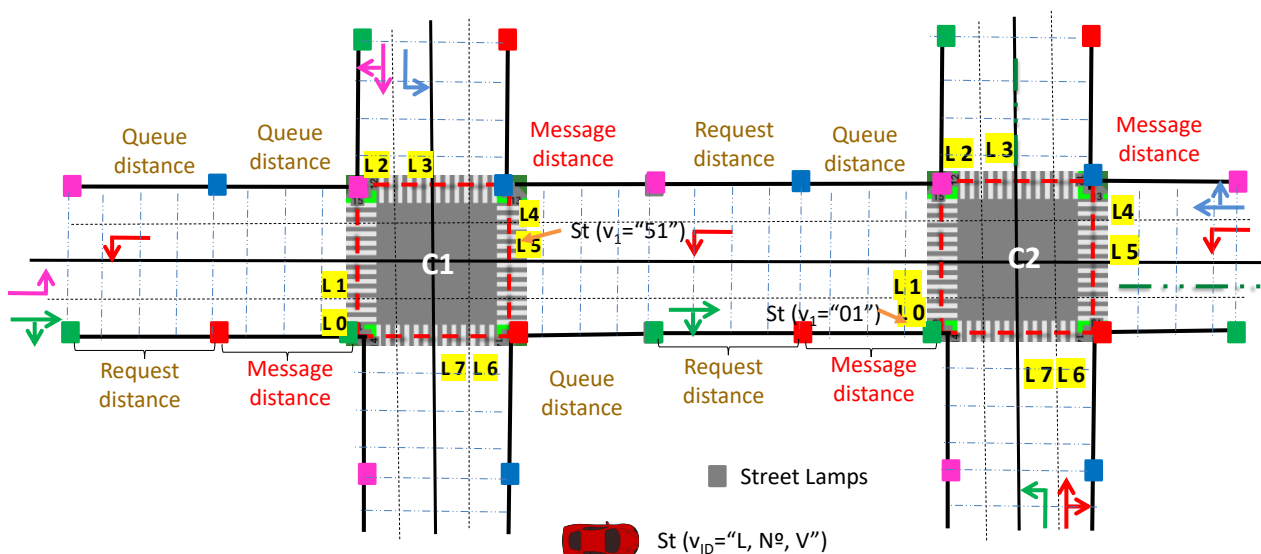

Figure 8. Agent state space grid representation with spatial information about vehicle presence and discretized cells.

The selection of the action space is a pivotal aspect of the RL model's effectiveness. In this scenario, a discrete action space is utilized, where the agent chooses a phase to execute at each time step t. The potential phases and their sequence for each intersection are predefined, as depicted in Figure 2.

The reward (*r*) signifies the environment's feedback to the agent's decision, serving as a measure of how beneficial or detrimental the agent's action was in terms of achieving specific objectives or optimizing performance metrics. This reward signal plays a crucial role in guiding reinforcement learning algorithms, shaping the agent's learning process, and enhancing its decision-making capabilities over time [31,32].

In this context, the total waiting time metric is employed, and a suboptimal action is defined by introducing more vehicles to queues in the current time step (*t*) compared to the previous time step (*t* − 1). This results in an increased cumulative waiting time compared to the previous time step, leading to a negative reward. The degree of negativity in the reward (r_t) corresponds to the magnitude of additional vehicles introduced to queues at time step *t*, reflecting a more unfavorable evaluation of the agent's action. Conversely, positive rewards are associated with good actions, where minimizing waiting times contributes to an improved traffic flow. This positive feedback incentivizes the agent to make traffic-light control decisions that improve overall traffic conditions. The training process is divided into multiple episodes, with the total number of episodes determined by the user, where 300 episodes are utilized in this instance. Each episode acts as a training iteration. During an episode, actions are executed based on the activation of specific lanes by the traffic light system, following predetermined timings during the green phases as depicted in Figure 2. This iterative training approach allows the RL agent to gradually learn optimal traffic control policies across multiple episodes, refining its decision-making based on feedback from the environment, particularly concerning waiting times and traffic conditions. The duration of the yellow phase is standardized at four seconds, while the green phase persists for eight seconds.

When the action taken in the current agent step (*t*) matches the action from the previous step (*t* − 1), no yellow phase is introduced, and the ongoing green phase is extended. On the contrary, when the action chosen differs from the previous one, a 4-s yellow phase is introduced between the two actions. This strategy ensures smoother transitions between distinct actions and allows vehicles ample time to adjust to evolving traffic signals. It is important to mention that in the SUMO simulation, each simulation step corresponds to one second, leading to eight simulation steps between two identical actions.

4.3. Implementing Symmetric Homogeneous Rewards in Training

In this study, two adjacent intersections within a (1 × 2) road network topology are examined, a setup previously used in dynamic system analyses. This configuration introduces nuanced considerations, particularly regarding the connecting roadways between the intersections. These roads serve as vital links for balancing traffic flow. Unlike scenarios involving a single intersection, traffic on these roads is influenced by the agent's decision to activate a phase allowing vehicle flow. However, a decision benefiting one intersection may detrimentally affect the other, potentially increasing pressure and wait times, and reducing overall traffic flow.

The observation made by the agent at each intersection is identical concerning the roadways and the occupancy of their cells. The distinction between the two intersections lies precisely in the decisions made by the agent. For instance, when the agent decides, at the first intersection (C1), to activate a green phase for the west direction in all directions, giving vehicles the possibility of going straight or turning right or left, this action will have a different impact on the environment when applied to the second intersection (C2), as can be seen in Figure 9a. At the first intersection when this phase is activated, the cars that do not go straight will leave the environment, while those that do go straight will take a critical lane, heading for the adjacent intersection. When this phase is active at the second intersection, regardless of which direction the cars are traveling, they will all leave the

environment and will not return to it. This difference will cause problems when training the network, as the experiences observed at the first intersection will not be identical to those at the second. To address this issue, a phase relationship has been proposed between the first and second intersections, ensuring that both become entirely identical and homogeneous.
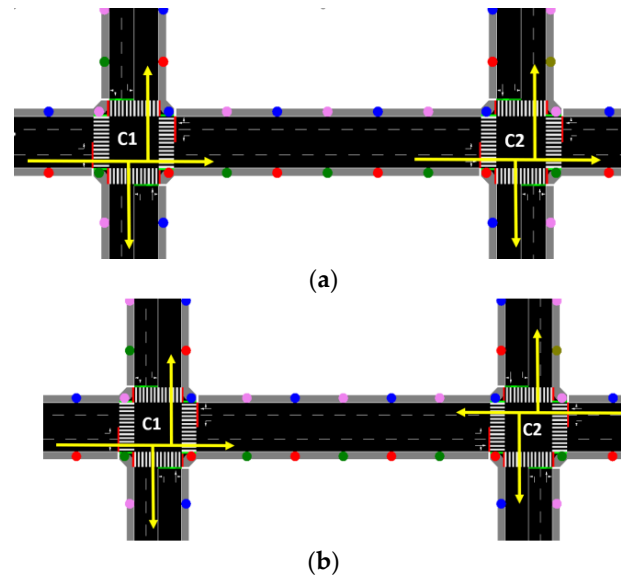

Figure 9. Agent's perception of (**a**) C1 and C2 intersections with north–south directions in both and of (**b**) C1 with a north–south direction and C2 with a south–north direction.

This approach allows for the attainment of an adjacent symmetric homogeneous reward, where actions taken at the first intersection have the same impact as those at the second, significantly contributing to reward improvement. The west all-direction action activated at the first intersection becomes equivalent to the east all-direction action at the second intersection (Figure 9b). The adjacent intersections with identical structures give rise to what is known as an adjacent symmetric homogeneous reward. This cooperative mechanism aids in the balancing of traffic flow between intersections and facilitates improved learning in both intersections, each with one agent.

Training typically involves multiple episodes (or epochs) to ensure effective learning from the data and convergence to an optimal solution. An "episode" refers to a single run or sequence of interactions that an agent undergoes with its environment from start to finish. The cumulative negative reward acts as a metric for evaluating the performance of the RL agent(s) in optimizing traffic control strategies throughout the training episodes.

Figure 10 displays cumulative negative rewards across successive episodes for intersections C1 and C2 in a 160 m (1×2) topology. States for training were obtained with either a single agent in C1 or C2, or with two agents, one in each intersection. This setup evaluates the RL model under different scenarios, including single-agent setups per intersection and the coordination of two agents, each managing one intersection (C1 or C2).

The results demonstrate that introducing a second agent accelerates the learning process with reduced oscillations towards the end of training. This behavior indicates the effective training of the network and validates the proposed solution's benefits for the traffic environment. Therefore, subsequent analyses and discussions assume the involvement of two agents in the learning process. This implies that collaborative efforts between agents in both intersections, C1 and C2, positively influence the learning dynamics, potentially leading to the more effective and efficient optimization of traffic control strategies in the multi-intersection environment.
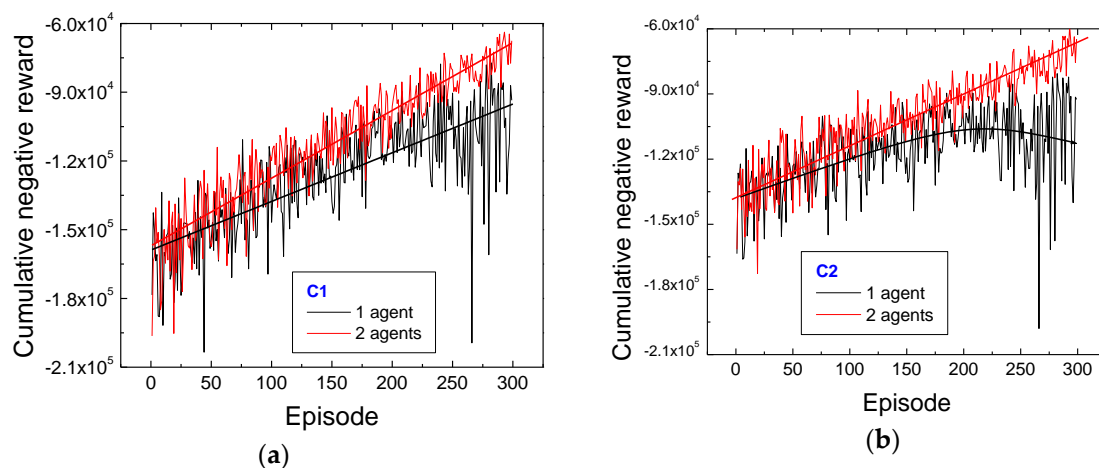
Figure 10. Cumulative negative rewards as a function of successive episodes acquired using a single agent or two separate agents at (**a**) intersection C1 and (**b**) intersection C2.

4.4. Analyzing the Performance of Neural Networks in High- and Low-Traffic Environments: A Study of a 160 m (1 × 2) Road Topology

Two scenarios were analyzed in a 160 m (1 × 2) topology: one with 2300 cars and the other with 1800 vehicles. The aim was to compare and contrast these scenarios with dynamic system findings, validating the feasibility of dispatching these car quantities within an hour. Neural networks for each scenario were trained over 300 episodes, each lasting 3600 s.

To characterize the scenarios, various traffic-related variables were utilized to assess the system's performance. These variables included queue sizes, with individual intersections in each scenario scrutinized to compare car flow. Additionally, the average queue size for each scenario was computed to gauge the impact of car numbers on the environment and the system's responsiveness in each instance. The average car speed was also considered, as it offers insights into traffic fluidity. Lastly, the number of cars halting (waiting) was analyzed to provide insights into the influence of vehicle volume on the environment.

Figure 11 depicts the queue length graph at both intersections (C1 and C2) for the scenario with 1800 and 2300 vehicles. It can be observed that until approximately 800 s, there is a significant increase in vehicles in the waiting queues, akin to a real-world rush hour scenario. There is a substantial influx of cars at both intersections, which gradually diminishes over time. During the neural network training, agents learn to make optimal decisions based on the observed environment. In testing, when agents are prompted to make these same decisions based on their observations, they respond accordingly, as evidenced by the decreasing number of cars in waiting queues over time. This results in clearing most of the vehicles from the intersections within the one-hour timeframe.

In the low-traffic scenario, with fewer vehicles in waiting queues, the intersections are less congested, aiding the agent in making better decisions and increasing the fluidity of vehicle movement throughout the environment. This translates to less time spent in waiting queues and more time in motion. Here, at around 3200 s, there were no longer any cars in the environment.

Figure 12a,b present a comparison that highlights the average speed and halting of vehicles in two distinct scenarios: one with 1800 vehicles per hour and the other with 2300 vehicles per hour.

By analyzing these factors, we aim to discern how varying vehicle volumes impact traffic dynamics and congestion levels.

As illustrated by the graphs, an evident peak in speed is noticeable during the initial phases of the halting simulations. This peak gradually diminishes over time as the simulation progresses. The initial flow in speed is attributed to the absence of vehicles at the intersections, allowing for smoother and faster movement. However, as the number of cars

entering the intersections increases, there is a significant decline in average speed. Towards the end of the simulation, as cars start to clear out, the average speed experiences an upturn due to reduced congestion. This trend reflects the dynamic nature of traffic, where higher volumes of waiting cars lead to decreased speed, while lower volumes result in increased speed, in accordance with expected traffic patterns.
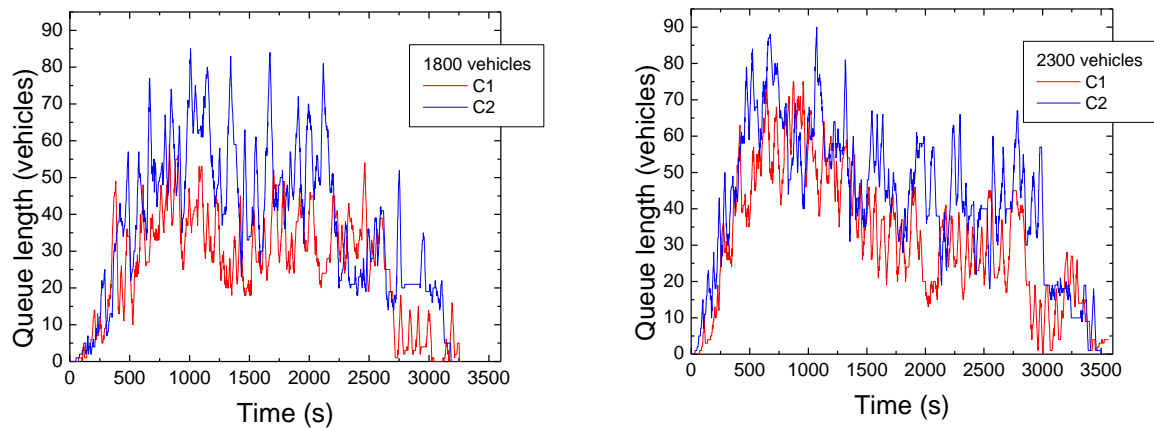

Figure 11. Queue length as a function of time in a scenario of 1800 (**left**) and 2300 vehicles (**right**) for both intersections during the training.

Figure 12. Comparative analysis of average speed and halting: (**a**) scenario with 1800 vehicles/hour and (**b**) scenario with 2300 vehicles/hour.

4.5. Inter-Intersection Roads: 160 m (1 × 2), 250 m (1 × 2), and 400 m (1 × 2) Road Network Topology

After examining the environmental impact of varying the number of vehicles, our focus shifts to investigating the size of critical lanes connecting two junctions. Each agent oversees its junction, monitoring lanes and car volumes through cell occupation. Following optimization in terms of intersection phase relationships, both intersections become homogeneous, rendering the experience identical. Despite this, inadequate communication among agents may elevate car volumes on critical roads. Agent decisions generate rewards based on vehicle wait times at respective junctions. When an action facilitates vehicle movement to target roads, the agent perceives it as beneficial locally, but this may adversely affect the adjacent intersection. Enhanced communication could manage actions based on neighboring intersection pressure. However, implementing this communication might escalate system complexity, potentially requiring a neural network for information exchange and facing scalability issues with more adjacent intersections.

Figure 13 illustrates the cumulative negative reward across successive episodes for the high-traffic scenario, where 2300 vehicles per hour are considered, across different target road lengths for both intersections. This depiction allows for an analysis of how varying road lengths impact the performance of the system in terms of negative rewards over time.
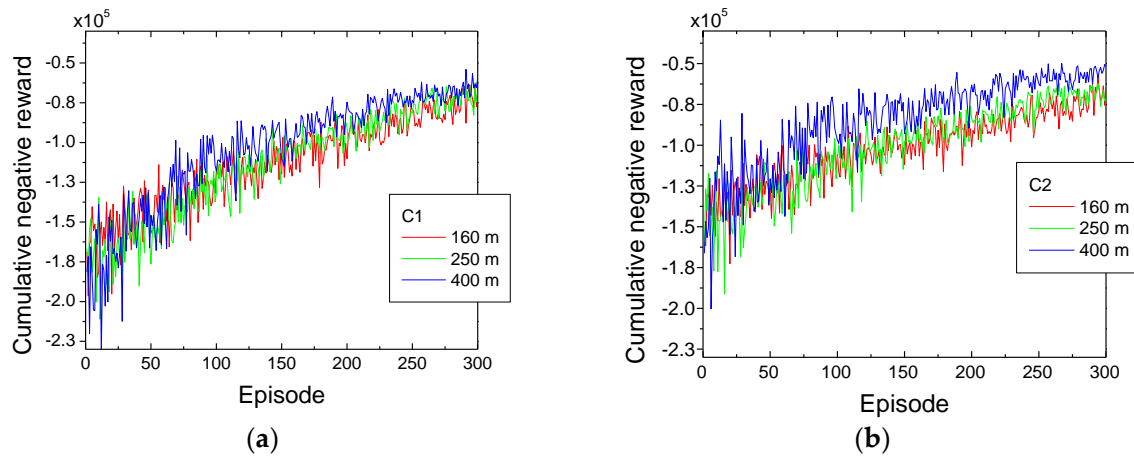

(**a**)   (**b**)

Figure 13. Cumulative negative reward through successive episodes in the high-traffic scenario (2300 vehicles/hour) and different target road lengths. (**a**) C1 intersection. (**b**) C2 intersection.

Here, the neural networks trained with different lane sizes exhibit expected reward behaviors. The findings indicate that with a higher target road length, waiting times decrease, leading to reduced queue sizes. This alleviates the pressure on the agent's junction and ensures sufficient space for vehicle circulation.

Figure 14a,b illustrate average queue sizes during network training episodes. The 400 m lane exhibits fewer queued cars than the other two, indicating minimal need for communication due to ample space for circulation. Conversely, for the 160 m and 250 m lanes, communication remains essential, as queue sizes are comparable to the 400 m lane, necessitating coordination to manage traffic effectively.
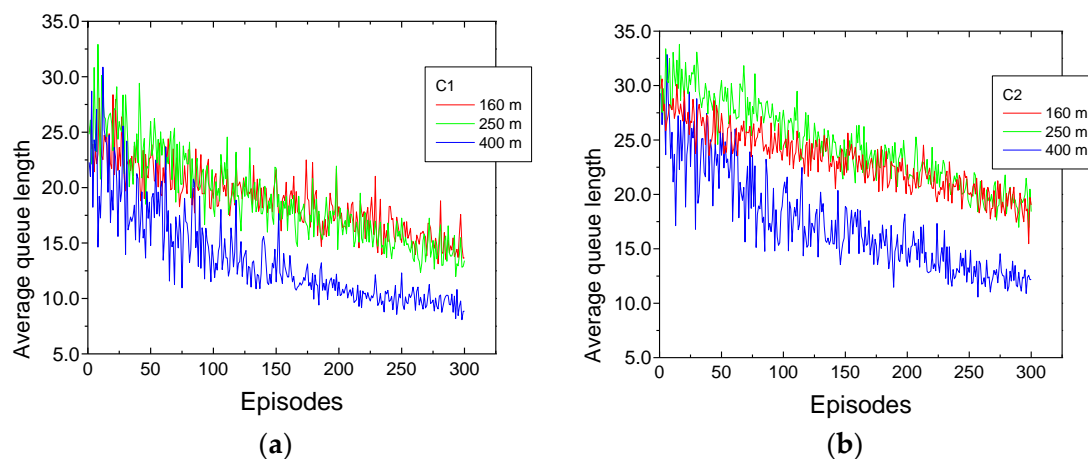

(**a**)   (**b**)

Figure 14. Average queue length (number of vehicles) across successive episodes in the high-traffic scenario (2300 vehicles/hour) and different target road lengths. (**a**) C1 intersection. (**b**) C2 intersection.

After completing the reinforcement learning (RL) training, we observed fluctuations in the learning curve, indicating challenges in achieving convergence. Nevertheless, the model demonstrated gradual improvement, reaching a moderate level of performance over the training period.

The results revealed consistent trends in both cumulative negative reward and average queue length at both intersections. Importantly, there was no significant separation between the cumulative rewards for the three types of road networks, highlighting the scalability of our distributed approach across road networks of varying sizes. The observed stability in these metrics, with a decreasing amplitude of oscillations as training progressed, suggests an enhancement in decision-making capabilities. Interestingly, in the shorter path, learning was faster initially but was later surpassed by longer paths as training advanced.

As anticipated, the average number of vehicles in the queue decreased at both intersections. Notably, the reduction in queue lengths was more pronounced and stable in the longer path at C1 compared to C2. This discrepancy can be attributed to the decreasing resistance of traffic flow with increasing path length, contributing to the observed effects.

In Figure 15, the average queue length across the time was tested for both intersections (C1 and C2) and different target road lengths.
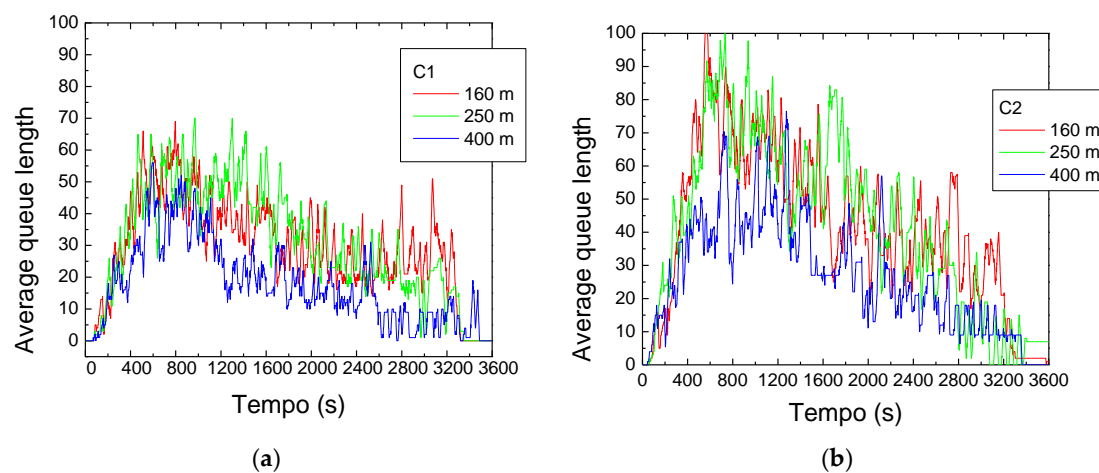

Figure 15. Average queue length (number of vehicles) test as a function of time in the high-traffic scenario (2300 vehicles/hour) and different target road lengths. (**a**) C1 intersection. (**b**) C2 intersection.

The observed average queue length can be explained by noting a notable surge in the number of vehicles in waiting queues until approximately 15 min, resembling a real-world rush hour scenario. Both intersections experience a substantial influx of cars, which gradually diminishes over time. Notably, as the road length increases, the queue length decreases at both intersections. Around the 45-min mark of training, at C1, there are no cars waiting, while at C2, the queue disappears only at the end. So, as the road length increases, fewer vehicles remain in waiting queues, resulting in less congestion at the intersections. This reduction in congestion aids the agent in making better decisions, ultimately enhancing the fluidity of vehicle movement throughout the environment. Consequently, less time is spent in waiting queues, allowing for more time in motion.

Throughout training, agents learn to make optimal decisions based on the observed environment. In testing, when agents are prompted to make these same decisions, they respond accordingly. This is evident in the decreasing number of cars in waiting queues over time, leading to the clearance of most vehicles from the intersections within almost half an hour.

Results show that reinforcement learning can optimize traffic flow by dynamically adjusting traffic signals, pedestrian crossing times, and other traffic management parameters. This adaptability helps reduce congestion, improve overall traffic efficiency, and minimize delays for both pedestrians and vehicles. Reinforcement learning is particularly effective in adaptive traffic signal control. Traffic signal timings can be dynamically adjusted based on current traffic conditions, reducing wait times and improving the overall throughput of intersections. In summary, reinforcement learning offers a flexible and adaptive approach to traffic management, providing the potential for significant improvements in efficiency,

safety, and sustainability in both pedestrian and vehicle traffic scenarios. While RL offers several advantages, it is focal to consider potential challenges such as safety concerns, ethical considerations, and the need for careful validation and testing before deploying RL-based traffic control systems in real-world scenarios.

Comparison with previous works is hard to achieve since there is not a well-established benchmark or traffic scenario that allows a fair comparison between different solutions. Previous works consider different environments, different traffic conditions, different reward metrics, etc.

Considering some previous recent works, in [22], a formal analysis of the queue problem is addressed. The number of vehicles is very low and not comparable to the problem under study. The work proposed in [23] does not control traffic lights. It considers a cross-road and autonomous vehicles controlled by an intersection manager using 5G technology. Similarly, work [24] does not control traffic lights. It considers wireless communication to detect vehicles and take transfer information. The work from [25] has a similar solution but only focuses on the RL algorithm. There is no associated communication technology. In this case, there are four lanes per arm. The number of cars as well as the number of lanes doubles. Comparing both, our cumulative reward is lower.

5. Advancements in Urban Traffic Management through Integrated Technologies and Innovative Strategies

This study involves the integration of emerging technologies, the enhancement of intersection efficiency, the development of multi-intersection traffic control strategies, and the application of reinforcement learning algorithms. These advancements have the potential to significantly impact urban traffic management and contribute to the development of more efficient and sustainable transportation systems.

The integration of VLC into dynamic traffic control systems has represented a novel approach to improving urban intersections [22]. VLC technology offers advantages such as high data transmission rates, low latency, and immunity to electromagnetic interference. Incorporating this emerging technology has also contributed to advancing our study in the field of intelligent transportation systems. Also using VLC technology, here we have added a proposal for an intelligent traffic control system leveraging advanced algorithms and artificial intelligence to optimize traffic management strategies. This system, in the future, can analyze large datasets collected from various sources, including VLC-enabled infrastructure, vehicles, and pedestrians, to predict traffic patterns and optimize traffic flow proactively. Intelligent traffic control systems can anticipate traffic congestion before it occurs and implement preemptive measures to mitigate its impact. They may also incorporate features such as predictive modeling, machine learning, and optimization algorithms to continuously improve traffic management strategies over time. While dynamic traffic control focuses on real-time adjustments to optimize traffic flow, intelligent traffic control systems using VLC technology take a more proactive and data-driven approach, utilizing advanced algorithms and predictive analytics to optimize traffic management strategies and improve overall traffic efficiency.

The primary aim is to enhance the efficiency of urban intersections. Improving intersection efficiency can lead to shorter travel times, reduced congestion, and enhanced overall traffic flow, thereby benefiting both commuters and cities. By leveraging VLC for communication between vehicles and infrastructure, coupled with RL algorithms for traffic signal optimization, the research addresses a critical need in urban traffic management.

The development of a multi-intersection traffic control system is essential for managing complex urban traffic networks. By optimizing traffic signals across multiple intersections simultaneously, it addresses the challenges associated with urban traffic congestion and coordination. This approach demonstrates a holistic perspective on traffic management, contributing to the advancement of urban mobility solutions.

The utilization of a reinforcement learning scheme for traffic signal scheduling represents an innovative approach. RL enables the traffic control system to adapt and learn

from real-time traffic conditions, leading to dynamic and adaptive signal control strategies. This adaptive nature enhances the system's responsiveness to changing traffic patterns, ultimately improving intersection efficiency and overall traffic management effectiveness.

By demonstrating the feasibility and efficacy of the integrated VLC-based traffic control system with reinforcement learning, the research provides evidence of its contribution. Real-world validation will enhance the credibility and applicability of the findings, showcasing the potential for practical implementation and impact.

6. Conclusions and Future Work

This paper sets the stage for future advancements in intelligent traffic management by emphasizing the potential of VLC technology in enhancing safety and efficiency at urban intersections through RL. The integration of VLC technology across pedestrians, vehicles, and surrounding infrastructure marks a significant breakthrough in optimizing traffic signals and vehicle trajectories. This integration facilitates the direct monitoring of critical factors such as queue formation, dissipation, relative speed thresholds, inter-vehicle spacing, and pedestrian corner density, ultimately leading to improved road safety.

Our dynamic control system model, designed to securely manage vehicular and pedestrian traffic at intersections, underwent detailed analysis under both high- (120 s) and low-traffic cycles (90 s) using the SUMO simulator. We introduced a SUMO extension for pedestrian modeling and made modifications to various tools within the SUMO package to facilitate the generation, simulation, and analysis of multi-modal traffic scenarios. The study aimed to assess the effective management of vehicles and pedestrians within a one-hour timeframe, taking into account various road network topologies.

In the realm of effective traffic optimization learning, our intelligent state representation incorporates environmental information, vehicle distribution from V-VLC messages, and a proposed phasing diagram guiding agent actions. A reinforcement learning model utilizing VLC technology to control traffic in dynamic scenarios was developed. Placing an agent at each intersection, the system optimizes traffic lights based on VLC-ready vehicle communication, calculating optimal strategies to enhance flow, and communicating with other agents to optimize overall traffic. The introduction of adjacent symmetric homogeneous rewards during training significantly improved the model's performance. Through training and testing, the reinforcement learning model showcased its ability to adapt to varying scenarios, emphasizing the importance of continuous learning in dynamic traffic environments. A comparative analysis of cumulative negative rewards across successive episodes and neural network tests for high and low vehicular scenarios using different road network topologies provided valuable insights into the model's efficiency and adaptability.

The improved results obtained with RL when compared to a traditional traffic control approach are traded-off by a higher computational cost since the RL requires the inference calculation of a neural network model. An optimized design is important to guarantee real-time computation in embedded systems near the sensors.

Future work will involve introducing the pedestrian phase, an aspect previously overlooked in the intelligent system. This addition aims to scrutinize agents' behavior, particularly regarding decision-making and environmental observations, with a focus on optimizing the activation timing of the pedestrian phase to ensure safety patterns for pedestrians. Relevant case studies will include analyzing the number of cars at intersections before initiating the pedestrian phase, pedestrian clearance time, and the number of individuals in waiting zones. Optimizing these factors will be crucial to ensuring an efficient system without a high concentration of people in designated areas.

Author Contributions: Conceptualization, M.A.V. and G.G.; validation: M.V. (Mário Véstias) and P.V.; formal analysis, P.L.; investigation, writing, and editing, M.V. (Manuela Vieira). All authors have read and agreed to the published version of the manuscript.

References

1. O'Brien, D.; Le Minh, H.; Zeng, L.; Faulkner, G.; Lee, K.; Jung, D.; Oh, Y.; Won, E.T. Indoor Visible Light Communications: Challenges and prospects. *Proc. SPIE* **2008**, *7091*, 60–68.
2. Parth, H.; Pathak, X.; Pengfei, H.; Prasant, M. Visible Light Communication, Networking and Sensing: Potential and Challenges. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 2047–2077.
3. Memedi, A.; Dressler, F. Vehicular Visible Light Communications: A Survey. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 161–181. [CrossRef]
4. 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]
5. Vieira, M.A.; Vieira, M.; Louro, P.; Vieira, P. Cooperative vehicular communication systems based on visible light communication. *Opt. Eng.* **2018**, *57*, 076101. [CrossRef]
6. 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; IEEE: Piscataway, NJ, USA, 2011; pp. 1–4.
7. Elliott, D.; Keen, W.; Miao, L. Recent advances in connected and automated vehicles. *J. Traffic Transp. Eng.* **2019**, *6*, 109–131. [CrossRef]
8. Bajpai, J.N. Emerging vehicle technologies & the search for urban mobility solutions. *Urban Plan. Transp. Res.* **2016**, *4*, 83–100.
9. Wang, N.; Qiao, Y.; Wang, W.; Tang, S.; Shen, J. Visible Light Communication based Intelligent Traffic Light System: Designing and Implementation. In Proceedings of the 2018 Asia Communications and Photonics Conference (ACP), Hangzhou, China, 26–29 October 2018. [CrossRef]
10. Cheng, N.; Lyu, F.; Chen, J.; Xu, W.; Zhou, H.; Zhang, S.; Shen, X. Big data driven vehicular networks. *IEEE Netw.* **2018**, *32*, 160–167. [CrossRef]
11. Singh, P.; Singh, G.; Singh, A. Implementing Visible Light Communication in intelligent traffic management to resolve traffic logjams. *Int. J. Comput. Eng. Res.* **2015**, *5*, 1–5.
12. Oskarbski, J.; Guminska, L.; Miszewski, M.; Oskarbska, I. Analysis of Signalized Intersections in the Context of Pedestrian Traffic. *Transp. Res. Procedia* **2016**, *14*, 2138–2147. [CrossRef]
13. Han, G.; Zheng, Q.; Liao, L.; Tang, P.; Li, Z.; Zhu, Y. Deep Reinforcement Learning for Intersection Signal Control Considering Pedestrian Behavior. *Electronics* **2022**, *11*, 3519. [CrossRef]
14. Fruin, J.J. *Designing for Pedestrians a Level of Service Concept*; Polytechnic University: Kowloon, China, 1970.
15. Eskandarian, A.; Chaoxian, W.; Chuanyang, S. Research Advances and Challenges of Autonomous and Connected Ground Vehicles. *J. IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 683–711. [CrossRef]
16. Pribyl, O.; Pribyl, P.; Lom, M.; Svitek, M. Modeling of smart cities based on ITS architecture. *IEEE Intell. Transp. Syst. Mag.* **2019**, *11*, 28–36. [CrossRef]
17. Miucic, R. *Connected Vehicles: Intelligent Transportation Systems*; Springer: Cham, Switzerland, 2019.
18. Yousefpour, A.; Fung, C.; Nguyen, T.; Kadiyala, K.; Jalali, F.; Niakanlahiji, A.; Kong, J.; Jue, J.P. All one needs to know about fog computing and related edge computing paradigms: A complete survey. *J. Syst. Archit.* **2019**, *98*, 289–330. [CrossRef]
19. Galvão, G.; Vieira, M.; Louro, P.; Vieira, M.A.; Véstias, M.; Vieira, P. Visible Light Communication at Urban Intersections to Improve Traffic Signaling and Cooperative Trajectories. In Proceedings of the 2023 7th International Young Engineers Forum (YEF-ECE), Caparica/Lisbon, Portugal, 7 July 2023; pp. 60–65. [CrossRef]
20. Vieira, M.A.; Vieira, M.; Louro, P.; Vieira, P.; Fantoni, A. Vehicular Visible Light Communication for Intersection Management. *Spec. Issue Adv. Wirel. Sens. Netw. Signal Process. Signals* **2023**, *4*, 457–477. [CrossRef]
21. Zhang, J.; Wang, F.Y.; Wang, K.; Lin, W.H.; Xu, X.; Chen, C. Data-driven intelligent transportation systems: A survey. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 1624–1639. [CrossRef]
22. Vieira, M.A.; Galvão, G.; Vieira, M.; Louro, P.; Vestias, M.; Vieira, P. Enhancing Urban Intersection Efficiency: Visible Light Communication and Learning-Based Control for Traffic Signal Optimization and Vehicle Management. *Symmetry* **2024**, *16*, 240. [CrossRef]
23. Elbaum, Y.; Novoselsky, A.; Kagan, E. A Queueing Model for Traffic Flow Control in the Road Intersection. *Mathematics* **2022**, *10*, 3997. [CrossRef]
24. Antonio, G.-P.; Maria-Dolores, C. AIM5LA: A Latency-Aware Deep Reinforcement Learning-Based Autonomous Intersection Management System for 5G Communication Networks. *Sensors* **2022**, *22*, 2217. [CrossRef]
25. Shi, Y.; Liu, Y.; Qi, Y.; Han, Q. A Control Method with Reinforcement Learning for Urban Un-Signalized Intersection in Hybrid Traffic Environment. *Sensors* **2022**, *22*, 779. [CrossRef]
26. Kaelbling, L.P.; Littman, M.L.; Moore, A.W. Reinforcement learning: A survey. *J. Artif. Intell. Res.* **1996**, *4*, 237–285. [CrossRef]

27. Shokrolah Shirazi, M.; Chang, H.-F.; Tayeb, S. Turning Movement Count Data Integration Methods for Intersection Analysis and Traffic Signal Design. *Sensors* **2022**, *22*, 7111. [CrossRef] [PubMed]
28. Genders, W.; Razavi, S. Using a deep reinforcement learning agent for traffic signal control. *arXiv* **2016**, arXiv:1611.01142.
29. Vidali, A.; Crociani, L.; Vizzari, G.; Bandini, S. A Deep Reinforcement Learning Approach to Adaptive Traffic Lights Management. In Proceedings of the WOA 2019, the 20th Workshop "From Objects to Agents", Parma, Italy, 26–28 June 2019; pp. 42–50.
30. Kővári, B.; Tettamanti, T.; Bécsi, T. Deep Reinforcement Learning based approach for Traffic Signal Control. *Transp. Res. Procedia* **2022**, *62*, 278–285.
31. Lopez, P.A.; Behrisch, M.; Bieker-Walz, L.; Erdmann, J.; Flötteröd, Y.; Hilbrich, R.; Lücken, L.; Rummel, J.; Wagner, P.; Wiessner, E. 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. [CrossRef]
32. Touhbi, S.; Babram, M.A.; Nguyen-Huu, T.; Marilleau, N.; Hbid, M.L.; Cambier, C.; Stinckwich, S. Adaptive traffic signal control: Exploring reward definition for reinforcement learning. *Procedia Comput. Sci.* **2017**, *109*, 513–520. [CrossRef]