



Article

Real-Time Adaptive Traffic Signal Control in a Connected and Automated Vehicle Environment: Optimisation of Signal Planning with Reinforcement Learning under Vehicle Speed Guidance

Saeed Maadi ^{1,2,*}, Sebastian Stein ³ , Jinhyun Hong ⁴ and Roderick Murray-Smith ³ ¹ Urban Big Data Centre, Department of Urban Studies, University of Glasgow, Glasgow G12 8QQ, UK² School of Engineering, Damghan University, Damghan 36716-41167, Iran³ School of Computing Science, University of Glasgow, Glasgow G12 8QQ, UK⁴ Smart City Department, University of Seoul, Seoul 02504, Korea

* Correspondence: saeedmoadi@gmail.com

Abstract: Adaptive traffic signal control (ATSC) is an effective method to reduce traffic congestion in modern urban areas. Many studies adopted various approaches to adjust traffic signal plans according to real-time traffic in response to demand fluctuations to improve urban network performance (e.g., minimise delay). Recently, learning-based methods such as reinforcement learning (RL) have achieved promising results in signal plan optimisation. However, adopting these self-learning techniques in future traffic environments in the presence of connected and automated vehicles (CAVs) remains largely an open challenge. This study develops a real-time RL-based adaptive traffic signal control that optimises a signal plan to minimise the total queue length while allowing the CAVs to adjust their speed based on a fixed timing strategy to decrease total stop delays. The highlight of this work is combining a speed guidance system with a reinforcement learning-based traffic signal control. Two different performance measures are implemented to minimise total queue length and total stop delays. Results indicate that the proposed method outperforms a fixed timing plan (with optimal speed advisory in a CAV environment) and traditional actuated control, in terms of average stop delay of vehicle and queue length, particularly under saturated and oversaturated conditions.

Keywords: connected and automated vehicles; adaptive traffic signal control; reinforcement learning; microscopic traffic simulation



Citation: Maadi, S.; Stein, S.; Hong, J.; Murray-Smith, R. Real-Time Adaptive Traffic Signal Control in a Connected and Automated Vehicle Environment: Optimisation of Signal Planning with Reinforcement Learning under Vehicle Speed Guidance. *Sensors* **2022**, *22*, 7501. <https://doi.org/10.3390/s22197501>

Academic Editor: Joel J. P. C. Rodrigues

Received: 19 July 2022

Accepted: 24 September 2022

Published: 3 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Traffic congestion has been a key urban issue, causing high economic costs in many cities worldwide. The report carried out by the Institute of Economic Affairs (IEA) claims that just a two-minute delay to every car journey costs the economy approximately 16 billion GBP a year, or nearly one percent of GDP (gross domestic product) in the UK [1]. Traffic lights play an essential role in controlling traffic flow and minimising delay, especially in an urban area. As a result, the number of traffic lights in England has increased by 25% since 2000, while the number of cars on roads grew by just 5% [1]. However, most of them use offline control systems based on the historical traffic flow and cannot respond to unexpected traffic situations (e.g., car accidents) efficiently or predict future traffic flows. Thus, adaptive traffic signal control (ATSC) approaches have been developed by numerous researchers [2–4]. Recently, advancements in computer performance and new optimisation methods have allowed researchers and practitioners to adopt heuristic methods for the ATSC process (e.g., real-time hierarchical optimising distributed effective system, RHODES [5]). These systems try to optimise traffic signal parameters in real time without considering a cyclic time interval. Thus, the signal plan could change at any time step depending on rapidly changing traffic conditions [6].

In recent years, autonomous vehicles (AVs) and connected vehicles (CVs) have been introduced due to the new generation of automation and connectivity technologies. Reference [7] reported that AVs could save the US economy roughly 211 billion USD annually with a 50% AV market penetration rate. They also suggested that the combination of automation and vehicle-to-everything (V2X) communication (vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication) is called connected and autonomous vehicles (CAVs), and they could further reduce congestion and fuel consumption. These will not only be an opportunity for transport planners to improve current capacity utilisation but will also be a research challenge for them in order to adapt the current transport systems to this new traffic environment [8]. As CAVs are gradually developed, they would also call for further research on intelligent signal control in the near future traffic environment [9]. Considering CAV technology, vehicles and traffic infrastructure can communicate with each other. For example, sending real-time trajectory data on speed, location and headway from CAVs to signal controls enables planners to optimise better traffic signal control plans based on accurate current and future traffic flow predictions [10–14]. In addition, CAVs can drive more efficiently by controlling vehicle speed and acceleration in response to signal timing plans, further improving the overall network performance (reducing the number of stops, fuel/energy consumption, and emission) [15,16]. Unfortunately, most existing CAV studies assumed fixed traffic signal timings to optimise the trajectory of CAVs [17–20].

Learning-based methods such as reinforcement learning (RL) have achieved promising results in lane-change behaviour under CAVs environment with self-learning and high-efficiency calculation [21]. In this work, we use RL in signal plan optimisation under the CAVs environment. In an RL approach, the agent (vehicles, traffic signal control) learns from the state (traffic environment) by taking actions (e.g., lateral control, signal plan changing) and observing the feedback rewards (e.g., lane changing time consumed, queue length). That can improve the lane-change manoeuvres or optimise the signal plan. In the signal plan case, agents learn by replicating traffic signal plans in a closed-loop system, leading to a model-free, self-learning framework [22] that can adapt to network-wide real-time traffic changes.

This study adopted an RL-based approach to consider both the optimisation of traffic signal control plans and CAV environments, given the above motivations and context. The highlight of this work is the speed guidance algorithm that helps to reduce the stopped delays of vehicles. The idea of combining a speed guidance algorithm with a reinforcement learning-based traffic signal control system is new in this area. In our RL approach, the intersection, as an agent, adopts traffic signal plans to optimise the total queue length. Simultaneously, the CAVs adjust their speed based on a fixed timing plan to decrease total stop delays, which are the waiting time for a vehicle when its speed is zero (practically when its speed is less than 5 mph) [23]. As a result, the agent adapts the traffic signal control to an adjusted speed of CAVs in real-time (updated traffic flow). Based on this framework, two different performance measures—minimising total queue length and total stop delays—are combined into a performance index (PI). The signal controller (agent) was trained in the fully dynamic traffic environment (traffic flows and CAV speeds) under different demand levels (unsaturated, saturated, oversaturated) and CAV penetration rate scenarios (0%, 25%, 50%, 75%, 100%) to show potential interaction effects between signal plan, CAV penetration rate and traffic flow. Results indicate that the proposed approach outperforms a fixed timing plan as well as conventional vehicle actuated control in terms of the average stop delay of vehicles and queue length (performance index). This work uses a microsimulation model in the VisVAP module of the Vissim 20 program for actuated control that is based on the guidelines of the Federal Highway Administration (FHWA) [24]. The actuated control prioritises the phase with the primary approach (main road), and detector actuation partially controls each phase's time.

The rest of this paper is organised as follows. Section 2 presents the relevant literature review. Section 3 presents our integrated reinforcement learning (RL) and speed guidance approach under the CAVs environment. In Section 4, the simulation setup is presented,

followed by our evaluation results in VISSIM microsimulation software. Finally, in Section 5, we conclude and provide ideas for future work.

2. Related Work

Adaptive traffic signal control approaches, including SCOOT [25] and SCATS [26], have been widely used in real-world traffic networks. They are based on an open-loop control system that does not consider feedback control in the traffic network (Figure 1). They use a cyclic system with pre-determined time intervals, which means the controller updates the signal timing plan (cycle length, green signal ratio, and phase difference) at a specific time interval [27]. Studies show that the traffic flow at intersections may vary significantly in major cities due to the fluctuation of traffic demand [28,29]. Nevertheless, these typical adaptive traffic signal control systems cannot respond to such travel demands (traffic flow varies at shorter time intervals) [6] and require complex computation schemes that make their implementation costly [30]. This can increase travel delays for road users.

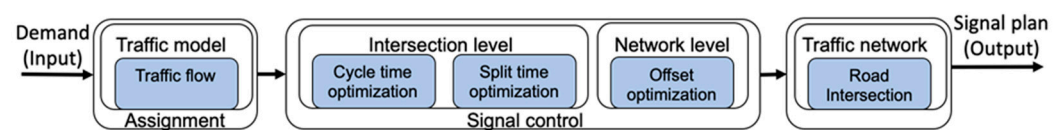


Figure 1. Open-loop traffic signal control system.

As mentioned in the introduction section, adaptive traffic signal control in a connected vehicle environment has shown a positive effect on the improvement in network efficiency [10–13]. Connected vehicle technology is a mobile data platform that allows real-time information to be exchanged among vehicles and between vehicles and infrastructure [31].

2.1. Traffic Signal Control under CAVs

As mentioned in [15], there are many novel approaches for CAV-based traffic control. One of the most common methods, which can be seen in various studies, is the ‘advanced driver guidance’. In this approach, vehicle speeds and positions are adjusted to minimise some performance measures. Reference [32] proposed an integrated traffic control model to optimise the total delay, and the decision variables for this research were vehicle arrival time and signal timing. They simulated the speed guidance model of CVs in VISSIM microsimulation software. They concluded that this method could significantly decrease vehicle delays and the number of stops. Reference [33] introduced GLOSA and assessed its benefits in reducing vehicles’ stop time behind a traffic light and fuel consumption using an integrated cooperative ITS simulation platform. Under two simulated scenarios, reference [34] investigated the positive impacts of the speed guidance on fuel consumption and driving behaviour for multiple signalised intersections. Other methods can be found in the literature, for instance, planning-based traffic signal control [35], platoon-based traffic signal control [36], and signal vehicle coupled control (SVCC) [15,37]. Unfortunately, most existing CAV studies assumed fixed traffic signal timings to optimise the trajectory of CAVs [18]. Unlike the current research methods, this paper presents an adaptive traffic signal control under the CAVs with the advanced traffic control (speed guidance model).

2.2. Closed-Loop Signal Control

Compared to typical adaptive traffic signal control (open-loop) systems, two general closed-loop signal control approaches exist. The traditional signal control approaches (i.e., non-learning-based approach) use a simple feedback loop control system and utilise only the current traffic flow but not historical traffic flow data. In addition, these approaches do not have underlying models for state prediction and optimisation. A learning-based method such as reinforcement learning can learn from the traffic environment by taking actions (i.e., cycle length and phase split) and observing the feedback that can help us predict the traffic flow and optimise the signal plan [32].

2.2.1. Non-Learning-Based Approach

Some studies propose a simple closed-loop system (non-learning-based approach) in traffic signal control. Reference [38] presented a decentralised feedback control mechanism aiming to equalise the degree of saturation and queue length on different approaches toward intersections in a network. Reference [39] showed a localised feedback speed control for mainline traffic on motorways. The backpressure controller is a distributed feedback system that does not require knowledge of global network inflow. Reference [40] introduced the traffic-responsive urban control (TUC) strategy as a network-wide feedback approach. Based on store-and-forward modelling of the urban network traffic and using the linear-quadratic regulator theory, the design of TUC leads to a multivariable regulator for traffic-responsive coordinated network-wide signal control that is also particularly suitable for saturated traffic conditions. Reference [6] proposed an upgrade closed-loop feedback signal control strategy, which takes the total amount of instantaneous stopped delay (ISD Total) as input detector data for measurement frame-by-frame in traffic flow video and realised real-time switches of the signal status when the amount reaches the threshold, and adaptively distributes the green interval to the most needed approaches (east–west and north–south) without the regular traffic signal cycle time. In other words, they still used a non-learning-based method but with new forms of data.

2.2.2. Learning-Based Approach (Reinforcement Learning)

Reinforcement learning-based adaptive traffic signal control changes traffic signals based on the feedback from the traffic demand, which can be hypothetical dynamic [41–44] or based on real-world data [45,46]. The existing literature on using the reinforcement learning approach can be categorised into two groups: networks consisting of CVs [44,47,48] and non-CVs environments [45,46,49]. Moreover, two general classifications (i.e., vehicle positions and queue length) are available for state representation. References [41,43,49,50] proposed the state as discrete values such as the position of vehicles or different levels of queue length. Nevertheless, this kind of state needs massive storage space for solving large problems. Therefore, recent studies recommend the use of continuous states such as queue length [42,45,51], average delay [44], and waiting time [30,42,48]. Furthermore, all of the papers relevant to this topic have used simulation platforms in order to obtain their desired results. SUMO [43,44,49,50], VISSIM [46,47], AIMSUN [45], and PARAMICS [30] are the most common software packages in which the combination of traffic simulation and RL can be executed appropriately. Finally, it is worth mentioning that all the previous papers took the signal controller as an agent for their RL algorithm except [47], which used connected vehicles as its agents.

This research aims to bridge the gap between the RL-based adaptive traffic signal control and the advanced traffic control (speed guidance model) under CAVs. The RL based approach focuses on the intersection as an agent in order to optimise total queue length while allowing the CAVs to adjust their speed based on a fixed timing plan to minimise the total stop delays as an agent adapts the traffic signal control to an adjusted speed of CAVs in real-time (updated traffic flow).

3. Methods

This section discusses the proposed framework for RL, PTV VISSIM microsimulation platform and implementation details. The RL framework is implemented in Python and integrated into VISSIM through the component object model (COM) interface (COM interface of PTV VISSIM).

3.1. Reinforcement Learning (RL)

RL is an area of adaptive control that encompasses algorithms for learning optimal behaviour policies in sequential decision-making problems from scalar rewards. It is assumed that a decision-making problem is time-discrete, and it can be represented with a Markov decision process (MDP) (S, A, P_a, R_a) , where S is the state space, A is the space

of possible actions, $P_a(s_t, s_{t+1}) = P(s_{t+1} | s_t, a_t)$ is the probability of transitioning into state s_{t+1} by taking action a_t in-state s_t , and $R_a(s_t, s_{t+1})$ is the immediate reward received after transitioning s_t into s_{t+1} by taking action a_t . The agent is specified by a policy $\pi(a | s)$ mapping each state to a probability distribution over actions. An optimal policy maximises discounted return $G_{t=0}$, which is defined as the sum of discounted future rewards (see (1)). The state-value function $V_\pi(s)$ is the expected return of a state s , assuming the agent follows policy π in all time steps (see (2)).

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots = R_{t+1} + \gamma G_{t+1} \quad (1)$$

$$V_\pi(s) = \mathbb{E}_{a \sim \pi} [P_a(s, s') R_a(s, s') + \gamma V_\pi(s')] \quad (2)$$

Model-based RL algorithms assume P_a and R_a to be known, while model-free RL algorithms learn about them implicitly from the interaction between agents and the MDP. This paper uses the advantage actor-critic algorithm (A2C), a model-free RL algorithm suitable for stochastic environments with partially observed states. The policy $\pi(a | s; \theta)$ and state-value function $V(s; \theta_v)$ are approximated by neural networks with parameters θ and θ_v , respectively. Initialised at random, the policy is optimised through iterative policy evaluation (gathering experience by interacting with the environment using the current policy) and policy improvement (reinforcing actions that led to greater than expected rewards and discouraging others) by gradient descent using (3) and (4).

$$d\theta = \nabla \log \pi(a | s; \theta) (G_t - V(s_t; \theta_v)) \quad (3)$$

$$d\theta_v = \partial / (\partial \theta_v) (G_t - V(s_t; \theta_v)) \quad (4)$$

In order to apply A2C to traffic signal control, we need to define the state space S and the action space A . Monte Carlo samples of $P_a(s, s')$ and $R_a(s, s')$ are generated by applying the agent policy in an episodic microscopic traffic simulation. We defined the action space of an agent that controls a single intersection as the set of possible signal phases. In this paper, all intersections have four approaches controlled in three phases. At every decision time step, the agent selects the next signal phase. An inter-stage is executed (switching the active signal group to red and the selected signal group to green) if required, and the chosen phase is applied for a minimum duration before the agent receives its immediate reward and the subsequent environment state. The state is represented as the concatenation of the following features:

- a. a vector encoding the current queue lengths on all incoming lanes,
- b. a one-hot vector encoding of the last chosen signal phase at time $t - 1$,
- c. the elapsed time since the last signal phase change,
- d. for each signal phase, the elapsed time since it was last active.

Elapsed time here is measured from the agent's perspective in the number of decision time steps. The queue lengths along all approaches represent a partial observation of current traffic demand, and all other state information summarises aspects of recent signalling history. The latter allows the agent to anticipate demand from elapsed time or learn a deterministic signal program (if that was an optimal solution). The instantaneous reward $R_a(s, s')$ is defined as the difference between the average queue length across all incoming lanes in state s and the average queue length in state s' . Thus, a positive instantaneous reward is given if the average queue length after an action is reduced and vice versa. We also explored using average vehicle delay as immediate rewards in preliminary experiments but discovered a simulation artifact that resulted in reward hacking. As the vehicle delay in PTV VISSIM is only counted after a vehicle passes the intersection, agents would converge to suboptimal policies that prevented vehicles from crossing the intersection to avoid the associated penalty. The policy network $\pi(s; \theta)$ consisted of three dense hidden layers with 64 units each and leaky ReLU activations ($\alpha = 0.05$), followed by a Softmax layer with one unit per signal phase. The value network $V(s; \theta_v)$ also consisted of three dense hidden

layers with 64 units each and leaky ReLU activations, followed by a single-unit linear layer. The state was normalised to $[0, 1]$ independently along each dimension before it was fed into these networks. During training, the agent gathered experience by interacting simultaneously with two environment instances, which has been shown to stabilise training by de-correlating batches of experience. We used the Adam optimiser with a learning rate 1×10^{-4} , gradient clipping at value 1.0, and entropy regularisation. The contributions of the policy loss, value loss and entropy loss were weighted with weights $w_\pi = 1.0$, $w_v = 0.5$, and $w_h = 0.01$, respectively. The agent was able to choose a signal phase once every 5 s of green time (ignoring the time spent changing signal phases). With this configuration, all agents were trained for 100 simulation episodes, each simulating two hours of traffic.

3.2. Simulation Platform

In this paper, PTV VISSIM microsimulation software was used to investigate the impact of the proposed framework on performance measures such as queue length and stop delay. Therefore, akin to some relevant studies [47,52], an isolated four-leg signalised intersection was employed in this study (north entry has one line, whereas other approaches have two lines). This intersection has been taken from a sample example in VISSIM (used as a template to show the benefits of three-stage vehicle actuated signal control over fixed time) to represent the functionality of the proposed method in dynamic traffic demand. As shown in Figure 2, time-varying arrival rates were generated based on the VISSIM example at 15-min intervals for two hours to consider the fluctuation of traffic demand in a real-world network.

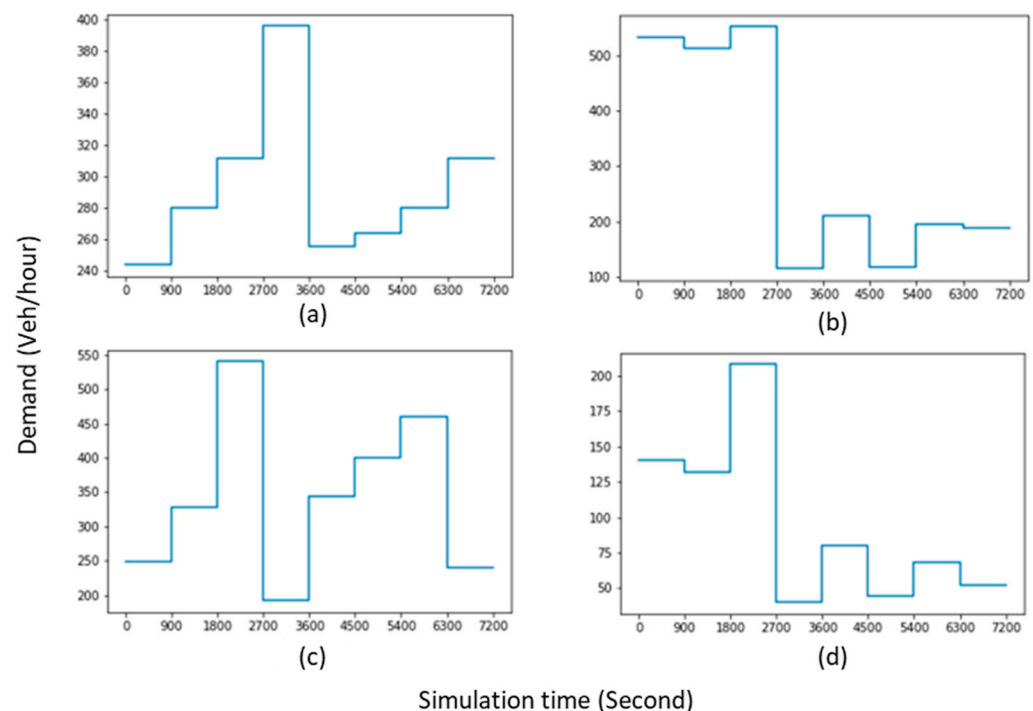


Figure 2. Time-varying arrival rates at 15-min interval. (a) East entry; (b) West entry; (c) South entry; (d) North entry.

3.3. Driving Behaviours

In this study, two vehicle classes were defined for the simulation:

- (1) Conventional vehicles: This type of vehicle has typical characteristics of a human-driven car. The default VISSIM car-following model (Wiedemann 74) was used. Furthermore, the uniform distribution with a minimum value of 45 km per hour and the maximum value of 55 km per hour was utilised to generate the speed of conventional vehicles.

- (2) Connected and automated vehicles (CAVs): driving behaviour for this vehicle class consists of two major components, autonomous behaviour and connected behaviour, which will be explained below.

3.3.1. Autonomous Behaviour

Numerous studies have investigated the characteristics of autonomous vehicles (AVs) through various experiments [53–55]. AVs have some common features which should be considered for the simulation. For instance, AVs can accept a smaller headway than conventional cars [53,55]. They can also keep their speed constant without any fluctuation during free-flow [56], so the constant speed of 50 km per hour was considered for CAVs in this research, which is the mean speed of conventional vehicles. Furthermore, autonomous vehicles accelerate more smoothly than conventional cars. Unlike conventional cars, every particular speed has a unique acceleration and deceleration rate for this type of vehicle, whose acceleration ranges between the minimum and maximum values [57]. Figure 3 better depicts the last cases.

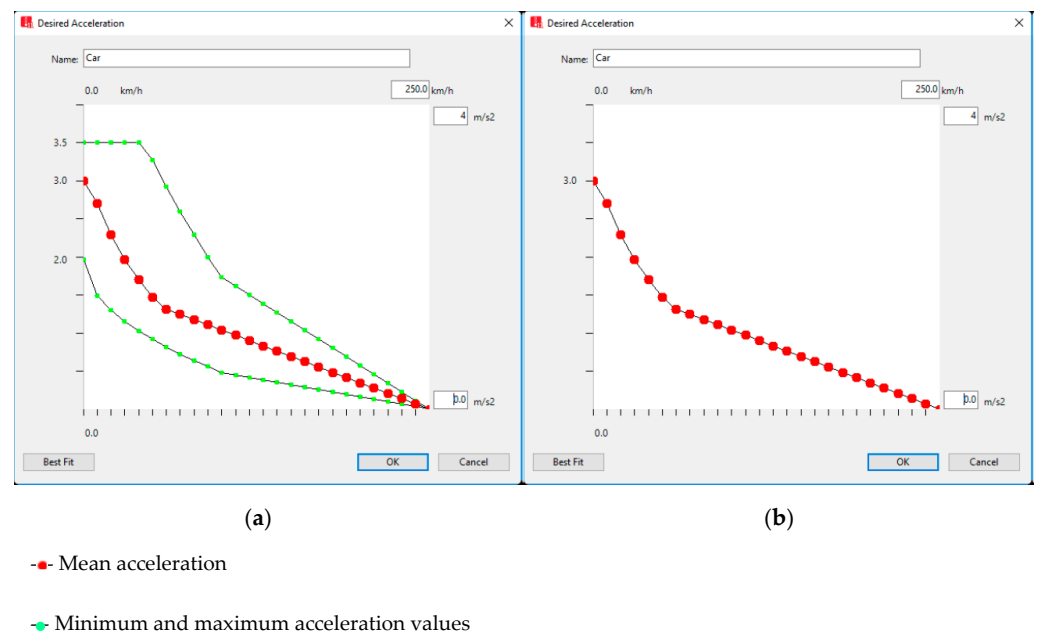


Figure 3. Desired acceleration functions of a conventional vehicle and an autonomous vehicle in VISSIM [58]. (a) Conventional vehicle; (b) Autonomous vehicle.

In this study, the default AV behaviour of VISSIM, which is based on a European project called Coexist [59], was utilised. Coexist has tested its cars with four driving patterns (Rail Safe, Cautious, Normal, and All-Knowing). The main difference between these types of autonomous vehicles is their capability to accept headways. Cautious AVs are more conservative than conventional cars. Thus, they keep larger headways than other AV types. Cautious AVs were selected for simulating the autonomous behaviour of CAVs in this paper because this vehicle class may be the first generation of highly automated vehicles and can penetrate in transportation networks sooner than other AV types.

3.3.2. Connected Behaviour

CAVs continuously pay attention to the data transmitted to them from other vehicles (V2V) and traffic infrastructures (V2I). In particular, CAVs receive information about the upcoming signal and modify their speed to arrive at the green phase without stopping in signalised intersections. Therefore, it is essential to use the internal script which reads this information from PTV VISSIM [58] and updates the desired speed of CAVs to arrive within green at the signal. This speed guidance is a rule-based algorithm [34] that developed mainly includes the following steps.

- Step 1. The first question that should be asked of all vehicles entering the network is if the car is able to receive signal data or not. Therefore, conventional vehicles will proceed with movement at their desired speed (the speed at which the driver wants to drive). However, if the vehicle is connected, Step 2 is executed.
- Step 2. The vehicle will continue with its current speed if it passes the intersection or no signal controller can be found ahead of this vehicle; otherwise, Step 3 must be performed.
- Step 3. In this step, the following question should be answered. "Is the signal at its green phase?". If the response is negative and the signal controller is at its red phase, the vehicle speed must be adjusted (5). Otherwise, go to Step 4.

$$V_{opt} = \max(\min(V_{max \text{ for green start}}, V_{des}) - V_{diff}, V_{min}) \quad (5)$$

In this equation, V_{opt} is the optimal speed of the vehicle, and V_{des} is the desired speed of the vehicle. Moreover, the functionality of V_{diff} is to adjust the vehicle speed so that the vehicle arrives shortly before the signal head. It was assumed to be 2 km per hour in the simulation. V_{min} is the least feasible speed of the vehicles in the network, which was considered 5 km per hour in this paper. Finally, the vehicle should not drive above the $V_{max \text{ for a green start}}$ in order to arrive just when the next green starts. This speed can be obtained from (6).

$$V_{max \text{ for green start}} = \frac{\text{Vehicle distance to signal head}}{\text{Time until the next green phase starts}} \quad (6)$$

- Step 4. If $V_{min \text{ for a green end}}$ (a minimum speed required to arrive at the intersection during the current green) is lower than the desired speed of the vehicle, then V_{opt} should be equal to V_{des} . Conversely, Step 5 is executed. $V_{min \text{ for a green end}}$ can be calculated by (7).

$$V_{min \text{ for green end}} = \frac{\text{Vehicle distance to signal head}}{\text{Time until the next red phase starts}} \quad (7)$$

- Step 5. If $V_{max \text{ for a green start}}$ is greater than the desired speed of the vehicle, then V_{opt} should be equal to V_{des} . Otherwise, $V_{opt} = V_{max \text{ for a green start}}$. Therefore, the optimal speed of all CAVs in the network can be calculated through this procedure.

The proposed behaviours were entered in VISSIM for both conventional vehicles and CAVs. Now, the designated scenarios for this study should be explained.

3.4. Simulation Scenarios

It is important to perform sensitivity analysis on various parameters to achieve more accurate and comprehensive findings. Therefore, several simulation scenarios were defined in this paper. Different market penetration rates of 0%, 25%, 50%, 75%, and 100% for CAVs were the first set of scenarios considered for this study. Moreover, three scenarios for the total demand were designated. The first scenario was called saturated, which has the travel demands equal to the capacity for each signal phase. This capacity can be calculated by (8). The second and third scenarios were called oversaturated and unsaturated, with travel demands equal to 1.1 and 0.7 of saturated conditions, respectively.

$$\text{Capacity} = s \times \frac{g}{C} \times N \quad (8)$$

where s is the saturation flow rate was considered to be 1900 veh/h for each lane based on the Highway Capacity Manual [60]. Furthermore, g is the effective green time duration for the phase in seconds, C is the total cycle length in seconds, and N is the lane number for each lane group.

Finally, different signal plans should be compared to measure the relative improvements of our approach compared to other widely used methods. First, the intersection with fixed signal timing alongside the AVs was investigated. This scenario assumes that vehicles cannot receive any information from the signal controller (AV scenarios do not include speed guidance approaches). This situation was called “Fixed AV”. Furthermore, the fixed signal timing was considered for the intersection again, but AVs, in this case, have the ability to receive signal data, and they can adjust their speed based on the speed guidance approach (i.e., CAVs were used for this condition instead of simple AVs). This situation was called “Fixed CAV”. The actuated signal plan was examined with different AV penetration rate scenarios (Not CAV). This case was named “Actuated”.

Thus, 60 scenarios (five different vehicle compositions \times three different demands \times four controlling conditions: Fixed AV, Fixed CAV, Actuated (AV), and RL (CAV)) were generated. Each scenario was typically executed in VISSIM for up to five various random seeds. This process was also accomplished by the RL, and the expected improvements in performance measures were assessed under the CAV environment.

4. Results

Two different performance measures are implemented: minimising total queue length and total stop delays. The main purpose of this paper is to develop a framework to minimise the performance measure. The minimisation of queue length is the first part of the performance index (PI). Queues happen when vehicles cannot pass the intersection at their green share, and they must stop behind the red light. The average queue length of all entries in the whole simulation period is assessed to optimise an RL feedback loop as a reward factor. The second part of PI is the stop delay of vehicles, a common performance measure in an optimal speed guidance system under a connected environment [28]. This can be described as a delay (in seconds), the mean stop time of vehicles waiting at the intersection behind red lights. In order to use the two aforementioned performance measures simultaneously in one expression, the transformation strategy has been used to convert each measure into a non-dimensional term. According to [61], the average queue length (q) can be divided by its maximum value (Q_{max}).

Moreover, stop delay (d) can be investigated per cycle length, so it should be divided by signal cycle time (C) discussed in the previous section. Therefore, (9) to (11) illustrate how to determine the PI value for each scenario. In these equations, n represents different replications for the simulation.

$$\text{Queue Length ratio} = \frac{1}{n} \sum_{i=1}^n \frac{q_i}{Q_{max_i}} \quad (9)$$

$$\text{Stop Delay ratio} = \frac{1}{n} \sum_{i=1}^n \frac{d_i}{C} \quad (10)$$

$$\text{PI} = \text{Queue Length ratio} + \text{Stop Delay ratio} \quad (11)$$

In the first step, it is worth comparing PI values of various signal plans for both saturated and oversaturated conditions. The results for the average PI value are shown in Figures 4 and 5.

As can be observed, the speed guidance approach has worked appropriately due to the fact that PI values for the fixed CAV condition are less than their corresponding values for the fixed AV condition in each scenario. In other words, a 100% penetration rate of CAVs compared to a conventional environment (0% penetration rate of CAVs) in the saturated scenario can decrease PI by 18% and 43% for fixed AV and fixed CAV conditions, respectively. Moreover, the fixed CAV situation has performed better than the actuated case. Therefore, the fixed CAV condition is the best case among all non-learning approaches, and the RL should be compared with this condition in order to assess its functionality.

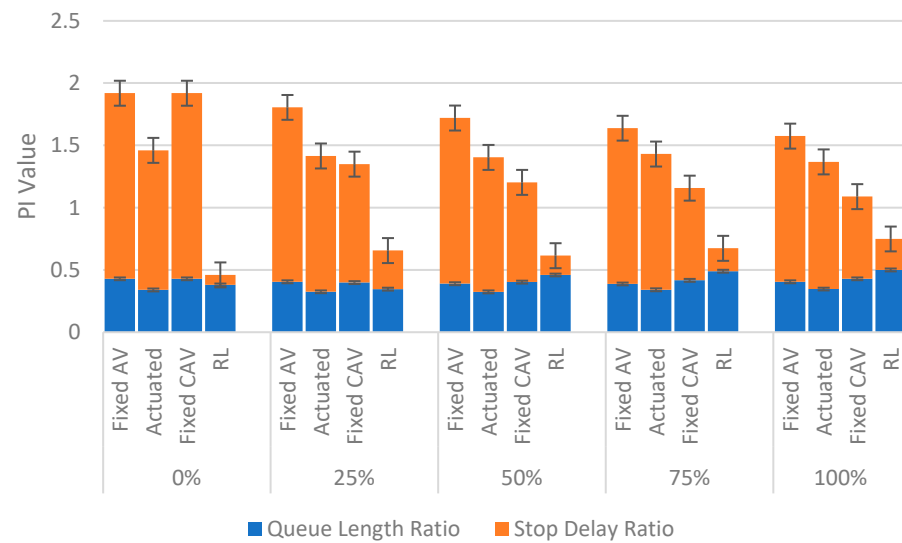


Figure 4. Comparing performance index values of the saturated scenarios.

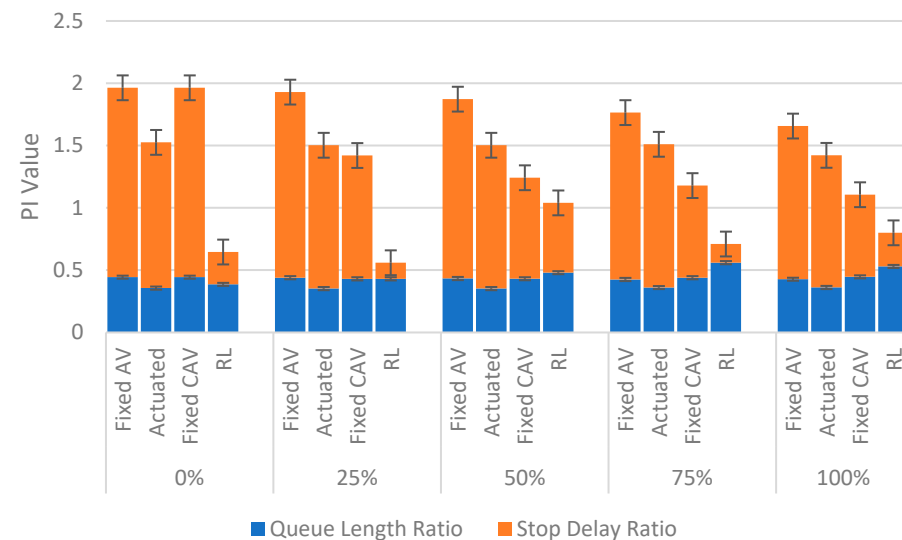


Figure 5. Comparing performance index values of the oversaturated scenarios.

As shown, the RL can noticeably reduce the PI compared to the fixed CAV condition, especially in 0% and 25% scenarios. However, a slight improvement has occurred in the last three vehicle composition scenarios (i.e., 50%, 75% and 100%), particularly for the oversaturated scenario. This is mainly because of the assumption that CAVs in this paper only can obtain the signal data. This assumption makes the simulation more straightforward and more rapid. Results show that the total network performance could be improved under this simple CAVs environment. Furthermore, future works should consider the CAVs that can receive and interpret the information transmitted from other CAVs (V2V). In other words, the driving behaviour implemented in this study for CAVs only can model V2I, and it cannot cover the V2V connection whose impact has been pinpointed from the 50% penetration rate for CAVs.

Another point to note is that in most cases, the queue length ratio increases when the CAVs penetration rate also increases for the RL. For instance, the maximum value of the queue length ratio for the RL in the saturated scenario has a growth rate of 43% compared to its minimum value in the 25% scenario. Although it seems far from the expectation at first glance, it may stem from the fact that the speed guidance approach defined for this research does not consider vehicle positions, and CAVs receive the signal data as soon as they enter the network. Moreover, this algorithm does not specify different ranges for

the minimum speed of CAVs, and it has been assumed to be 5 km per hour for this study. Therefore, such a wide range for receiving the signal information in conjunction with this tiny speed value can cause long queues behind CAVs.

PI values were analysed for both saturated and oversaturated scenarios. One of the unforeseen outcomes which can be highlighted in Figure 4 is that the PI value suddenly escalates when moving from 25% to 50% for the RL. In order to better perceive this trend, we present PI variation in the unsaturated condition for these two aforementioned CAV scenarios. The final results are displayed in Figure 6.

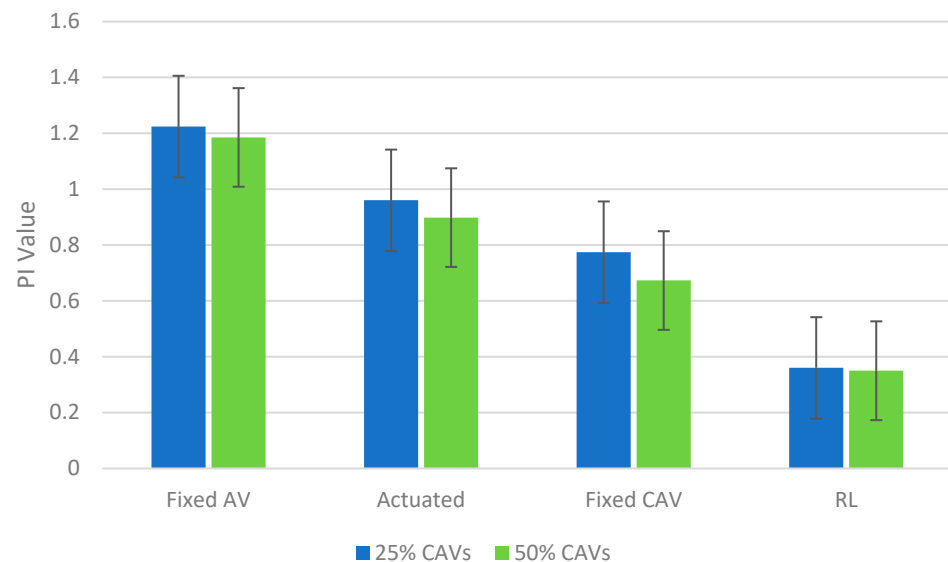


Figure 6. Comparing performance index values of the unsaturated scenario for all signal planning approaches.

As expected, fixed CAV, actuated, and RL approaches can improve the conditions of this intersection compared to the simple fixed AV situation. The average improvements are 23%, 40%, and 70%, respectively. Therefore, RL has the best performance in unsaturated conditions among all other algorithms. The main reason for this is that despite using optimal fixed time for CAVs speed guidance, the proposed real-time RL could optimise the signal plan based on the time-varying traffic demand. This means that more vehicles can pass the intersection in their green time, and each entry will be vacant at its green share. That is why the queue length and stop delay will be minimised, and the intersection will perform at the ideal condition, which is unrivalled.

Furthermore, all approaches show a decreasing PI rate by transitioning from 25% CAVs to 50% CAVs. Hence, it is concluded that, unlike saturated and oversaturated conditions, vehicle automation can play an apt role in unsaturated traffic conditions.

The results of this analysis also highlighted the fact that the PI value for a proposed RL framework with speed guidance under a CAV environment is much less than other control systems at all penetration rates of CAVs (AVs) and demand scenarios. The proposed framework is a novel way of framing the RL state based on a CAV environment with speed guidance that could work more flexibly than RL under a conventional environment or speed guidance system under CAVs with fixed timing signal control.

5. Conclusions

This study develops a learning-based framework that optimises the signal plan by training a traffic signal control as an agent in RL to minimise total queue length (reward in RL) when CAVs receive speed guidance in a fixed-time strategy to minimise the total stop delays. The signal controller (agent) was trained in the fully dynamic traffic environment (traffic flows and CAV speeds) under different demand levels (unsaturated, saturated,

oversaturated) as well as CAV penetration rates (0%, 25%, 50%, 75%, 100%) scenarios to show potential interaction effects between signal timing plan, CAV penetration rate and traffic flow.

An isolated four-leg signalised intersection with three phases (training example in VISSIM as a template of vehicle actuated signal control) was modelled to test the performance. The results were compared to a well-tuned fixed timing plan (with optimal speed advisory in a CAV environment) and actuated signal control. Two objective functions are implemented in a performance index (queue length and stop delay). An important finding from this study is that the proposed framework reduced stop delay significantly under all scenarios and was comparable to other control strategies in queue length.

Finally, the proposed framework only considers simple V2I communication in an isolated intersection. The framework is designed to handle urban networks, and the interaction in operations between adjacent intersections (for example, interchanges) under the V2X environment and evaluation for these conditions is planned. The main limitation of a network-wide application is the computational time for the RL training. In future works, offset optimisation could be added to the signal timing optimisation to implement the proposed approach in an arterial corridor with multiple traffic signals or a city traffic network. It can improve the overall performance of the proposed approach but increase the training process complexity for the RL approaches. Future research will also make use of V2V communication and dynamic speed guidance strategies. Lastly, with dynamic speed guidance strategies which have different communication and control levels compared to simple V2I, it might be helpful to analyse reward and state options of RL with different performance measures.

Author Contributions: Conceptualisation, S.M., S.S., J.H. and R.M.-S.; methodology, S.M. and S.S.; software, S.M. and S.S.; validation, S.M.; formal analysis, S.M., S.S. and J.H.; investigation, S.M. and S.S.; resources, S.M. and S.S.; data curation, S.M. and S.S.; writing—original draft preparation, S.M. and S.S.; writing—review and editing, S.M., S.S., J.H. and R.M.-S.; visualisation, S.M.; project administration, R.M.-S.; funding acquisition, R.M.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by EPSRC grant UK EP/R018634/1 entitled Closed-Loop Data Science for Complex, Computationally- and Data-Intensive Analytics.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are not publicly available due to the project requirements.

Acknowledgments: The authors would like to express their deep gratitude to Mohammad Khashayarfar for his inspirational guidance and professional assistance.

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

References

1. Institute of Economic Affairs (IEA) Discussion Paper, No.68, 2016. SEEING RED Traffic Controls and the Economy. Available online: <https://iea.org.uk/wpcontent/uploads/2016/07/IEA%20Seeing%20Red%20%20Traffic%20Controls%20and%20the%20Economy.pdf> (accessed on 11 January 2021).
2. Gartner, N.H. OPAC: A demand-responsive strategy for traffic signal control. *Transp. Res. Rec. J. Transp. Res. Board* **1983**, *906*, 75–81.
3. Head, K.L.; Mirchandani, P.B.; Sheppard, D. Hierarchical framework for real-time traffic control. *Transp. Res. Rec.* **1992**, *1360*, 82–88.
4. Luyanda, F.; Gettman, D.; Head, L.; Shelby, S.; Bullock, D.; Mirchandani, P. ACS-Lite algorithmic architecture: Applying adaptive control system technology to closed-loop traffic signal control systems. *Transp. Res. Record* **2003**, *1856*, 175–184. [[CrossRef](#)]
5. Mirchandani, P.; Head, L. A real-time traffic signal control system: Architecture, algorithms, and analysis. *Transp. Res. Part C Emerg. Technol.* **2001**, *9*, 415–432. [[CrossRef](#)]
6. Ran, Q.; Yang, J. A novel closed-loop feedback traffic signal control strategy at an isolated intersection. In Proceedings of the 2012 IEEE International Conference on Information Science and Technology, ICIST, Wuhan, China, 23–25 March 2012. [[CrossRef](#)]

7. Fagnant, D.J.; Kockelman, K. Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transp. Res. Part A Policy Pract.* **2015**, *77*, 167–181. [CrossRef]
8. Kockelman, K.; Boyles, S. Smart Transport for Cities & Nations: The Rise of Self-Driving & Connected Vehicles. The University of Texas at Austin, 2018. Available online: <https://bit.ly/3gDpEaa> (accessed on 15 March 2021).
9. Elliott, D.; Keen, W.; Miao, L. Recent advances in connected and automated vehicles. *J. Traffic Transp. Eng.* **2019**, *6*, 109–131. [CrossRef]
10. Feng, Y.; Head, K.L.; Khoshmashgham, S.; Zamanipour, M. A real-time adaptive signal control in a connected vehicle environment. *Transp. Res. Part C Emerg. Technol.* **2015**, *55*, 460–473. [CrossRef]
11. Yao, H.; Jin, Y.; Jiang, H.; Hu, L.; Jiang, Y. CTM-based traffic signal optimisation of mixed traffic flow with connected automated vehicles and human-driven vehicles. *Phys. A: Stat. Mech. Its Appl.* **2022**, *603*, 127708. [CrossRef]
12. Yao, Z.; Shen, L.; Liu, R.; Jiang, Y.; Yang, X. A Dynamic Predictive Traffic Signal Control Framework in a Cross-Sectional Vehicle Infrastructure Integration Environment. *Proc. IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 1455–1466. [CrossRef]
13. Saldivar-Carranza, E.; Li, H.; Mathew, J.; Fisher, C.; Bullock, D. Signalized Corridor Timing Plan Change Assessment Using Connected Vehicle Data. *J. Transp. Technol.* **2022**, *12*, 310–322. [CrossRef]
14. Yao, Z.; Jiang, Y.; Zhao, B.; Luo, X.; Peng, B. A dynamic optimization method for adaptive signal control in a connected vehicle environment, *Journal of Intelligent Transportation. Systems* **2020**, *24*, 184–200. [CrossRef]
15. Guo, Q.; Li, L.; Ban, X. Urban traffic signal control with connected and automated vehicles: A survey. *Transp. Res. Part C Emerg. Technol.* **2019**, *101*, 313–334. [CrossRef]
16. Mintsis, E.; Vlahogianni, E.I.; Mitsakis, E.; Ozkul, S. Enhanced speed advice for connected vehicles in the proximity of signalized intersections. *Eur. Transp. Res. Rev.* **2021**, *13*, 2. [CrossRef]
17. Jiang, Y.; Zhao, B.; Liu, M.; Yao, Z. A Two-Level Model for Traffic Signal Timing and Trajectories Planning of Multiple CAVs in a Random Environment. *J. Adv. Transp.* **2021**, *2021*, 9945398. [CrossRef]
18. He, X.; Liu, H.X.; Liu, X. Optimal vehicle speed trajectory on a signalised arterial with consideration of queue. *Transp. Res. Part C Emerg. Technol.* **2015**, *61*, 106–120. [CrossRef]
19. Wang, M.; Daamen, W.; Hoogendoorn, S.P.; van Arem, B. Rolling horizon control framework for driver assistance systems. Part I: Mathematical formulation and non-cooperative systems. *Transp. Res. Part C Emerg. Technol.* **2014**, *40*, 271–289. [CrossRef]
20. Wu, X.; He, X.; Yu, G.; Harmandayan, A.; Wang, Y. Energy-optimal speed control for electric vehicles on signalised arterials. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2786–2796. [CrossRef]
21. Wang, P.; Chan, C.Y.; de La, F.A. A reinforcement learning-based approach for automated lane change maneuvers. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu/Suzhou, China, 26–30 June 2018; pp. 1379–1384.
22. Abdulhai, B.; Kattan, L. Reinforcement learning: Introduction to theory and potential for transport applications. *Can. J. Civ. Eng.* **2003**, *30*, 981–991. [CrossRef]
23. Shatnawi, I.; Yi, P.; Khliefat, I. Automated intersection delay estimation using the input-output principle and turning movement data. *Int. J. Transp. Sci. Technol.* **2018**, *7*, 137–150. [CrossRef]
24. Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software. US Department of Transportation, 2010. Available online: https://ops.fhwa.dot.gov/trafficanalysistools/tat_vol3/vol3_guidelines.pdf (accessed on 16 March 2021).
25. Hunt, P.B.; Robertson, D.I.; Bretherton, R.D. The SCOOT online traffic signal optimisation technique (Glasgow). *Traffic Eng. Control.* **1982**, *23*, 190–192.
26. Sims, A.G.; Dobinson, K.W. The Sydney Coordinated Adaptive Traffic (SCAT) System Philosophy and Benefits. *IEEE Trans. Veh. Technol.* **1980**, *29*, 130–137. [CrossRef]
27. Athmaraman, N.; Soundararajan, S. Adaptive predictive traffic timer control algorithm. In Proceedings of the 2005 Mid-Continent Transportation Research Symposium, Ames, IA, USA, 18–19 August 2005. Available online: <https://www.cs.uic.edu/~nathmaraman/AthmaramanTraffic.pdf> (accessed on 1 March 2021).
28. Wang, Y.; Yang, X.; Liang, H.; Liu, Y. A Review of the Self-Adaptive Traffic Signal Control System Based on Future Traffic Environment. *J. Adv. Transp.* **2018**, *2018*, 1096123. [CrossRef]
29. Ke-Zhao, B.; Rui-Xiong, C.; Mu-ren, L.; Ling-Jiang, K.; Rong-sen, Z. Study of the grade roundabout crossing. *Comput. Sci.* **2009**. [CrossRef]
30. El-tantawy, S.; Member, S.; Abdulhai, B.; Abdelgawad, H. Multiagent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers Application on Downtown Toronto. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1140–1150. [CrossRef]
31. Jing, P.; Huang, H.; Chen, L. An Adaptive Traffic Signal Control in a Connected Vehicle Environment: A Systematic Review. *Information* **2017**, *8*, 101. [CrossRef]
32. Wu, W.; Huang, L.; Du, R. Simultaneous Optimization of Vehicle Arrival Time and Signal Timings within a Connected Vehicle Environment. *Sensors* **2020**, *20*, 191. [CrossRef] [PubMed]
33. Katsaros, K.; Kernchen, R.; Dianati, M.; Rieck, D. Performance study of a Green Light Optimized Speed Advisory (GLOSA) application using an integrated cooperative ITS simulation platform. In Proceedings of the IWCMC 2011 7th International Wireless Communications and Mobile Computing Conference, Istanbul, Turkey, 4–8 July 2011. [CrossRef]
34. Tang, T.Q.; Yi, Z.Y.; Zhang, J.; Wang, T.; Leng, J.Q. A speed guidance strategy for multiple signalised intersections based on car-following model. *Phys. A: Stat. Mech. Its Appl.* **2018**, *496*, 399–409. [CrossRef]

35. Beak, B.; Larry Head, K.; Feng, Y. Adaptive coordination based on connected vehicle technology. *Transp. Res. Rec.* **2017**, *2619*, 1–12. [[CrossRef](#)]
36. Pandit, K.; Ghosal, D.; Zhang, H.M.; Chuah, C.N. Adaptive traffic signal control with vehicular Ad Hoc networks. *IEEE Trans. Veh. Technol.* **2013**, *62*, 1459–1471. [[CrossRef](#)]
37. Yu, C.; Feng, Y.; Liu, H.X.; Ma, W.; Yang, X. Integrated optimisation of traffic signals and vehicle trajectories at isolated urban intersections. *Transp. Res. Part B Methodol.* **2018**, *112*, 89–112. [[CrossRef](#)]
38. Lämmer, S.; Helbing, D. Self-control of traffic lights and vehicle flows in urban road networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P04019. [[CrossRef](#)]
39. Carlson, R.C.; Papamichail, I.; Papageorgiou, M. Local feedback-based mainstream traffic flow control on motorways using variable speed limits. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 1261–1276. [[CrossRef](#)]
40. Diakaki, C.; Papageorgiou, M.; Aboudolas, K. A multivariable regulator approach to traffic-responsive network—Wide signal control. *Control Eng. Pract.* **2002**, *10*, 183–195. [[CrossRef](#)]
41. Chin, Y.K.; Lee, L.K.; Yang, S.S.; Tze, K.; Teo, K. Exploring Q-Learning Optimisation in Traffic Signal Timing Plan Management. In Proceedings of the 2011 Third International Conference on Computational Intelligence, Communication Systems and Networks, Bali, Indonesia, 26–28 July 2011; pp. 269–274. [[CrossRef](#)]
42. Chu, T.; Wang, J.; Codecà, L.; Li, Z. Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 1086–1095. [[CrossRef](#)]
43. Kim, D.; Jeong, O. Cooperative Traffic Signal Control with Traffic Flow Prediction in Multi-Intersection. *Sensors* **2019**, *20*, 137. [[CrossRef](#)]
44. Yang, K.; Tan, I.; Menendez, M. A reinforcement learning-based traffic signal control algorithm in a connected vehicle environment. In Proceedings of the 17th Swiss Transport Research Conference (STRC 2017), Ascona, Switzerland, 17–19 May 2017.
45. Aslani, M.; Seipel, S.; Saadi, M.; Wiering, M. Advanced Engineering Informatics Traffic signal optimisation through discrete and continuous reinforcement learning with robustness analysis in downtown Tehran. *Adv. Eng. Inform.* **2018**, *38*, 639–655. [[CrossRef](#)]
46. Prabuchandran, K.J.; Hemanth Kumar, A.N.; Bhatnagar, S. Decentralised Learning for Traffic Signal Control. In Proceedings of the Intelligent Transportation System Workshop, COMSNETS, Bangalore, India, 6–10 January 2015. [[CrossRef](#)]
47. Chen, J.; Xue, Z.; Fan, D. Deep Reinforcement Learning Based Left-Turn Connected and Automated Vehicle Control at Signalized Intersection in Vehicle-to-Infrastructure Environment. *Information* **2020**, *11*, 77. [[CrossRef](#)]
48. Liu, W.; Qin, G.; He, Y.; Jiang, F. Distributed Cooperative Reinforcement Learning-Based Traffic Signal Control That Integrates V2X Networks' Dynamic Clustering. *Proc. IEEE Trans. Veh. Technol.* **2017**, *66*, 8667–8681. [[CrossRef](#)]
49. Mousavi, S.S.; Schukat, M.; Howley, E. Traffic light control using deep policy- gradient and value-function-based reinforcement learning. *IET Intell. Transp. Syst.* **2017**, *11*, 417–423. [[CrossRef](#)]
50. van der Pol, E.; Oliehoek, F.A. Coordinated deep reinforcement learners for traffic light control. In Proceedings of the 30th Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016.
51. Nishi, T.; Otaki, K.; Hayakawa, K.; Yoshimura, T. Traffic Signal Control Based on Reinforcement Learning with Graph Convolutional Neural Nets. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 877–883. [[CrossRef](#)]
52. Zhao, J.; Li, W.; Wang, J.; Ban, X. Dynamic traffic signal timing optimisation strategy incorporating various vehicle fuel consumption characteristics. *IEEE Trans. Veh. Technol.* **2015**, *65*, 3874–3887. [[CrossRef](#)]
53. Aria, E.; Olstam, J.; Schwietering, C. Investigation of Automated Vehicle Effects on Driver's Behavior and Traffic Performance. *Transp. Res. Procedia* **2016**, *15*, 761–770. [[CrossRef](#)]
54. ATKINS. Research on the Impacts of Connected and Autonomous Vehicles (CAVs) on Traffic Flow. 2016.
55. Stanek, D.; Huang, E.; Milam, R.T. Measuring Autonomous Vehicle Impacts on Congested Networks Using Simulation. In Proceedings of the Transportation Research Board 97th Annual Meeting, Washington, DC, USA, 7–11 January 2018.
56. Asadi, F.E.; Anwar, A.K.; Miles, J.C. Investigating the potential transportation impacts of connected and autonomous vehicles. In Proceedings of the 2019 8th IEEE International Conference on Connected Vehicles and Expo, ICCVE, Graz, Austria, 4–8 November 2019. [[CrossRef](#)]
57. Zeidler, V.; Buck, S.; Kautzsch, L.; Vortisch, P.; Weyland, C. Simulation of Autonomous Vehicles Based on Wiedemann's Car Following Model in PTV Vissim. In Proceedings of the Transportation Research Board 98th Annual Meeting, Washington, DC, USA, 13–17 January 2019.
58. PTV Group. PTV Vissim 10 User Manual. Ptv Ag. Available online: <https://usermanual.wiki/Document/Vissim20102020Manual.1098038624.pdf> (accessed on 1 March 2021).
59. CoEXist. Working towards a Shared Road Network. 2020. Available online: <https://www.h2020-coexist.eu/> (accessed on 12 January 2020).
60. Council, N.R. *TRB. Highway Capacity Manual*; National Research Council: Washington, DC, USA, 2000.
61. Marler, R.T.; Arora, J.S. Function-transformation methods for multi-objective optimisation. *Eng. Optim.* **2005**, *37*, 551–570. [[CrossRef](#)]