Multi-agent Reinforcement Learning vs. Fixed-Time Control for Traffic Signal Optimization: A Simulation Study

Saahil Mahato
Independent Researcher
Bhaktapur, Nepal
mahatosaahil4@gmail.com

***Abstract*—Urban traffic congestion, particularly at intersections, significantly impacts travel time, fuel consumption, and emissions. Traditional fixed-time signal control systems often lack the adaptability to manage dynamic traffic patterns effectively. This study explores the application of multi-agent reinforcement learning (MARL) to optimize traffic signal coordination across multiple intersections within a simulated environment. Utilizing Pygame, a simulation was developed to model a network of interconnected intersections with randomly generated vehicle flows to reflect realistic traffic variability. A decentralized MARL controller was implemented, in which each traffic signal operates as an autonomous agent, making decisions based on local observations and information from neighboring agents. Performance was evaluated against a baseline fixed-time controller using metrics such as average vehicle wait time and overall throughput. The MARL approach demonstrated statistically significant improvements, including reduced average waiting times and improved throughput. These findings suggest that MARL-based dynamic control strategies hold substantial promise for improving urban traffic management efficiency. More research is recommended to address scalability and real-world implementation challenges.**

***Index Terms*—Multi-Agent Reinforcement Learning, Traffic Signal Control, Intelligent Transportation Systems**

I. Introduction

Urban traffic congestion is a significant challenge faced by cities worldwide, resulting in increased travel times, higher fuel consumption, and greater environmental pollution. Traditional traffic signal control methods, such as fixed-time control, operate on predetermined schedules that fail to adapt to real-time traffic conditions. Although these systems are easy to implement and maintain, they often do not perform well in dynamic environments characterized by unpredictable traffic patterns. To overcome these limitations, adaptive traffic control strategies have been developed, utilizing real-time data to adjust signal timings for optimal traffic flow. Recent advancements in artificial intelligence, particularly in Reinforcement Learning (RL), have introduced Multi-Agent Reinforcement Learning (MARL), wherein each traffic signal acts as an autonomous agent capable of learning optimal control policies through environmental interactions. This study aims to conduct a comparative analysis of the performance of a traditional fixed-time controller versus a MARL-based controller in a simulated multi-intersection traffic environment, using the Pygame platform to model a network with randomly spawned vehicles that emulate real-world traffic variability. The research specifically examines the effectiveness of MARL in reducing average vehicle travel time, minimizing queue lengths, and improving overall traffic throughput compared to fixed-time control systems.

II. Research Objective

The primary objective of this research is to conduct a comparative analysis between a fixed-time traffic signal controller and a multi-agent reinforcement learning (MARL)-based controller within a simulated multi-intersection traffic network. The study aims to evaluate the performance of each control strategy by simulating realistic traffic flow scenarios, systematically recording key performance metrics such as average waiting time and throughput. Through statistical comparison of the results, the research seeks to determine the relative effectiveness of each approach and assess the significance of any observed differences, thus providing empirical insight into the potential advantages of adaptive, learning-based traffic control systems over traditional fixed-time methods.

III. Related Work

Optimization of traffic signal control has been a central theme of research for decades, evolving significantly from traditional static timing strategies to more sophisticated intelligent systems. This section critically evaluates the progression of traffic signal control methodologies, elucidating the limitations of conventional approaches while underscoring the transformative advancements introduced by artificial intelligence, with a particular focus on Multi-agent Reinforcement Learning (MARL).

A. Fixed-Time and Actuated Signal Control

Conventional traffic signal control methods, particularly fixed-time control, operate based on pre-established schedules that fail to adjust to real-time traffic conditions. While these systems are relatively straightforward in implementation, their static nature often results in inefficiencies in dynamic traffic environments [1]. Contrasting this, actuated signal control introduced a degree of responsiveness by adjusting signal phases based on data obtained from sensors. However, despite this advancement, such systems still lack the requisite adaptability to effectively manage the complexities inherent in urban traffic scenarios.

B. Adaptive Traffic Control Systems (ATCS)

To address the deficiencies associated with static control methods, Adaptive Traffic Control Systems (ATCS) emerged as a progressive alternative. Prominent examples of ATCS include the Split Cycle Offset Optimization Technique (SCOOT) and the Sydney Coordinated Adaptive Traffic System (SCATS). SCOOT employs real-time traffic flow data to adaptively adjust signal timings, demonstrating performance enhancements of approximately 15% relative to fixed-time systems [2]. SCATS, which has been successfully implemented in over 180 cities, dynamically modifies signal phases in response to actual traffic data, thereby improving traffic flow and mitigating congestion more effectively than earlier methodologies [3].

C. Artificial Intelligence in Traffic Signal Control

The integration of artificial intelligence (AI) into the realm of traffic management has catalyzed substantial advancements in efficiency and performance. For instance, systems like SURTRAC utilize a decentralized, schedule-driven approach for intersection control, achieving average reductions in travel time that exceed 25% in pilot deployments [4]. Furthermore, Google's Green Light project exemplifies the application of AI in optimizing traffic signal timings, yielding reductions of up to 30% in idle time at red lights—accomplished without necessitating new hardware installations [5]. This highlights the potential of AI-driven solutions to revolutionize traffic signal control by enabling a more fluid and responsive traffic management system.

D. Reinforcement Learning Approaches

Reinforcement Learning (RL) has gained prominence as a potent tool for the adaptive control of traffic signals. While single-agent RL methodologies have demonstrated their effectiveness in optimizing the operations of individual intersections, they often encounter scalability challenges when applied to larger traffic networks. To mitigate these limitations, Multiagent Reinforcement Learning (MARL) frameworks have been proposed, facilitating decentralized decision-making among multiple agents.

In a significant advancement, Chu et al. (2019) introduced a scalable MARL algorithm that utilizes the Advantage Actor-Critic (A2C) approach, specifically designed for large-scale traffic networks. Their findings indicate that this method exhibits both robustness and sample efficiency, performing effectively in both synthetic environments and real-world applications [6]. Building on this foundation, Zhang et al. (2019) developed QCOMBO, a MARL algorithm that merges independent and centralized learning strategies. Their research demonstrates that QCOMBO achieves competitive performance across a variety of road topologies, thereby enhancing the versatility of MARL applications in traffic control [7]. Additionally, Devailly et al. (2020) contributed to this field by creating IG-RL, which leverages graph-convolutional networks to facilitate the development of transferable adaptive traffic signal control policies adaptable to diverse network structures [8].

Furthering the exploration of cooperative strategies, Li et al. (2021) presented KS-DDPG, an innovative framework that integrates a knowledge-sharing communication protocol among agents. This enhancement significantly improves cooperation and control efficiency within large-scale transportation networks [9]. Moreover, the exploration of Mean Field Reinforcement Learning methods has revealed promising results, showcasing improved convergence rates and diminished vehicle loss times in comparison to traditional deep RL methodologies [10].

E. Multimodal and Safety-Oriented Traffic Control

Recent investigations have increasingly turned their attention to multimodal traffic environments and the imperative of safety optimization in signal control. A noteworthy contribution in this domain is the decentralized MARL approach known as eMARLIN-MM, which was specifically designed to optimize signal timings for both vehicular and transit traffic. This approach has successfully achieved significant reductions in total person delays [11]. In a complementary study, Essa and Sayed developed RS-ATSC, an adaptive traffic signal control algorithm that is capable of optimizing intersection safety in real-time, utilizing data derived from connected vehicles [12].

F. Comprehensive Reviews and Future Directions

Comprehensive reviews of the literature have underscored the transformative potential of integrating AI, IoT, and predictive analytics into adaptive traffic control systems. Various techniques, including fuzzy logic, deep neural networks, and hybrid models that fuse Long Short-Term Memory networks (LSTMs) with Convolutional Neural Networks (CNNs), have been employed to analyze complex traffic patterns. These methodologies enable dynamic adjustments of signal timings, thereby effectively alleviating congestion and minimizing queuing issues [13]. As the field continues to evolve, ongoing research will likely further refine these approaches, promoting smarter and safer traffic management solutions.

G. Summary

The evolution of traffic signal control methodologies signifies a critical shift towards the implementation of intelligent and adaptive systems that can respond effectively to real-time traffic dynamics. Multi-Agent Reinforcement Learning (MARL) approaches, in particular, have emerged as promising solutions that offer scalable and decentralized strategies for managing the complexities inherent in urban traffic networks. This research endeavors to build upon these advancements by rigorously comparing the performance of a traditional fixed-time traffic signal controller with that of a MARL-based controller in a simulated multi-intersection environment. The findings of this comparative study aim to contribute valuable insights to the ongoing development of efficient traffic management strategies, ultimately enhancing urban mobility and reducing congestion.

IV. Methodology

To assess the efficacy of fixed-time control in comparison to Multi-agent Reinforcement Learning (MARL) for optimizing traffic signal management, we developed a bespoke traffic simulation environment utilizing the Pygame library. This environment enables systematic experimentation across a spectrum of traffic scenarios and control strategies. The source code for this simulation is publicly accessible in our GitHub Repository [14].

A. Simulation Environment

The simulation environment is meticulously crafted to replicate a simplified urban traffic network, providing a platform for analyzing traffic flow dynamics under various signal control methodologies.

1) Simulation Framework: At the core of our simulation is Pygame, a versatile cross-platform suite of Python modules specifically designed for game development. Pygame offers extensive capabilities for rendering graphics and managing real-time events [15]. Its inherent flexibility renders it particularly suitable for constructing custom simulations necessitating dynamic visualizations and interactive user engagement.

2) Road Network Configuration: The simulated road network is structured as a 900×900 pixel grid, designed to represent a four-way intersection characterized by four traffic lights positioned at the coordinates $(300, 300)$, $(600, 300)$, $(300, 600)$, and $(600, 600)$. Each intersection operates independently, enabling the implementation and comparative analysis of diverse traffic signal control strategies.

3) Traffic Signal Parameters: Traffic signals function on a predetermined fixed-time cycle, which comprises three distinct phases:

- **Green Phase**: Lasting for 5 seconds, during this time vehicles traveling in the horizontal direction are allowed to proceed, while vertical vehicle movement is halted.
- **Red Phase**: Also spanning 5 seconds, this phase permits vehicles moving in the vertical direction to advance, whereas vehicles in the horizontal direction must remain stationary.
- **Yellow Phase**: This critical 2-second interval mandates that all vehicles, irrespective of their direction, come to a complete stop.

The operational behavior of the traffic lights is uniformly applied across all intersections in the fixed-time control simulations. Conversely, in simulations employing Multi-Agent Reinforcement Learning (MARL), the signal phases are dynamically adjusted per the policies developed by the learning agents.

4) Vehicle Generation and Movement: To simulate realistic traffic patterns, vehicles are generated at eight designated entry points at precise intervals of 0.5 seconds. The specific spawn locations and their corresponding movement directions are delineated as follows:

- $(0, 280)$ moving right
- $(0, 580)$ moving right
- $(900, 320)$ moving left
- $(900, 620)$ moving left
- $(320, 0)$ moving down
- $(620, 0)$ moving down
- $(280, 900)$ moving up
- $(580, 900)$ moving up

Each vehicle adheres to a predetermined trajectory, observing crucial traffic rules such as halting at red signals and proceeding during green phases. For the sake of simplicity and to maintain consistency in the simulation environment, vehicles are modeled to travel at a constant speed and are restricted to linear movement. Furthermore, vehicles are programmed to stop when a vehicle traveling in the same direction comes to a halt immediately ahead of them.

5) Traffic Light Vehicle Count Display: To augment the analytical capabilities of the simulation framework, each traffic light is outfitted with a display mechanism that provides a real-time count of approaching vehicles. This feature serves as a crucial tool for monitoring local traffic density at each intersection, thereby facilitating a more comprehensive assessment of traffic flow and potential congestion levels. The implementation of such displays has been shown to enhance the optimization of signal timings, as documented in previous traffic simulation studies [16], [17].

6) Simulation Dynamics: The simulation operates at a frame rate of 60 frames per second (FPS) over a total duration of 10 minutes for each experimental run. Throughout each simulation, data is systematically collected on critical performance metrics, specifically:

- The total number of vehicles that successfully navigate through the intersection
- The average wait time experienced by each vehicle

These metrics serve as fundamental quantitative indicators for assessing the effectiveness of various traffic signal control strategies.

7) Visualization: A representative screenshot of the simulation environment is displayed in Figure 1. This figure captures the intricate layout of the road network, including the positioning of traffic signals and vehicles, as well as the vehicle count displays observed during a typical run of the simulation.

8) Relevance and Applications: Employing a Pygame-based simulation facilitates a controlled and adaptable environment for the rigorous testing and comparative analysis of diverse traffic control strategies. This approach aligns with methodologies utilized in prior studies aimed at modeling and scrutinizing traffic systems [17], [16]. The insights derived from such simulations hold the potential to significantly enhance the development of more efficient traffic management systems in real-world urban contexts.

B. Traffic Signal Control Strategies

Effective management of urban traffic flow is essential for optimizing transportation systems. In this study, we explore

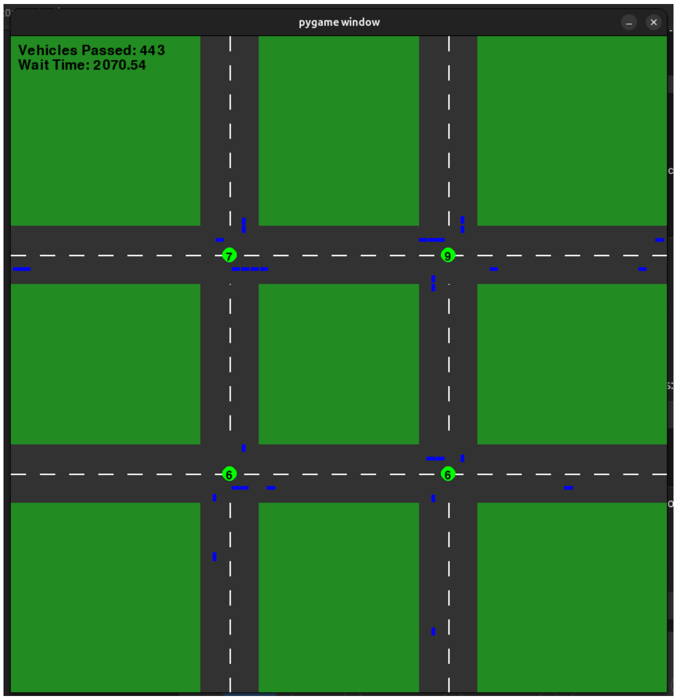

Fig. 1. Screenshot of the Pygame-based traffic simulation environment

two principal traffic signal control strategies: the traditional fixed-time controller and a Multi-Agent Reinforcement Learning (MARL) based controller. This section provides an in-depth examination of the fixed-time control approach utilized in our simulation.

1) Fixed-Time Controller: The fixed-time control strategy is characterized by its operation on a pre-established schedule, wherein traffic signal phases transition through fixed durations without regard for real-time traffic conditions. This method is commonly adopted in traffic management frameworks due to its inherent simplicity and ease of implementation, particularly in contexts where traffic patterns exhibit relative consistency or where infrastructural limitations preclude the adoption of adaptive control mechanisms [18], [19].

a) Determination of Fixed Timings: In our simulation, the fixed timings were determined following established traffic engineering principles and guidelines. Specifically, the cycle lengths and phase durations were designed to simulate typical urban intersection scenarios, thereby ensuring a balanced and equitable allocation of green time across conflicting traffic movements.

The total cycle length was calibrated to 24 seconds, incorporating the following phases:

- **Green Phase**: 5 seconds
- **Yellow Phase**: 2 seconds
- **Red Phase**: 5 seconds

This sequence facilitates alternating green phases for perpendicular traffic flows, while the yellow and red phases provide crucial clearance intervals. The selected durations are consistent with recommendations from the Federal Highway Administration (FHWA), which advocate for green times between 20 to 40 seconds for local roads, alongside yellow intervals typically ranging from 3 to 5 seconds, contingent upon approach speeds and intersection widths [18].

b) Implementation in Simulation: In the simulation environment, each traffic signal operates on a predetermined fixed-time schedule, functioning independently of adjacent signals. The controller systematically cycles through predefined phases, updating the status of the traffic lights at each interval. Vehicles react to the current state of the traffic signal at their respective intersections—advancing during green phases while halting during red and yellow phases.

This methodology establishes a fundamental baseline for assessing the efficacy of more sophisticated control strategies, such as the Multiagent Reinforcement Learning (MARL)-based controller. By providing a consistent and predictable traffic signal behavior, the fixed-time approach enables a clear framework against which the improvements offered by advanced methodologies can be evaluated.

2) Multiagent Reinforcement Learning Controller: To promote enhanced adaptability and responsiveness in traffic signal management, a Multiagent Reinforcement Learning (MARL) approach has been adopted. Within this framework, each traffic light is designed to function as an autonomous learning agent, employing Deep Q-Networks (DQN) to optimize traffic signal phases per the observed traffic conditions.

Agent Definition: In this context, each traffic light at an intersection is conceptualized as an independent agent. These agents are tasked with making decisions autonomously, with the overarching goal of optimizing traffic flow specific to their respective intersections. This decentralized framework facilitates scalability and enhances resilience within intricate traffic networks [20].

State Space: The state representation for each agent incorporates a comprehensive array of both local and neighboring traffic information, encompassing:

- **Queue Lengths:** The number of vehicles approaching from each direction, categorized as North, South, East, or West.
- **Average Distances:** The mean distances of approaching vehicles relative to the intersection.
- **Movement Status:** The proportion of vehicles in motion within each lane.
- **Spatial Features:** The relative positions of vehicles to effectively capture traffic density and distribution.

This extensive state representation empowers agents to make well-informed decisions by incorporating insights not only from their immediate surroundings but also from the broader traffic context [21].

Action Space: In this study, the actions available to the agents are categorized into three discrete phases corresponding to typical traffic light operations:

- **Green Phase:** This phase permits the uninterrupted flow of vehicular traffic in the designated primary direction, thereby facilitating efficient movement through the intersection.

- **Yellow Phase:** The yellow phase serves as a transitional indicator, signaling drivers of an upcoming change in light status, and urging them to prepare for a potential halt in movement.
- **Red Phase:** During this phase, traffic flow in the primary direction is halted, ensuring the safety of all road users and allowing cross traffic to proceed.

These selected actions are meticulously designed to optimize traffic management while prioritizing safety during transitions between phases, which is vital for urban traffic systems.

Reward Function: The reward function in this framework is constructed to balance diverse traffic management objectives, thereby creating an effective incentive structure:

$$R = M - \alpha Q - \beta S,$$

Where the variables are defined as follows:

- M: The number of vehicles that successfully navigate through the intersection, representing a measure of throughput.
- Q: The cumulative queue length across all approaches to the intersection, indicative of congestion levels.
- S: The total number of vehicles that are brought to a stop during the traffic light phases.
- α, β: Weighting factors that have been empirically set to values of 0.1 and 0.2, respectively, to reflect their relative importance in the reward calculation.

This formulation is strategically designed to incentivize agents to reduce congestion and minimize delays while simultaneously maximizing the throughput of vehicles, as referenced in [22].

Algorithm: The agents utilize Deep Q-Networks (DQN) to derive optimal traffic light policies through reinforcement learning techniques. Each DQN architecture encompasses the following components:

- **Input Layer:** This layer processes the state representation of the traffic environment, effectively converting environmental inputs into a format suitable for learning.
- **Hidden Layers:** Comprising two fully connected layers equipped with ReLU activation functions, these layers facilitate the learning of complex representations.
- **Output Layer:** The output layer generates Q-values corresponding to each possible action, providing the basis for the agent's decision-making process.

To enhance stability during the learning process, a target network is employed, while experience replay buffers retain past experiences that are utilized for reinforcing the training [23].

Training Process: The agents are trained within a simulated environment that accurately reflects real-world traffic conditions, thereby ensuring the validity of the learning outcomes. The key parameters governing the training regimen include:

- **Episodes:** A total of 10,000 training episodes are conducted to ensure comprehensive exploration of the action space.
- **Episode Duration:** Each training episode lasts for 300 seconds, providing sufficient time for the agents to learn effective policies.
- **Optimizer:** The Adam optimizer is employed, with a learning rate set at 1×10^{-3} to facilitate efficient convergence.
- **Discount Factor (γ):** A discount factor of 0.99 is utilized, prioritizing long-term rewards and fostering strategic foresight in decision-making.
- **Batch Size:** A mini-batch size of 64 is adopted for gradient descent optimization, striking a balance between computational efficiency and stability.
- **Epsilon-Greedy Strategy:** The epsilon value is systematically decayed from 1.0 to 0.05 throughout training, effectively balancing the exploration of new strategies with the exploitation of known rewarding actions.

This structured training environment is intentionally designed to mirror the evaluation conditions, thereby ensuring consistency in performance assessment and validation of the agents' learned behaviors.

Visualizations: To provide a clearer understanding of the learning dynamics and the behavior of the agent, we present the following visualizations:

- **DQN Architecture:** As illustrated in Figure 2, the architecture of the Deep Q-Network (DQN) employed by each agent consists of multiple layers, which enable the agent to process and learn from the high-dimensional input space effectively.
- **Q-Values Heatmap:** Figure 3 presents a sample heatmap representing the Q-values assigned to different actions across various states. This visualization serves to illustrate the agent's preferences, showcasing areas where certain actions are favored over others based on the agent's learned experiences.

C. Experimental Design

1) Simulation Framework: To rigorously evaluate the performance of the Fixed-Time (FT) controller in comparison to the Multi-Agent Reinforcement Learning (MARL) controller, we established a discrete-event traffic simulation model. Within this framework, vehicles are generated randomly at one of eight predefined locations at intervals of 0.5 seconds. Notably, all other elements of the simulation—including vehicle dynamics and traffic signal operations—are deterministic. This methodological structure ensures that the sole source of stochasticity lies in the vehicle spawning process, thereby permitting a controlled and valid comparison between the two controller types.

2) Handling Randomness and Ensuring Fairness: The inherent randomness associated with vehicle spawning necessitates the execution of multiple simulation runs to attain statistically significant findings. Relying on a singular simulation run risks yielding misleading conclusions, potentially skewed by outlier effects [24]. Accordingly, we undertook 20 independent simulation runs for each type of controller, each utilizing a

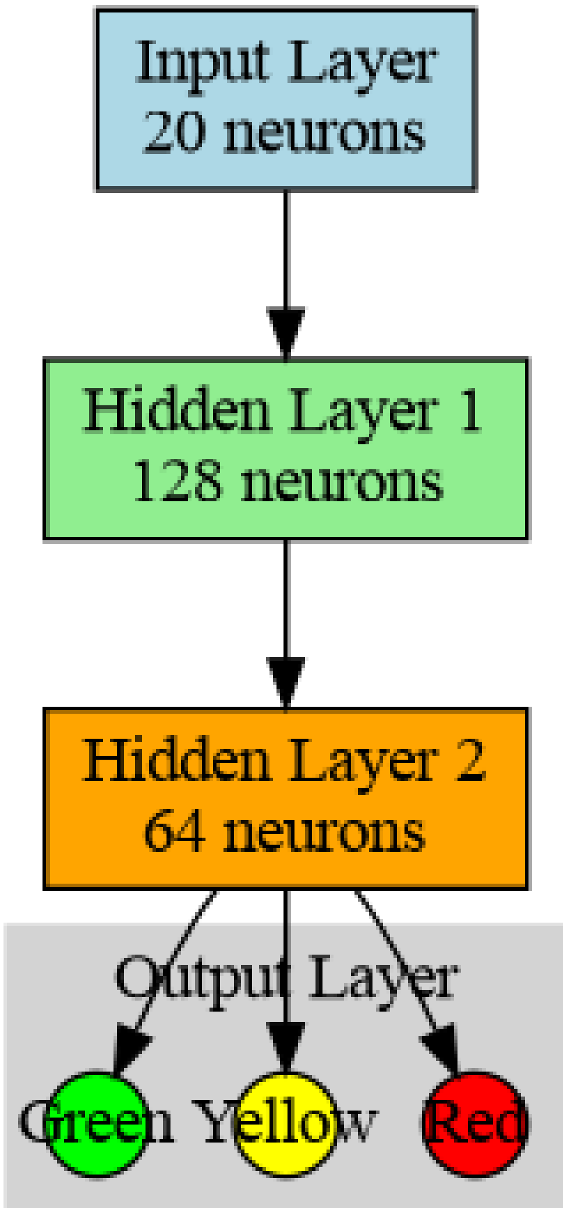

Fig. 2. Deep Q-Network Architecture

Fig. 3. Sample Q-Values Heatmap

distinct random seed to foster variability in vehicle spawning patterns.

This methodology aligns with established best practices in traffic simulation, which advocate for conducting multiple runs to adequately capture the variability intrinsic to stochastic systems [25]. By averaging the outcomes across these runs, we effectively mitigated the influence of random fluctuations, thereby facilitating a fair and equitable comparison between the FT and MARL controllers.

3) Justification for Experimental Design: The decision to implement 20 simulation runs per controller finds strong support in the literature, which highlights the critical importance of multiple iterations in stochastic simulations to generate reliable and valid results [24]. Furthermore, isolating the randomness solely to the vehicle spawning process assures that any observed performance differentials are attributable directly to the controllers themselves, devoid of interference from extraneous variables.

By adhering to such rigorous methodological principles, we have positioned our experimental design to provide a robust and fair comparison between the FT and MARL controllers.

D. Performance Metrics

In our investigation of traffic signal control strategies, we identified two primary performance metrics that serve as critical indicators of efficacy:

- **Average Vehicle Delay:** This metric calculates the mean duration that vehicles are required to wait at intersections, thereby providing a direct reflection of traffic flow efficiency and the effectiveness of signal timing strategies. The Highway Capacity Manual (HCM) identifies average vehicle delay as a fundamental metric for assessing the Level of Service (LOS) at signalized intersections, highlighting its pivotal role in traffic performance evaluation [26].
- **Throughput:** Throughput is defined as the total count of vehicles that successfully traverse the traffic network within a specified time. This metric serves as a quantitative indicator of both the capacity and efficiency of the traffic system. Elevated throughput values are indicative of enhanced traffic movement and a reduction in congestion levels [27].

Together, these metrics provide a comprehensive framework for assessing the performance of traffic signals. While average vehicle delay emphasizes the temporal aspect of traffic efficiency, throughput addresses the volumetric capacity of the network.

Exclusion of Queue Length: Despite its common use in traffic studies, we chose to omit queue length as a performance metric in our analysis. This decision is rooted in several key considerations:

- **Redundancy with Wait Time:** Queue length and vehicle delay are closely related; an increase in queue length often correlates with extended wait times for vehicles. Consequently, the simultaneous measurement of both metrics may generate redundant data that fails to provide meaningful additional insights [18].
- **Practical Measurement Challenges:** Accurately determining queue lengths, particularly in simulation environments, presents numerous challenges and may not yield significantly different information in comparison to delay metrics. In contrast, delay measurements are generally more straightforward and have direct implications on driver experience and overall traffic flow efficiency [28].

By concentrating on average vehicle delay and throughput, we aim to present a robust and efficient evaluation of traffic signal control strategies, while avoiding the complexities and potential redundancies associated with queue length measurements.

E. Hypothesis Testing

1) Formulation of Hypotheses: To rigorously evaluate the performance discrepancies between the Fixed-Time (FT) controller and the Multi-Agent Reinforcement Learning (MARL) controller, we established the following hypotheses:

- **Null Hypothesis (H_0):** There exists no significant difference in the average wait time and the total number of vehicles passed between the FT and MARL controllers.
- **Alternative Hypothesis (H_1):** The MARL controller yields a significantly lower average wait time and a higher total number of vehicles passed when compared to the FT controller.

2) Statistical Tests Employed: Considering the nature of our data and the necessity to draw comparisons between two independent groups, we implemented a combination of both parametric and non-parametric statistical tests. This dual approach ensures a comprehensive and robust analysis of our results.

Normality and Homogeneity of Variance Tests: Before selecting the appropriate statistical tests, it was imperative to assess the underlying assumptions of normality and homogeneity of variances:

- **Shapiro-Wilk Test:** Utilized to evaluate the normality of the data distributions associated with each metric under investigation.
- **Levene's Test:** Employed to assess the equality of variances between the two groups, ensuring that the assumptions required for parametric testing are met.

The outcomes of these preliminary tests provided critical insights that guided our choice of subsequent statistical analyses, ultimately enhancing the rigor of our conclusions.

Independent Samples t-Test: In our analysis, we first applied the independent samples t-test for metrics where the assumptions of normality and equal variances were satisfied. This statistical test is deemed appropriate for comparing the means of two independent samples under the condition that the data are normally distributed and exhibit equal variances [29]. The t-test enables us to determine whether there is a statistically significant difference between the group means.

Mann-Whitney U Test: When the assumptions required for the t-test were not upheld, we turned to the Mann-Whitney U test. This non-parametric alternative is particularly valuable as it does not assume normality, making it suitable for datasets that are either ordinal or not normally distributed [30]. By employing the Mann-Whitney U test, we can effectively compare the differences between two independent groups without the limitations imposed by parametric tests.

3) Significance Level: All statistical tests within this study were conducted using a significance level of $\alpha = 0.05$. This threshold strikes a balance between the risks of Type I and Type II errors, serving as a standardized criterion for establishing statistical significance in hypothesis testing [31]. By adhering to this significance level, we aim to ensure the reliability of our findings.

4) Effect Size and Power Analysis: To quantify the magnitude of the observed differences, we calculated Cohen's d for each metric analyzed. This metric provides insight into the effect size, offering valuable context alongside p-values. Moreover, we conducted a power analysis to ascertain the probability of correctly rejecting the null hypothesis when it is indeed false. Conducting power analysis is essential for understanding the sensitivity of our statistical tests and confirming that our sample size was sufficient for detecting meaningful effects [32].

5) Bootstrap Confidence Intervals: To augment the robustness of our findings, we employed bootstrap resampling methods to construct 95% confidence intervals for the mean differences observed between groups. Bootstrap methods are advantageous due to their non-parametric nature, as they do not rely on the assumption of normality. This characteristic makes them particularly suitable for estimating the sampling distribution of our statistics [33].

6) Summary of Statistical Approach: Our comprehensive statistical approach integrates both parametric and non-parametric testing methods, effect size calculations, power analysis, and bootstrap confidence intervals. This multifaceted methodology ensures a robust and reliable comparison between the FT and MARL controllers. By accounting for potential violations of statistical assumptions, we provide an in-depth examination of the performance differences between these two traffic control strategies, contributing to a nuanced understanding of their efficacy in practice.

V. Results

A. Performance Comparison of Fixed-Time and MARL Controllers

To assess the effectiveness of the Multi-Agent Reinforcement Learning (MARL) controller in comparison to the conventional fixed-time traffic signal controller, we performed a comprehensive series of simulations aimed at evaluating two key performance metrics: the number of vehicles successfully traversing the intersection and the cumulative wait time incurred by vehicles. The findings of our analysis are presented in Table I, which provides a detailed summary of the simulation results.

TABLE I
DETAILED PERFORMANCE METRICS FOR FIXED-TIME AND MARL CONTROLLERS

Index	Vehicles Passed	Wait Time (s)	Controller
1	1146	5283.23	Fixed-Time
2	1148	5193.31	Fixed-Time
3	1146	5262.83	Fixed-Time
4	1146	5081.75	Fixed-Time
5	1144	5250.06	Fixed-Time
6	1146	5327.18	Fixed-Time
7	1146	5228.23	Fixed-Time
8	1146	5347.56	Fixed-Time
9	1146	5310.16	Fixed-Time
10	1149	5249.45	Fixed-Time
11	1148	5280.27	Fixed-Time
12	1148	5382.48	Fixed-Time
13	1145	5244.82	Fixed-Time
14	1143	5227.23	Fixed-Time
15	1147	5374.63	Fixed-Time
16	1145	5370.44	Fixed-Time
17	1146	5150.83	Fixed-Time
18	1147	5364.79	Fixed-Time
19	1149	5142.19	Fixed-Time
20	1147	5205.03	Fixed-Time
21	1154	1149.46	MARL
22	1155	1104.94	MARL
23	1154	1156.51	MARL
24	1153	1111.07	MARL
25	1155	1107.69	MARL
26	1152	1128.64	MARL
27	1152	1208.07	MARL
28	1153	1166.85	MARL
29	1150	1129.98	MARL
30	1153	1135.63	MARL
31	1154	1154.45	MARL
32	1152	1173.42	MARL
33	1152	1145.04	MARL
34	1152	1115.58	MARL
35	1155	1174.28	MARL
36	1155	1148.37	MARL
37	1153	1122.54	MARL
38	1153	1137.87	MARL
39	1153	1168.26	MARL
40	1153	1156.70	MARL

The Multi-Agent Reinforcement Learning (MARL) controller exhibited a marginal increase of approximately 0.59% in the average number of vehicles that successfully passed through the intersection. More significantly, it accomplished a substantial reduction of 78.26% in average wait time when compared to the traditional fixed-time controller.

B. Statistical Analysis of Performance Metrics

To ascertain the statistical significance of the differences observed in performance metrics, we conducted hypothesis testing for both the number of vehicles passed and average wait time.

1) Vehicles Passed/ Throughput:

- **Normality Tests:** The Shapiro-Wilk test was employed to evaluate the normality of the data distributions, yielding p-values of 0.225 for the fixed-time controller and 0.053 for the MARL controller. These results suggest that the data for both controllers adheres to an approximately normal distribution.
- **Variance Homogeneity:** Levene's test was performed to assess the homogeneity of variances, resulting in a p-value of 0.641. This indicates that the variances between the two groups can be considered equal.
- **Independent Samples t-Test:** The independent samples t-test revealed a t-statistic of -14.9612, accompanied by a p-value of 1.64×10^{-17}. This indicates a statistically significant difference in the number of vehicles passed between the two controllers.
- **Effect Size:** The effect size was calculated using Cohen's d, which yielded a value of 4.7311. This signifies a very large effect size, suggesting a meaningful impact of the MARL controller.
- **Mann-Whitney U Test:** To further corroborate the findings, the Mann-Whitney U Test was conducted, yielding a U-statistic of 0.0000 with a p-value of 5.22×10^{-8}. This result aligns with the conclusions drawn from the t-test.
- **Bootstrap 95% Confidence Interval:** Finally, a bootstrap analysis was performed, generating a 95% confidence interval for the mean difference, which ranged from [5.90, 7.60]. This interval reinforces the statistical significance found in the comparative assessments.

2) Wait Time: In this section, we analyze the wait times associated with the fixed-time controller and the MARL controller.

- **Normality Tests:** To assess the normality of the wait time distributions, we conducted the Shapiro-Wilk test. The results yielded p-values of 0.537 for the fixed-time controller and 0.679 for the MARL controller, indicating that both datasets follow a normal distribution.
- **Variance Homogeneity:** We performed Levene's test to evaluate the equality of variances between the two groups. The test produced a p-value of 3.50×10^{-4}, which suggests that the variances are significantly unequal.
- **Welch's t-Test:** Given the evidence of unequal variances, we utilized Welch's t-test to compare the mean wait times. The analysis resulted in a t-statistic of 209.1100 and a p-value of 8.59×10^{-39}, indicating a highly significant difference between the wait times of the two controllers.
- **Effect Size:** To quantify the magnitude of this difference, we calculated Cohen's d, which was found to be -66.1264. This value suggests an extremely large effect size, emphasizing the practical significance of the findings.
- **Mann-Whitney U Test:** Further supporting our results, the Mann-Whitney U test yielded a U-statistic of 400.0000 and a p-value of 3.40×10^{-8}. This non-parametric test corroborates the findings of the Welch's t-test, reinforcing the evidence for a significant difference.
- **Bootstrap 95% Confidence Interval:** Finally, we con-

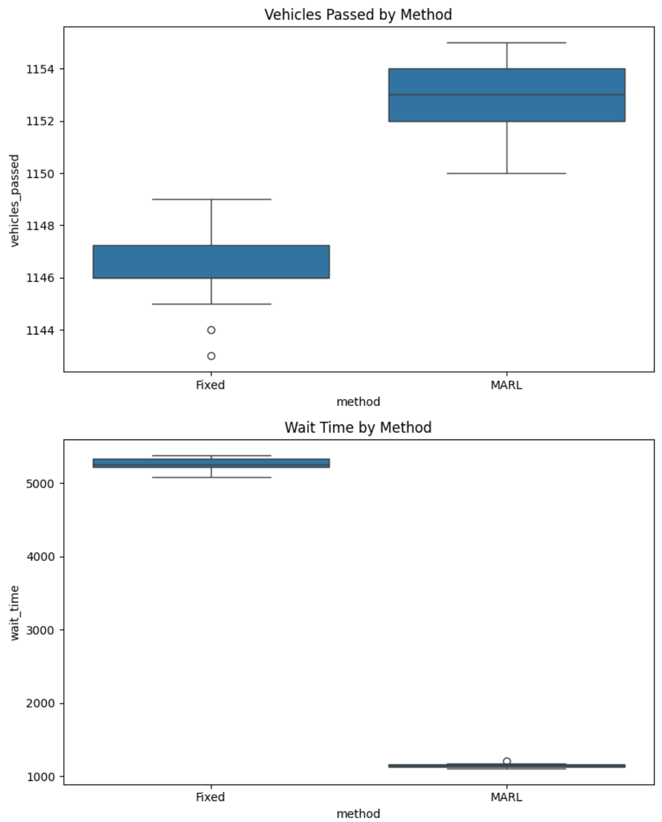

Fig. 4. Boxplots Comparing the Number of Vehicles Passed and Wait Times for Fixed-Time and MARL Controllers

structed a bootstrap 95% confidence interval for the mean difference in wait times, which yielded the interval [-4154.47, -4078.52]. This interval further confirms the substantial reduction in wait time associated with the MARL controller.

C. Visual Representation of Results

To enhance the comprehension of the performance disparities observed between the different controller types, we present several visualizations that elucidate the data:

These figures effectively depict the distribution and variability of the performance metrics under consideration. Notably, they underscore the superior efficacy of the MARL controller, particularly in its capacity to minimize wait times for vehicles. Such visual representations provide critical insights into the operational advantages associated with adopting MARL frameworks in traffic management.

VI. Discussion

A. Interpretation of Results

The results of the comparative analysis between the fixed-time traffic light controller and the Multi-Agent Reinforcement Learning (MARL) controller highlight striking differences in performance metrics. Specifically, the MARL controller facilitated a greater average number of vehicles passed, achieving

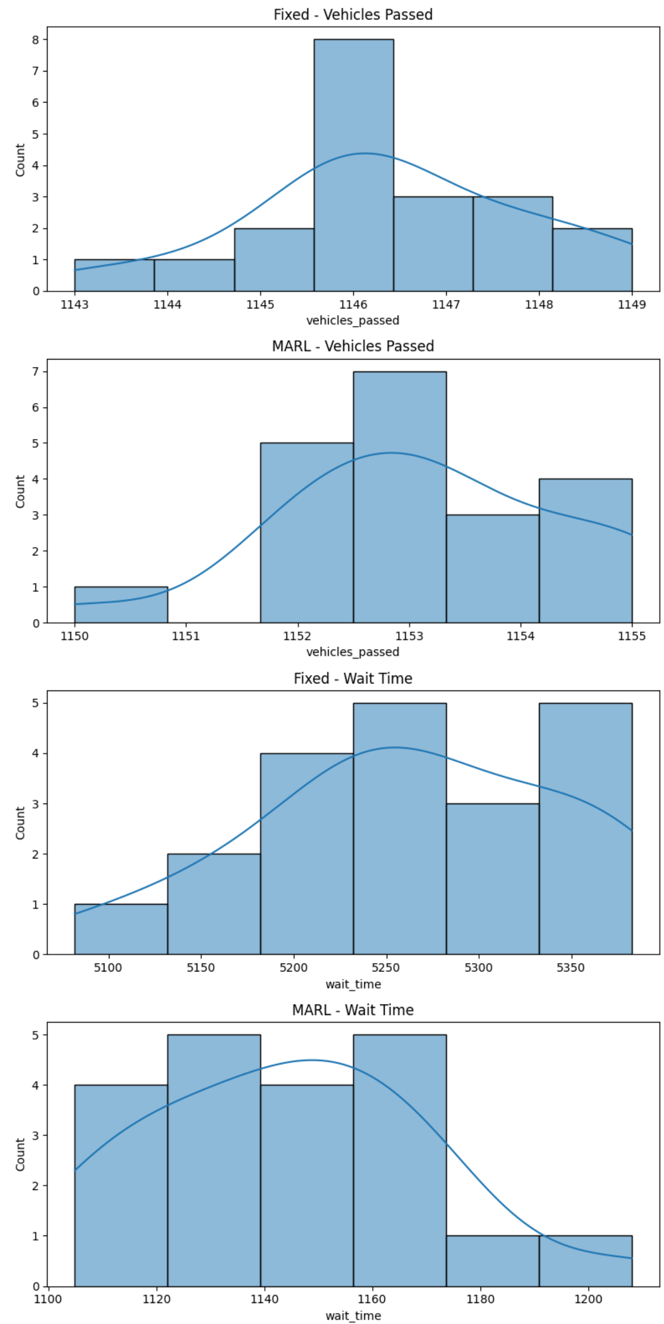

Fig. 5. Histograms Representing the Distribution of Vehicles Passed and Wait Times for Fixed-Time and MARL Controllers

a mean of 1153.15 vehicles compared to 1146.40 vehicles observed with the fixed-time controller. Statistical analysis using a t-test yielded a p-value of 1.64×10^{-17}, with a Cohen's d of 4.7311, thereby indicating a substantial effect size.

More significantly, the average wait time experienced by vehicles under the MARL controller was dramatically lower, recorded at 1144.77 seconds, in stark contrast to the 5263.82 seconds noted with the fixed-time controller. This substantial reduction in wait time achieves statistical significance, as evidenced by a t-test p-value of 8.59×10^{-39} and a Cohen's d of -66.1264. These findings underscore the superior efficiency of the MARL controller in minimizing vehicle wait times, presenting a compelling case for the adoption of learning-based approaches in traffic management.

B. Adaptability of MARL in Dynamic Traffic Conditions

The enhanced performance observed with the MARL controller can be attributed to its remarkable adaptability to dynamic traffic conditions. Unlike fixed-time controllers that operate on rigid, predetermined schedules, MARL agents enhance their policy decisions through continuous interaction with their environment. This capability allows them to respond effectively to real-time fluctuations in traffic patterns, a feature that is particularly crucial in urban settings where traffic distributions are inherently stochastic and subject to numerous unpredictable influences [6]. Such adaptability not only optimizes traffic flow but also aligns with the evolving demands of urban transportation systems.

C. Comparison with Related Work

The findings presented in this study resonate profoundly with the existing body of literature that underscores the effectiveness of Multi-Agent Reinforcement Learning (MARL) in the realm of traffic signal control. For instance, Chu et al. [6] illustrated that employing MARL methodologies can lead to a marked reduction in average vehicle delays when contrasted with traditional traffic management approaches. In a similar vein, Liu et al. [34] highlighted the critical role of adaptability in MARL systems, particularly in their capacity to manage the complexities of dynamic traffic scenarios. Our results not only corroborate these prior studies but also bolster the assertion that MARL embodies a powerful framework for traffic management within intricate urban environments.

D. Practical Implications for Real-World Traffic Management

The significant decrease in vehicle wait times, coupled with heightened throughput achieved through the MARL controller, suggests that there are considerable practical advantages to be gained from implementing MARL-based traffic signal control systems in real-world settings. Such implementations have the potential to enhance traffic flow, alleviate congestion, and ultimately improve the commuting experience for all users. Furthermore, the decentralized nature of MARL systems facilitates the development of scalable solutions that can be adapted to accommodate various urban layouts and varying traffic densities, making them a viable option for future urban traffic management strategies.

E. Limitations of the Study

While the findings of this study are promising, it is crucial to acknowledge several inherent limitations that could affect the interpretation of the results:

- **Simulation Environment**: The study was conducted within a controlled simulated environment, which may not fully encompass the complexities of real-world traffic systems. Factors such as the variability in driver behavior and unpredictable events were not considered, potentially impacting the external validity of the results.
- **Specific Road Network**: The particular road network utilized in the simulations may not be representative of all urban layouts. This specificity poses a challenge to the generalizability of the findings across different geographical and infrastructural contexts.
- **Simplified Vehicle Behavior**: The simulation employs a model that simplifies vehicle dynamics, which may not accurately mirror the complexities of real-world driving patterns. This simplification could lead to discrepancies between the simulation outcomes and real-world behaviors.
- **MARL Algorithm Constraints**: The performance of the Multi-Agent Reinforcement Learning (MARL) controller is heavily reliant on the selected algorithm and its associated hyperparameters. This reliance suggests that substantial tuning may be necessary to optimize performance across various scenarios.

F. Potential Influencing Factors

Several external factors may have influenced the results of this study, which warrant consideration:

- **Randomness in Vehicle Spawning**: The stochastic aspect of vehicle spawning within the simulation can introduce variability in traffic patterns. Such randomness may inadvertently affect the performance of the controller, leading to results that are not consistent across different runs.
- **Reward Function Design**: The design of the reward function used within the MARL framework is pivotal in shaping the behavior of the agents and ultimately the performance outcomes. Variations in reward structure could lead to divergent agent behaviors and effectiveness.
- **Training Hyperparameters**: The choice of training hyperparameters, including learning rates and exploration strategies, significantly impacts the learning efficiency and overall effectiveness of the MARL agents. Optimizing these parameters could enhance agent performance but requires careful exploration.

Future research should aim to mitigate these limitations by integrating more realistic traffic models, examining diverse urban layouts, and conducting empirical field tests. Such efforts would validate the applicability of MARL-based controllers in real-world settings, thus enhancing the robustness and applicability of the findings.

VII. Conclusion

This study conducted a comparative analysis between fixed-time traffic light controllers and Multi-Agent Reinforcement Learning (MARL) controllers within a Pygame-based simulation environment. The findings indicate that MARL controllers significantly outperform fixed-time systems in key performance metrics, including increased vehicle throughput and reduced average wait times. This enhanced performance is attributed to MARL's adaptive capabilities, allowing agents to learn optimal traffic signal timings through continuous interaction with dynamic traffic conditions. These results align with existing literature emphasizing the efficacy of MARL in traffic signal control. However, the study's limitations, such as the controlled simulation environment, the specific road network used, and the sensitivity of MARL performance to reward functions and training parameters may affect the generalizability of the findings. Future research should address these limitations by incorporating more realistic traffic models, exploring diverse urban layouts, and conducting empirical field tests to validate MARL-based controllers in real-world contexts. Additionally, integrating MARL with other intelligent transportation systems could further enhance urban traffic management strategies. Overall, this study provides compelling evidence supporting the adoption of MARL-based traffic signal controllers to improve urban traffic efficiency.

REFERENCES

[1] NACTO, "Fixed vs. actuated signalization," https://nacto.org/publication/urban-street-design-guide/intersection-design-elements/traffic-signals/fixed-vs-actuated-signalization/, n.d.

[2] Wikipedia contributors, "Split cycle offset optimisation technique," https://en.wikipedia.org/wiki/Split_Cycle_Offset_Optimisation_Technique, 2025.

[3] ——, "Sydney coordinated adaptive traffic system," https://en.wikipedia.org/wiki/Sydney_Coordinated_Adaptive_Traffic_System, 2025.

[4] ——, "Scalable urban traffic control," https://en.wikipedia.org/wiki/Scalable_Urban_Traffic_Control, 2025.

[5] C. Mims, "The smart, cheap fix for slow, dumb traffic lights," *WSJ*, 2024. [Online]. Available: https://www.wsj.com/tech/personal-tech/google-green-light-traffic-light-optimization-992e4252

[6] J. Wang, L. Codecà, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, 03 2019. [Online]. Available: https://www.researchgate.net/publication/331198452_Multi-Agent_Deep_Reinforcement_Learning_for_Large-Scale_Traffic_Signal_Control

[7] H. Z. Zhi Zhang, Jiachen Yang, "Integrating independent and centralized multi-agent reinforcement learning for traffic signal network optimization," *arXiv preprint arXiv:1909.10651*, 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1909.10651

[8] L. C. François-Xavier Devailly, Denis Larocque, "Ig-rl: Inductive graph reinforcement learning for massive-scale traffic signal control," *arXiv preprint arXiv:2003.05738*, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2003.05738

[9] Z. Li, H. Yu, G. Zhang, S. Dong, and C.-Z. Xu, "Network-wide traffic signal control optimization using a multi-agent deep reinforcement learning," *Transportation Research Part C: Emerging Technologies*, vol. 125, p. 103059, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X21000851

[10] Z. Zhang, W. Zhang, Y. Liu, and G. Xiong, "Mean field multi-agent reinforcement learning method for area traffic signal control," *Electronics*, vol. 12, p. 4686, 11 2023. [Online]. Available: https://www.mdpi.com/2079-9292/12/22/4686

[11] K. Othman, X. Wang, A. Shalaby, and B. Abdulhai, "Multimodal adaptive traffic signal control: A decentralized multiagent reinforcement learning approach," *Multimodal Transportation*, vol. 4, no. 1, p. 100190, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772586325000048

[12] M. Essa and T. Sayed, "Self-learning adaptive traffic signal control for real-time safety optimization," *Accident Analysis & Prevention*, vol. 146, p. 105713, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001457520305388

[13] C. Gheorghe and A. Soica, "Revolutionizing urban mobility: A systematic review of ai, iot, and predictive analytics in adaptive traffic control systems for road networks," *Electronics*, vol. 14, no. 4, 2025. [Online]. Available: https://www.mdpi.com/2079-9292/14/4/719

[14] S. Mahato, "Traffic light optimization," 2025. [Online]. Available: https://github.com/codesaahil/OptiTraffic

[15] P. Community, *Pygame v2.6.0 Documentation*, Pygame. [Online]. Available: https://www.pygame.org/docs/

[16] M. Gandhi, "Basic traffic intersection simulation," https://github.com/mihir-m-gandhi/Basic-Traffic-Intersection-Simulation.

[17] S. Tripathi, "Traffic optimisation with python!" https://medium.com/@sumit.tripathi/traffic-optimisation-with-python-be240d2fd863, 2023.

[18] F. H. Administration, *Traffic Signal Timing Manual*, U.S. Department of Transportation, https://ops.fhwa.dot.gov/publications/fhwahop08024/chapter5.htm.

[19] W. Su, C. Mu, L. Xue, X. Yang, and S. Zhu, "An improved traffic coordination control integrating traffic flow prediction and optimization," *Engineering Applications of Artificial Intelligence*, vol. 143, p. 109969, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197624021286

[20] Y. Liu, L. Liu, and W.-P. Chen, "Intelligent traffic light control using distributed multi-agent q learning," pp. 1–8, 10 2017. [Online]. Available: https://www.researchgate.net/publication/323786243_Intelligent_traffic_light_control_using_distributed_multi-agent_Q_learning

[21] T. Pan, "Traffic light control with reinforcement learning," 08 2023. [Online]. Available: https://www.researchgate.net/publication/373451408_Traffic_Light_Control_with_Reinforcement_Learning

[22] S. M. M. R. Swapno, S. Nobel, P. Meena, V. P. Meena, A. Azar, Z. Haider, and M. Tounsi, "A reinforcement learning approach for reducing traffic congestion using deep q learning," *Scientific Reports*, vol. 14, 12 2024. [Online]. Available: https://www.researchgate.net/publication/386987910_A_reinforcement_learning_approach_for_reducing_traffic_congestion_using_deep_Q_learning

[23] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, 2015. [Online]. Available: https://www.nature.com/articles/nature14236

[24] PTV Group, "Traffic simulation: Everything you need to know," https://www.ptvgroup.com/en-us/application-areas/traffic-simulation, 2023.

[25] R. D. Peters, "The application of simulation to traffic design and dispatcher testing," https://liftescalatorlibrary.org/paper_indexing/abstract_pages/00000049.html, 2013.

[26] T. R. Board, E. National Academies of Sciences, and Medicine, *Highway Capacity Manual 7th Edition: A Guide for Multimodal Mobility Analysis*. Washington, DC: The National Academies Press, 2022. [Online]. Available: https://nap.nationalacademies.org/catalog/26432/highway-capacity-manual-7th-edition-a-guide-for-multimodal-mobility

[27] I. Nizani, "Measuring traffic signal performance through digital twin technology," https://axilion.com/blog/measuring-traffic-signal-performance/, 2024.

[28] M. Coşkun, A. Baggag, and S. Chawla, "Deep reinforcement learning for traffic light optimization," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018, pp. 564–571.

[29] Student, "The probable error of a mean," *Biometrika*, vol. 6, no. 1, pp. 1–25, 1908. [Online]. Available: https://academic.oup.com/biomet/article-abstract/6/1/1/225634

[30] H. Mann and D. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.

[31] R. A. Fisher, *Statistical Methods for Research Workers*, S. Kotz and N. L. Johnson, Eds. New York, NY: Springer New York, 1992. [Online]. Available: https://doi.org/10.1007/978-1-4612-4380-9_6

[32] J. Cohen, *Statistical power analysis for the behavioral sciences*. Routledge, 1988. [Online]. Available: https://doi.org/10.1016/C2013-0-10517-X

[33] B. Efron and R. Tibshirani, *An introduction to the bootstrap*. CRC press, 1994. [Online]. Available: https://doi.org/10.1201/9780429246593

[34] W. Liu, C. Zhang, W. Fang, K. Zhou, Y. Li, F. Zhan, Q. Wang, W. Xue, and R. Chen, "Vehicle-level fairness-oriented constrained multi-agent reinforcement learning for adaptive traffic signal control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 4, pp. 4878–4890, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/10907776