Evaluating the Impact of Data Granularity on Deep Q-Network based Smart Traffic Signal Control

Michel Sauvage*†, Clément Leroy*, Pierrick Cochon Jaffres*, Jérôme Härri†
*Direction de l'Innovation, Alten, Rennes, France
†EURECOM, Sophia Antipolis, France
Emails: {michel.sauvage, clement.leroy, pierrick.cochonjaffres}@alten.com, jerome.haerri@eurecom.fr

***Abstract*—Deep Q-Networks (DQN) are a promising technology for AI-driven traffic signal control (TSC), but their training requires complex input data. Modeling can be conducted at either microscopic or macroscopic levels. While microscopic modeling captures detailed traffic dynamics, it requires extensive parameter calibration. In contrast, macroscopic modeling offers faster setup and reduced computational cost with less precision. To evaluate the trade-offs, the study compare models trained on both data types under two DQN configurations: one with fixed decision intervals, and another allowing decisions every second with enforced pauses after phase changes. All traffic data used in this study is synthetically generated using the SUMO traffic simulator, ensuring full control over experimental conditions and flow scenarios. Results show that macroscopic data enables faster convergence and comparable, if not better, performance. Although the microscopic model offers finer control, it suffers from instability when combined with coarse decision intervals. These findings highlight that high-fidelity data is not strictly necessary to train effective traffic signal control policies, which is particularly advantageous for large-scale urban simulations and city-scale digital twins.**

***Index Terms*—Traffic Signal Control, Reinforcement Learning, Traffic Flow Modeling, SUMO (Simulation of Urban MObility), Intelligent Transportation Systems**

I. Introduction

Urban congestion represents a major contemporary challenge, leading to significant delays for commuters and substantial environmental impact due to increased greenhouse gas emissions [1]. To address this growing issue, the development of intelligent traffic light systems based on Artificial Intelligence (AI) has emerged as a promising approach to improve traffic flow and reduce emissions [2].

However, designing and optimizing such systems is based on robust simulation tools capable of accurately modeling urban traffic dynamics. Despite a wide variety of traffic simulation models, there is currently no consensus on the most appropriate level of detail to use in AI-based traffic signal training [3]. Simulations vary in granularity, from microscopic models, which simulate individual vehicles [4], to macroscopic models, which treat traffic as aggregated flows [5]. Microscopic simulations offer high fidelity but require substantial calibration and computing resources. Macroscopic models, while coarser, are easier to configure and faster to run.

In recent years, reinforcement learning (RL) has shown strong potential for adaptive traffic light control, allowing agents to learn optimal policies by interacting with simulated environments [6]. RL performance critically depends on the quality and structure of simulation data used during training. Yet, little attention has been paid to the influence of simulation granularity, microscopic versus macroscopic, on the learning process and final performance of RL agents.

In practice, AI-based TSC systems commonly adopt one of two agent operation modes: fixed decision intervals or high-frequency decision-making. These modes reflect real-world constraints and design preferences in control systems. However, their interaction with the level of data granularity used during training has not been systematically explored.

This study addresses a key open question in the field: is high-fidelity microscopic data necessary to effectively train reinforcement learning agents for traffic light control? We investigate how data granularity and agent operation modes, such as per-second actions versus fixed intervals, jointly affect the performance of DQN, focusing on their impact on convergence speed and policy robustness.

Our objective is to clarify the trade-offs between data granularity and learning efficiency in RL-based traffic signal control. By comparing observation granularities and agent decision frequencies, this study seeks to identify configurations that provide strong performance while reducing simulation and training complexity. These findings contribute to ongoing efforts toward scalable and efficient RL training strategies, particularly relevant for future applications in large-scale networks and digital twin platforms [7].

The rest of this paper is organized as follows: Section II reviews granularity in traffic signal control modeling. Section III describes the methodology followed in this paper. Section IV provides performance comparisons. Section V discuss the impact of this study and presents future work.

II. State of the Art

A. Modeling Granularity

Traffic simulations can be categorized according to their level of granularity, with two main modeling scales commonly distinguished in the literature: macroscopic and microscopic [8].

1) Macroscopic Models: Macroscopic models describe traffic flow at an aggregate level, without representing individual vehicles. These models treat traffic similarly to fluid dynamics, using variables such as traffic density, average speed, and flow rate to characterize vehicle movement [5]. They are particularly efficient in terms of computational resources and are well-suited for large-scale network analysis. One of the most well-known macroscopic frameworks is the Cell Transmission Model (CTM) proposed by Daganzo [9], which discretizes both space and time. The model is derived from the Lighthill-Whitham-Richards (LWR) fluid-dynamic equations and applies them in a cellular structure. Each cell has a maximum vehicle capacity and a fundamental diagram governing the relationship between flow and density. Traffic propagation is modeled based on the supply and demand of adjacent cells.

2) Microscopic Models: Microscopic models provide a detailed representation of individual vehicle behavior. Each vehicle is explicitly modeled with its own position, speed, acceleration, and route. These models are highly accurate and suitable for evaluating local control strategies [4]. A common modeling approach in this category is based on car-following models, where the acceleration of a vehicle depends on its relative position and speed with respect to the leading vehicle. Popular examples include the Intelligent Driver Model (IDM) [10] and the Gipps model [11], which incorporate reaction times, desired speeds, and vehicle lengths. Another class of microscopic models focuses on lane-changing behavior, which can be motivated by either route requirements (e.g., needing to turn) or by the desire to maintain a higher speed.

Understanding the underlying traffic simulation models is essential, as the choice of modeling scale directly influences the type and precision of data available for training control strategies. Building upon this foundation, the next section focuses on the various approaches that have been developed to control traffic signals.

B. Traffic Signal Control Strategies

Traffic signal control is essential for managing urban mobility and alleviating congestion at intersections. Traditional strategies range from static fixed-time plans derived from offline traffic studies and invariant to real-time traffic conditions. In contrast, adaptive approaches dynamically adjust signal phases using real-time data from sensors such as inductive loops or cameras, typically following predefined rule-based logic to respond to traffic fluctuations. Coordinated systems further enhance flow by synchronizing multiple intersections, particularly along main corridors [12].

More recently, Artificial Intelligence (AI) has introduced new paradigms for traffic optimization. In particular, Reinforcement Learning (RL) has emerged as a promising alternative, enabling agents to learn optimal signal policies through interaction with a simulated environment [13]. Unlike rule-based systems, RL methods adapt autonomously to varying traffic patterns and aim to maximize long-term network performance. This has led to growing research interest and the development of diverse learning architectures, explored in the following sections.

C. Reinforcement Learning for Traffic Signal Control

Reinforcement Learning (RL) provides a powerful framework for sequential decision-making, where an agent learns to optimize its actions through interaction with an environment [14]. In traffic signal control, this environment represents the traffic network, and the agent aims to manage signal phases at intersections to minimize congestion-related metrics such as delay, queue length, or number of stops [3].

This interaction is modeled as a Markov Decision Process (MDP), defined by the tuple $\langle S, A, P, R, \gamma \rangle$, where S is the set of traffic states (e.g., densities, queues), A the set of signal phase actions, $P(s'|s,a)$ the state transition probabilities, $R(s,a)$ the reward function, and γ the discount factor for future rewards [15]. The agent seeks to learn a policy $\pi : S \to A$ that maximizes the expected cumulative reward:

$$R^t = \sum_{i=0}^{\infty} \gamma^i r^{t+i} \tag{1}$$

Among the most widely used algorithms is the Deep Q-Network (DQN), which approximates the optimal action-value function $Q^*(s,a)$ via deep neural networks and derives the policy as:

$$\pi^*(s) = \arg\max_a Q^*(s,a) \tag{2}$$

The learning process is guided by the minimization of the Bellman loss:

$$L(\theta) = \mathbb{E}_{(s,a,r,s')\sim D}\left[\left(r + \gamma \max_{a'} Q(s',a';\theta^-) - Q(s,a;\theta)\right)^2\right] \tag{3}$$

where experiences (s,a,r,s') are drawn from a replay buffer D, and θ^- denotes parameters of a target network used to stabilize learning.

The performance of RL agents in traffic environments depends critically on the type and quality of data provided during training. Modern simulators such as SUMO [16] provide access to a wide range of traffic indicators, but their availability depends on the level of simulation granularity. Microscopic models provide detailed, vehicle-level data (e.g., positions, speeds, and lane changes), which enable fine-grained control strategies for reinforcement learning agents) [17] [18]. However, they require substantial calibration effort and computational resources. In contrast, macroscopic models, which aggregate traffic features such as lane occupancy and average queue length, are often used by other RL approaches [19] [20] due to their ease of extraction and better scalability for large-scale simulations.

Despite growing interest in RL-based traffic control, there is no clear consensus on the most effective type of observation data [3]. While some approaches rely on detailed microscopic inputs, others suggest that coarser macroscopic data can achieve similar performance. To clarify this, we conduct a systematic comparison using a unified DQN architecture under consistent traffic scenarios, evaluating how

data granularity and agent decision frequency jointly affect learning efficiency and policy quality, a combination rarely explored in prior work.

III. Methodologies

A. Simulation tools

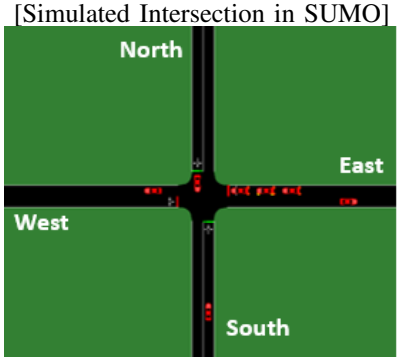
The simulation environment used in this study is SUMO (Simulation of Urban MObility), a microscopic traffic simulator widely adopted [3] with open-source integration. One of the key advantages of SUMO is its ability to generate both microscopic data such as individual vehicle positions and speeds and macroscopic data, including lane-level densities and queue lengths. This makes it particularly well-suited for evaluating the impact of observation granularity in reinforcement learning-based traffic control.

Custom traffic scenarios were created to simulate a single four-leg intersection controlled by a traffic light, fig 1. The simulation duration is fixed at one hour, during which varying traffic densities are applied to the four approaches (north, south, east, and west).

The SUMO-RL interface is used to allow real-time communication between the simulator and the RL agents, enabling dynamic decision-making during training.

subfig

[Simulated Intersection in SUMO]



[Vehicle Insertion Rates Over Time]

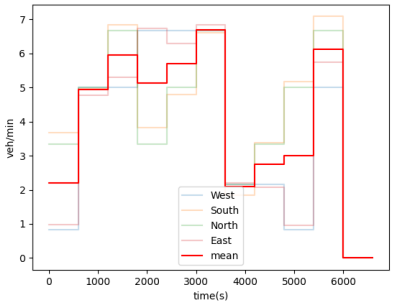


Fig. 1. Illustration of the traffic environment and demand setup.

B. Traffic Scenario

The training scenario consists of a four-leg intersection regulated by a single traffic light, operating with four distinct signal phases. Two of these are primary phases, allowing traffic flow from the north-south and east-west directions, respectively. Intermediate yellow phases are inserted between transitions to satisfy standard safety constraints.

Traffic demand is defined using several types of input flows, characterized by constant vehicle insertion rates ranging from 0 to 7 vehicles per minute. These flows are applied over a simulation period of 1200 seconds (Fig. 1). All vehicles are modeled as homogeneous passenger cars (5 meters in length), allowing the study to focus solely on the effect of control policies without the added complexity of multi-modal dynamics.

Using constant traffic flows during training ensures that the agent is exposed to well-defined, stable traffic densities. Because each decision by the agent alters the environment significantly, a consistent flow rate guarantees that the system remains within a specific density regime. This setup is designed to promote focused learning under controlled conditions, thereby facilitating the association between traffic density and optimal phase selection.

C. Decision Timing Strategies: Fixed vs. Reactive DQN

In reinforcement learning for traffic control, the timing of decision-making plays a crucial role in both learning dynamics and the effectiveness of the resulting policy. In this study, we evaluate two distinct decision-making Deep Q-Network (DQN) variants, differing in the temporal resolution at which they interact with the environment. For clarity, we introduce the names DQNFixed and DQNReact to distinguish these two agent types throughout the paper. These are not standardized terms in the literature but are defined here solely for the purpose of comparison.

The DQNFixed agent [13], [18], [21] selects an action at fixed intervals of duration D. This mirrors traditional fixed-time signal control, producing phase durations that are strict multiples of D. While simple to implement, this structure introduces temporal rigidity in duration phase. Moreover, such fixed-interval control can cause observation discontinuities, especially with microscopic data: the identity and position of vehicles may vary significantly between observations, making it difficult to learn stable state-action mappings. In contrast, macroscopic features like lane density or queue length evolve more gradually and are less sensitive to such temporal sampling gaps.

To address these limitations, this study adopts a reactive control protocol commonly used in the literature [17], referred to here as DQNReact. In this setup, the agent observes the environment at every second and can act at any time step. Once a phase change occurs, it becomes inactive for a fixed duration D, ensuring compliance with minimum green and yellow time requirements.

This frequent interaction improves temporal consistency in the observation space and allows for a greater variety of phase durations. It is particularly beneficial when using high-resolution (microscopic) inputs, as it reduces variability between consecutive states and enables the agent to better adapt signal timing to real-time traffic dynamics.

In our experiments, D is set to 7 seconds (5 seconds green + 2 seconds yellow) to meet safety requirements.

For benchmarking, the study use Webster's formula [22] to compute an optimal fixed cycle length:

$$C_{opt} = \frac{1.5L + 5}{1 - Y} \tag{4}$$

Where L is the total lost time per cycle (in seconds), and $Y = \sum \frac{q_i}{s_i}$, is the sum of flow ratios across phases. This reference enables us to assess whether RL policies converge toward near-optimal timing.

D. Replay Memory Filtering Strategy

Traditionally, reinforcement learning agents store all observed transitions, comprising state, action, reward, and next state in a replay memory buffer to stabilize training through experience replay [14]. However, in the context of traffic signal control, many environmental states may occur in the absence of any vehicles, particularly during low-traffic periods. Storing such transitions provides little to no learning signal and may introduce noise during training.

To enhance learning efficiency, this study filters the replay memory to retain only transitions involving at least one vehicle, ensuring relevance to decision-making. Additionally, the next state s' is limited to vehicles present at the time of the action, focusing learning on directly impacted traffic and avoiding uncertainty from unobserved future flows.

Overall, this selective replay strategy reduces learning noise, accelerates convergence, and reinforces policy learning on traffic states where control actions have meaningful impact.

E. Observation Space

Two observation vectors were implemented to investigate the impact of data granularity on learning performance. The SUMO simulator enables the extraction of both microscopic and macroscopic traffic information, allowing the agent to perceive the environment through either level of abstraction (see Table I).

The microscopic observation vector provides detailed vehicle-level information, while the macroscopic observation vector captures aggregated traffic states. The key characteristics of each observation type are summarized in Table I. Notably, microscopic features include fine-grained information such as individual vehicle speeds, positions, and blinkers, allowing the agent to anticipate maneuvers like left turns that could cause blockages. In contrast, macroscopic features offer a coarser but more stable view through lane occupancy densities and queue lengths, facilitating scalable learning.

F. Reward Function

The reward function is based on the cumulative waiting time of vehicles. After each action, the agent stores the IDs of the vehicles present at the intersection. At each decision-making, the environment calculates the total waiting time accumulated by these vehicles. This formulation encourages the agent to reduce waiting times in successive decisions.

$$R_t = \begin{cases} 1, & \text{if } V_t = 0 \\ R_t = \frac{1}{\sum_{v \in V_t} w_v}, & \text{otherwise} \end{cases} \tag{5}$$

TABLE I
COMPARISON OF OBSERVATION VECTORS USED IN THE DQN MODEL

Observation Type	Features Included in the Observation Vector
Microscopic	• One-hot encoding of the current traffic light phase • Speeds of the 10 closest vehicles • Positions of the 10 closest vehicles relative to the intersection • Turn indicators (blinkers) of the 10 closest vehicles
Macroscopic	• One-hot encoding of the current traffic light phase • Lane occupancy densities for all incoming lanes • Queue lengths on each approach to the intersection

where w_v is the cumulative waiting time of vehicle v observed over the delay interval D after the action is taken, and V_t is the set of vehicles present at the intersection at time t.

IV. EXPERIMENTATION

The experimental protocol is divided into three distinct studies, each targeting a specific aspect of the proposed reinforcement learning framework for traffic signal control. All experiments are conducted using the simulation environment described in Section III-B.

A. Experiment 1: Impact of Replay Memory Filtering

This experiment aims to evaluate the contribution of two specific mechanisms within the reinforcement learning framework: (i) selective storage of transitions in the replay memory, and (ii) targeted vehicle selection for computing rewards and next states, as described in Section III-D.

To this end, this experimentation compare two agents using the DQNFixed architecture. Both are trained under identical conditions using the microscopic observation vector (Section III-E), the simulation environment detailed in Section III-A, and the training traffic scenario presented in Section III-B. The baseline agent uses a standard replay memory that stores all observed transitions indiscriminately and computes rewards based on all vehicles present at the next state. The enhanced agent implements selective replay memory filtering and restricts reward computation to vehicles present at the moment the action was taken.

Results show in Fig. 2 that both agents successfully converge toward a good mean waiting time. However, the enhanced agent exhibits significantly more consistent performance across training epochs, with lower variance in average vehicle waiting times. This suggests that focusing the learning process on meaningful, vehicle-relevant transitions improves the stability and robustness of the learned policy.

Having established the benefits of selective replay memory and vehicle filtering on training stability, the next experiment focuses on evaluating the limitations of fixed decision intervals, particularly in the context of high-resolution microscopic observations.

B. Experiment 2: Impact of Observation Granularity with Fixed Decision Frequency

The working hypothesis is that the microscopic observation vector, due to its higher level of detail, would enable the

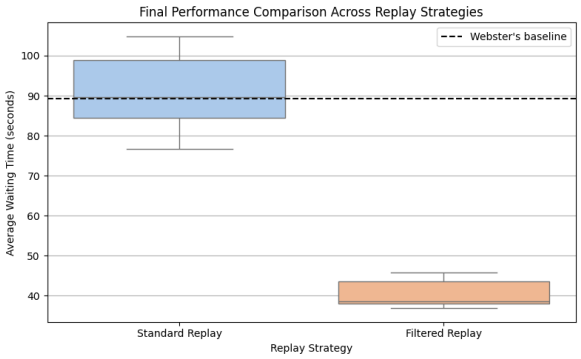


Fig. 2. Final performance comparison between DQNFixed agents with and without replay memory filtering. Each box represents the distribution of average waiting times over the last 10 episodes of 5 training runs of 100 episodes.

agent to learn more precise traffic control policies and reduce average vehicle waiting times. However, this increased granularity also introduces greater variability in the input space across successive time steps, particularly when decisions are made at coarse intervals. For instance, vehicle identities and positions may change drastically between two actions, reducing temporal coherence in the observed state.

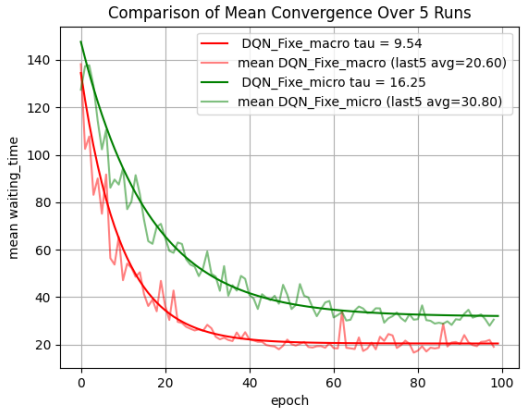


Fig. 3. Convergence comparison of DQNFixed using microscopic and macroscopic observation vectors over 5 training runs. Curves are approximated by an exponential function $y(x) = me^{-x/\tau} + b$, where the time constant τ indicates the convergence rate.

As shown in Fig. 3, the macroscopic model converges significantly faster, reaching 90% of optimal performance with a 41.29% improvement in convergence rate compared to the microscopic model. It also yields better overall results, with a 33.11% reduction in average vehicle waiting time. This outcome is likely due to the increased noise and instability introduced by high-dimensional microscopic features when combined with infrequent decision updates.

These findings confirm that the use of coarse decision intervals in DQNFixed impairs the agent's ability to exploit the richer but more volatile microscopic data. In contrast, macroscopic inputs offer more temporally stable representations, which are better suited to fixed-interval architectures.

To further investigate whether these limitations can be overcome through more frequent agent-environment interactions, the next experiment introduces a reactive agent architecture (DQNReact), which evaluates the environment at every time step while still respecting safety constraints on phase duration.

C. Experiment 3: Observation Granularity under DQNReact

The third experiment aims to evaluate whether high-frequency decision-making, enabled by the DQNReact architecture, improves traffic signal control performance and reduces learning variability. In particular, this experiment investigates whether the use of fine-grained microscopic observations is necessary to achieve high performance, or whether macroscopic inputs combined with reactive control can yield comparable results.

As in previous experiments, the two observation vectors defined in Section III-E are tested under identical traffic conditions and simulation settings. The DQNReact agent evaluates the environment every second and is capable of taking an action at any time step, while still respecting the minimum phase duration constraints (see Section III-C). Its ability to act more frequently than the DQNFixed agent is expected to alleviate the limitations of fixed control cycles identified in Experiment 2.

As shown in Table II, the DQNReact agent achieves lower average waiting times compared to its DQNFixed counterpart, regardless of the observation vector used. The microscopic observation provides only a marginal improvement over the macroscopic variant (2.19% gain), suggesting that high-resolution data may not be necessary when using a reactive control strategy.

More importantly, DQNReact demonstrates a substantial reduction in performance variability. The standard deviation of average waiting time across five runs drops to approximately 1 second, compared to 3.26 seconds for DQNFixed with microscopic inputs. This suggests that increasing the frequency of decision-making not only improves performance but also enhances the robustness and stability of the learning process.

These findings support the idea that macroscopic observations, when combined with a responsive control architecture, are sufficient to train effective reinforcement learning models for traffic signal control, without the need for costly and complex microscopic simulation.

D. Discussion and Insights

Overall, the experimental results demonstrate, with selective replay memory, that macroscopic observation vectors, despite being less detailed, are sufficient to train efficient traffic signal controllers with DQN. In fact, they lead to faster convergence and more stable behavior, particularly when combined with a reactive decision-making strategy.

The comparison between DQNFixed and DQNReact also reveals that frequent decision-making significantly improves policy robustness and reduces performance variability across

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT DQN STRATEGIES AGAINST OPTIMIZED FIXED-TIME TRAFFIC LIGHT (BASELINE/WEBSTER'S FORMULA).

Metric	**Webster's formula**	**DQN_Fixe Micro**	**DQN_Fixe Macro**	**DQN_React Micro**	**DQN_React Macro**
Average waiting time (seconds)	89.26	30.80	20.60	11.42	13.37
% Reduction from baseline	–	65.49%	76.92%	87.21%	85.02%
Standard deviation over the last 10 epochs (5 runs)	–	3.26	4.03	1.00	1.13

training runs. While microscopic data may offer marginal improvements in average waiting time, its benefits do not justify the added complexity and computational cost, especially when simpler, aggregated features can yield near-optimal performance.

These findings support the idea that lower-resolution data can be both effective and practical for reinforcement learning-based traffic control, which opens the door to more scalable training strategies using macroscopic simulation environments.

V. CONCLUSION

This study investigated the impact of data granularity on reinforcement learning-based traffic signal control (TSC), using the Deep Q-Network (DQN) architecture as a learning framework. We compared two agent configurations: DQN-Fixed, which makes decisions at regular fixed intervals, and DQNReact, which operates at a higher frequency with enforced pauses after phase changes. Our experiments evaluated how these configurations perform when trained on either microscopic (vehicle-level) or macroscopic (flow-level) observation data.

The results show that fine-grained microscopic data is not essential for effective policy learning. In particular, DQNReact, benefiting from more frequent decision-making, achieved robust and stable performance even with coarse macroscopic inputs. These findings challenge the common reliance on detailed, heavily calibrated microscopic simulations, and suggest that macroscopic modeling provides a more scalable and computationally efficient alternative for training RL agents in large-scale or real-time TSC scenarios Future work will aim to translate these findings into real-world contexts by applying DQN-based models to multi-intersection traffic networks using real data. A promising direction is to first learn decentralized control strategies from real-world aggregated data using macroscopic models, and then transfer the learned policies to more detailed environments.

REFERENCES

[1] S. Shahid, A. Minhans, and O. Che Puan, "Assessment of Greenhouse Gas Emission Reduction Measures in Transportation Sector of Malaysia," *Jurnal Teknologi*, vol. 70, no. 4, Sep. 2014. [Online]. Available: https://journals.utm.my/index.php/jurnalteknologi/article/view/3481

[2] A. Agrahari, M. M. Dhabu, P. S. Deshpande, A. Tiwari, M. A. Baig, and A. D. Sawarkar, "Artificial Intelligence-Based Adaptive Traffic Signal Control System: A Comprehensive Review," *Electronics*, vol. 13, no. 19, p. 3875, Sep. 2024. [Online]. Available: https://www.mdpi.com/2079-9292/13/19/3875

[3] M. Noaeen, "Reinforcement learning in urban network traffic signal control: A systematic literature review," *Expert Systems With Applications*, 2022.

[4] P. A. Ehlert and L. J. Rothkrantz, "Microscopic traffic simulation with reactive driving agents," in *ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No. 01TH8585)*. IEEE, 2001, pp. 860–865.

[5] G. Costeseque, "Modélisation et simulation dans le contexte du trafic routier," in *Modéliser et simuler. Epistémologies et pratiques de la modélisation et de la simulation*, edited by F. Varenne and M. Silberstein, Editions Matériologiques, 2013, 2013, available: https://enpc.hal.science/hal-00965010.

[6] H. Wei, G. Zheng, V. Gayah, and Z. Li, "Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation," *ACM SIGKDD explorations newsletter*, vol. 22, no. 2, pp. 12–18, 2021.

[7] V. K. Kumarasamy, A. J. Saroj, Y. Liang, D. Wu, M. P. Hunter, A. Guin, and M. Sartipi, "Integration of decentralized graph-based multi-agent reinforcement learning with digital twin for traffic signal optimization," *Symmetry*, vol. 16, no. 4, p. 448, 2024.

[8] V. L. Knoop, "Introduction to traffic flow theory: An introduction with exercises," *Delft University of Technology: Delft, The Netherlands*, 2017.

[9] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation research part B: methodological*, vol. 28, no. 4, pp. 269–287, 1994.

[10] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical review E*, vol. 62, no. 2, p. 1805, 2000.

[11] P. G. Gipps, "A behavioural car-following model for computer simulation," *Transportation research part B: methodological*, vol. 15, no. 2, pp. 105–111, 1981.

[12] M. Eom and B.-I. Kim, "The traffic signal control problem for intersections: a review," *European transport research review*, vol. 12, pp. 1–20, 2020.

[13] C. Ouyang, Z. Zhan, and F. Lv, "A comparative study of traffic signal control based on reinforcement learning algorithms," *World Electric Vehicle Journal*, vol. 15, no. 6, p. 246, 2024.

[14] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[15] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[16] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "Sumo–simulation of urban mobility: an overview," in *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*. ThinkMind, 2011.

[17] P. Chen, Z. Zhu, and G. Lu, "An adaptive control method for arterial signal coordination based on deep reinforcement learning," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 3553–3558.

[18] J. Ault and G. Sharon, "Reinforcement learning benchmarks for traffic signal control," in *Proceedings of the Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS 2021) Datasets and Benchmarks Track*, December 2021.

[19] Z. Qu, Z. Pan, Y. Chen, X. Wang, and H. Li, "A distributed control method for urban networks using multi-agent reinforcement learning based on regional mixed strategy nash-equilibrium," *IEEE Access*, vol. 8, pp. 19 750–19 766, 2020.

[20] W. Genders and S. Razavi, "Policy analysis of adaptive traffic signal control using reinforcement learning," *Journal of Computing in Civil Engineering*, vol. 34, no. 1, p. 04019046, 2020.

[21] T. Pan, "Traffic light control with reinforcement learning," *arXiv preprint arXiv:2308.14295*, 2023.

[22] F. V. Webster, "Traffic signal settings," Tech. Rep., 1958.