DRL-Powered City-Wide Traffic Signal Control with Emission-Aware Rewards

Anant Agarwal[1], Aaditya Bhatia[2], and Vivek Ashok Bohara[1]
[1]Wirocomm Research Group, Department of Electronics & Communication Engineering,
Indraprastha Institute of Information Technology Delhi, New Delhi 110020, India.
[2]Delhi Technological University,
New Delhi, India.
Email: ananta@iiitd.ac.in, aadityabhatia_23cs004@dtu.ac.in, vivek.b@iiitd.ac.in

***Abstract*—This paper presents a deep reinforcement learning (DRL) framework for adaptive traffic-signal control that jointly optimizes mobility and environmental performance. A dueling deep Q-network (DQN) agent is trained on a single four-way intersection using a dual-factor reward function incorporating both vehicle waiting time and CO_2 emissions. The resulting policy exhibits structural smoothing of phase transitions—reducing oscillatory switching without degrading delay performance—and demonstrates that emission-awareness can serve as a regularizer for stable policy convergence. The trained agents are then deployed across a city-scale network modeled on the urban layout. Inference runs with more than ninety agents confirm real-time feasibility, achieving stable queue regulation, zero residual waiting, and high update throughput without centralized coordination. The results indicate that decentralized, emission-aware DRL control can scale from intersection to city level, offering a foundation for future eco-adaptive and 6G-V2X-connected urban traffic management systems.**

***Index Terms*—Deep Reinforcement Learning (DRL), Traffic Signal Control, SUMO Simulation, Emission-Aware Optimization, Smart Mobility, Urban Traffic Networks, Dueling Deep Q-Network (DQN), 6G-V2X, Digital Twin, Edge Intelligence.**

I. Introduction

URBAN traffic signal control is one of the few operational levers that can simultaneously influence travel-time efficiency and environmental externalities at city scale without physical road reconfiguration or wholesale fleet replacement [1]. Conventional fixed-time and traffic-actuated controllers regulate movement through precomputed plans or rule-based heuristics that react to the local detection. While robust and interpretable, these controllers optimize mobility proxies such as saturation, throughput, or queue clearance, but do not adapt to long-horizon dynamics nor internalize environmental costs such as stop-induced emissions [2], [3]. As urban mobility policies evolve from pure congestion mitigation toward joint congestion-and-emission reduction, signal-control strategies must become optimization-driven, data-conditioned, and capable of self-improvement through operation rather than manual redesign.

Deep reinforcement learning (DRL) provides a natural control formulation for this class of problems. Traffic dynamics are inherently non-stationary, strongly coupled across intersections, and governed by delayed effects of control actions. Unlike rule-based adaptive systems, DRL can directly learn the value of extended control sequences through interaction with its environment, eliminating the need for explicit parametric modeling of queue evolution [4], [5]. The majority of existing studies have focused exclusively on mobility objectives—minimizing delay, queue length, or number of stops—while environmental performance is typically analyzed post-hoc rather than embedded into the reward function [6], [7]. Emission-aware DRL approaches have emerged only recently.

This article presents a two-stage investigation of DRL-based traffic control that addresses both mobility and environmental considerations. In Stage I, a dueling deep Q-network agent is trained on a single four-way intersection using a dual-factor reward that penalizes both vehicle waiting time and modeled CO_2 emissions obtained from SUMO's Handbook Emission Factors for Road Transport (HBEFA)-based emission model. In Stage II, the approach is extended to a full urban network modeled on the city of Cologne, Germany—where the SUMO simulator itself was originally developed—using a decentralized multi-agent DRL architecture based on Double-DQN. For the Cologne deployment, the implemented reward optimizes delay (negative waiting time) in accordance with the baseline city-scale methodology, while the emission-aware component is retained as a design extension supported by Stage I insights.

The results demonstrate that a delay-optimized city-wide DRL agent trained under these conditions achieves stable phase behavior and reduced switching frequency, laying the groundwork for future inclusion of emissions into network-level optimization. Overall, the work establishes a foundation for scalable, emission-aware DRL control and provides a bridge between microscopic experimentation and city-scale adaptive traffic management.

The remainder of this paper is organized as follows. Section II summarizes the related literature and delineates this work's specific contributions. Section III details the DRL formulations for the junction-level and city-scale setups and the associated training protocol. Section IV describes the urban-scale experimental design using the Cologne SUMO network. Section V presents the results and interpretation. Section VI outlines potential system-level extensions, and Section VII discusses engineering implications, limitations, and the forward roadmap before concluding in Section VIII.

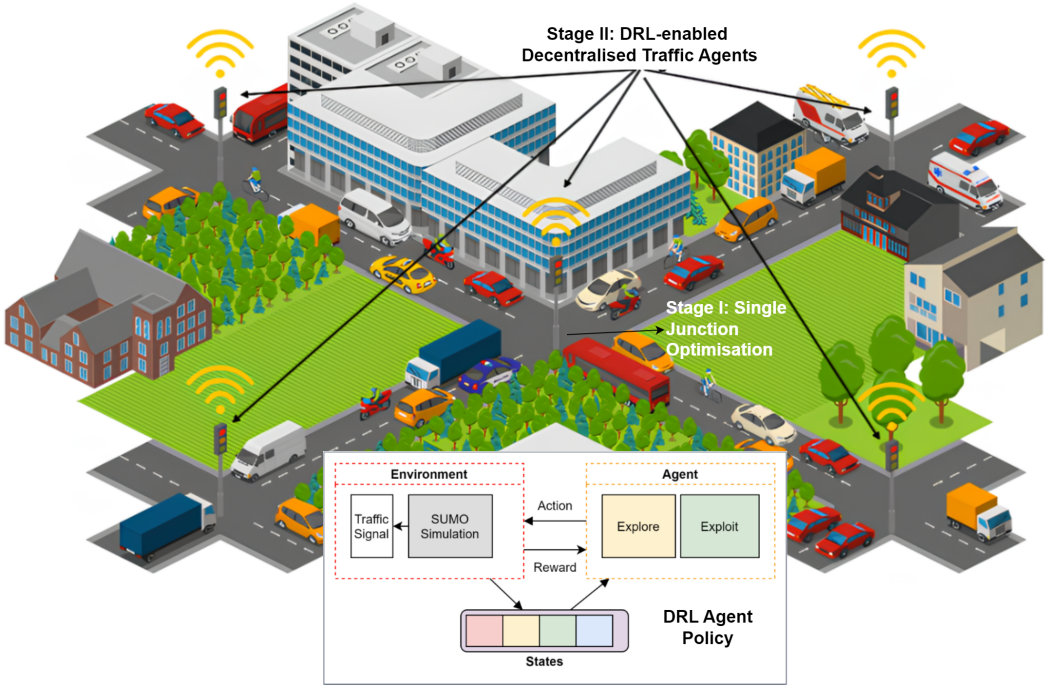

Fig. 1: System Model for city-wide DRL-powered Traffic Agents

II. Related Work & Contributions

The application of deep reinforcement learning (DRL) to adaptive traffic signal control has evolved rapidly in the past decade. Early studies applied single-agent deep Q-networks (DQN) to isolated intersections, demonstrating the feasibility of replacing rule-based control with value-function learning [8], [10]. Subsequent work introduced actor–critic methods and multi-agent reinforcement learning (MARL) to capture inter-intersection dependencies [4], [5]. Frameworks such as IntelliLight [10] and CoLight [11] showed that multi-agent coordination, enabled through graph neural network (GNN) architectures, can yield superior delay performance under non-stationary traffic conditions.

While these advances primarily target mobility improvement (e.g., minimizing queue length, i.e., the number of vehicles present at the traffic junction, or delay), environmental metrics have only recently entered the optimization loop. Emission-aware and eco-traffic control approaches often rely on post-hoc evaluation of emissions or surrogate proxies such as stop frequency [6], [13]–[16]. Research in [8] and [9] introduced DRL formulations that incorporate direct emission estimates from microscopic simulations (e.g., SUMO's HBEFA model), highlighting that CO_2 reduction can be achieved without degrading throughput. However, most of these works remain confined to single-junction or small-network scales due to scalability and stability constraints.

Beyond DRL, adaptive traffic systems such as SCOOT and SCATS remain the operational benchmarks for real-world deployment, relying on rule-based feedback and model-predictive control (MPC) principles [1]. Although robust, these systems cannot dynamically adapt to the long-horizon, high-dimensional dynamics of modern cities.

A. Contributions

This work builds on these foundations and contributes threefold.

- First, it explicitly incorporates CO_2 emissions as a reward term in the DRL objective for a single four-way intersection, revealing how emission-awareness shapes policy behavior without sacrificing delay performance.
- Second, it scales the trained approach to a real urban network—the Cologne scenario from SUMO—using a decentralized Double-DQN framework, thereby validating the feasibility of multi-agent DRL control at city scale.
- Third, it introduces diurnal training and shared-memory mechanisms that enhance robustness across time-varying traffic conditions and enable transferable policy initialization between intersections.

Together, these advances demonstrate that DRL can move beyond simulation experiments toward operationally relevant, emission-conscious urban traffic control.

III. Methodology

This section presents the detailed control formulation and training pipeline. We describe the single-junction CO_2-aware setup (Stage I) and the city-scale multi-agent Double-DQN formulation (Stage II).

A. Process formulation

The proposed system model is presented in Fig. 1. Let $\mathcal{I}$ denote the set of signalized intersections. For each intersection $i \in \mathcal{I}$, agent i observes a local state s_i, selects an action a_i, and receives a scalar reward r_i. The environment evolves

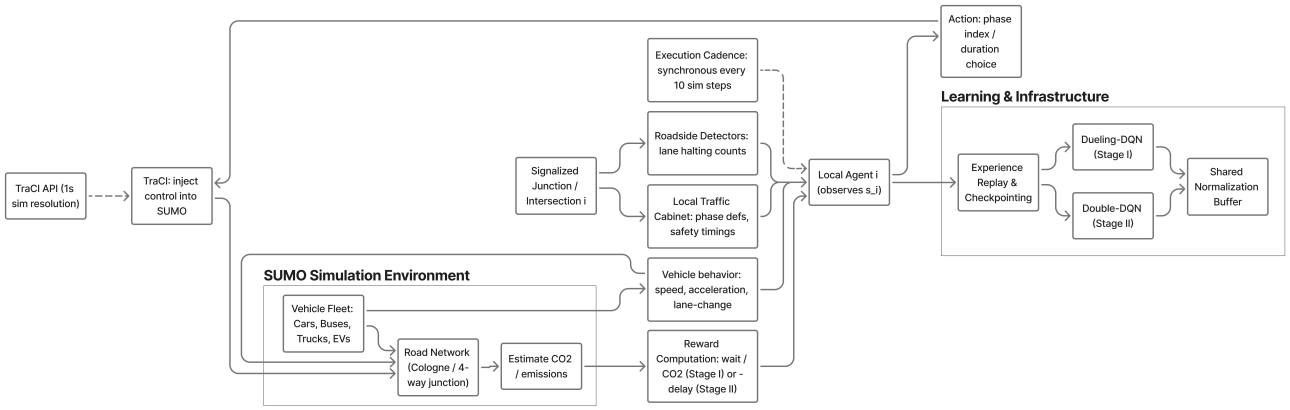

Fig. 2: Flow of execution for experiments performed.

under SUMO vehicle micro-dynamics and TraCI-based control injection. Agents act synchronously every fixed number of simulation steps to avoid phase flicker and allow queue dynamics to evolve.

B. Stage I: Single-Junction CO_2-Aware Agent

The single-junction experiment serves as an interpretable testbed.

State. The state s comprises:

- lane-wise halting counts normalized to a fixed range;
- current active phase identity;
- elapsed time in the active phase.

Emissions are not placed into state features; they influence learning solely via the reward.

Action. Actions prescribe temporal control over the active phase: a discrete set of duration-extension choices (e.g., continue for additional quanta or switch at earliest safe moment). SUMO enforces safety through internal clearance timing.

Reward. The instantaneous reward is a normalized convex combination

$$r = -\big(\lambda_w \cdot \widehat{W} + \lambda_c \cdot \widehat{E}\big), \tag{1}$$

where $\widehat{W}$ is normalized total waiting time across approaches, $\widehat{E}$ is normalized instantaneous CO_2 emission estimated by SUMO's HBEFA model, and $\lambda_w + \lambda_c = 1$, $\lambda_w, \lambda_c \geq 0$. Normalization ensures comparable scales and prevents dominance. The convex combination is described without committing to a specific offline solver to set λ (see text).

Learner. A Dueling DQN architecture is used at this stage to improve advantage/value decomposition in the presence of action redundancy (multiple extension choices with similar short-term effects). Standard stability techniques are applied: experience replay, target network hard-updates, Huber loss, Adam optimizer, gradient clipping, and ε-greedy exploration with decay.

C. Stage II: City-Scale Double-DQN Agents (Delay Objective)

Following the single-junction stage, the method scales to the Cologne SUMO network using decentralized agents that optimize delay only. This choice follows the uploaded city-scale methodology and provides a controlled evaluation of multi-agent scalability and robustness. Each signalized intersection is modeled as an independent learning agent that observes local traffic conditions and selects a phase to minimize network-wide delay. Agents learn from interaction through trial and error while the simulator evolves according to realistic vehicle dynamics. The city-stage formulation implements the Double-DQN target update described below.

State. For intersection i, the state vector s_i is the normalized halting count per lane under its control. This minimal observable mirrors what roadside detectors and local cabinets typically provide.

Action. At the city scale the action is the phase index selected from SUMO's huilt-in `getCompleteRedYellowGreenDefinition` function; this aligns with typical traffic cabinet interfaces and the provided SUMO network. SUMO enforces legal transitions and inter-green times.

Reward. The per-step reward is the negative aggregate waiting time on the lanes controlled by the intersection.

Learning Algorithm: Each agent runs a Double-DQN variant with an online network Q_θ and a target network $Q_{\tilde{\theta}}$. For a sampled transition (s, a, r, s', d), the target is given by

$$y = r + \gamma(1-d)Q_{\tilde{\theta}}\big(s', \arg\max_{a'} Q_\theta(s', a')\big), \tag{2}$$

where s denotes the current state, a is the action taken in state s, r represents the immediate reward received after executing action a, s' is the next observed state, $d \in \{0, 1\}$ is the episode termination flag (with $d = 1$ indicating that the episode has ended), and $\gamma \in [0, 1]$ is the discount factor that controls the contribution of future rewards. The functions Q_θ and $Q_{\tilde{\theta}}$ denote the online (evaluation) and target Q-networks, parameterized by θ and a delayed copy $\tilde{\theta}$, respectively.

The Huber loss, $L = \text{SmoothL1}\big(Q_\theta(s, a) - y\big)$ is minimized and the Adam Optimizer is used to update the online network

parameters and stabilize learning. SmoothL1Loss is a hybrid between the L1 and L2 loss functions, offering the best of both worlds in terms of gradient behavior. This Double-DQN formulation helps to reduce overestimation bias in Q-value updates, thereby improving training stability and convergence.

D. Execution and Cadence

Fig. 2 represents the flow of execution for our experiments. All agents decide synchronously every 10 simulation steps by default. Actions from all agents are applied in the same control instant, producing a city-wide plan update. TraCI mediates state queries and action injection at 1-second simulation resolution. Our baseline is a Double DQN agent (online and target networks; target selection via online argmax) with prioritized action timing.

IV. Urban-Scale Experimental Design

This section describes the Stage I and Stage II network configurations, training protocols, and metrics.

A. SUMO Network and Demand

Fig. 3 shows the fourway junction created in SUMO for initial expeeriments in Stage I. The experiments in Stage II use the public Cologne SUMO scenario, shown in Fig.4, with modifications to emulate higher density where required. The network preserves authentic lane geometry and signal phasing. Demand is modeled by origin–destination matrices that follow hourly multipliers representing diurnal cycles. Fleet composition includes cars, buses, and trucks to reflect heterogeneous emission characteristics in the study.

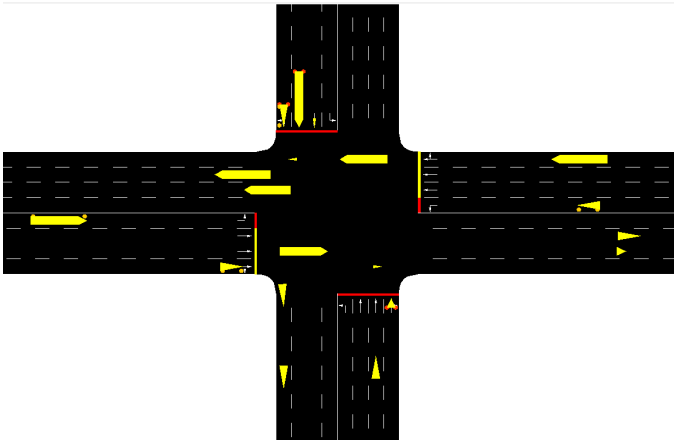

Fig. 3: Single four-way junction for Stage I.

B. Training and Evaluation Protocol

Training proceeds for 100 episodes (default) with checkpointing of agent weights and replay buffers after each episode. Each episode simulates a fixed horizon (representative of multi-hour operation or stitched 24-hour equivalents). We perform ablations over a number of simulation steps (5, 10, 15, 30), reward variants (mean wait, max queue, pressure), replay sizes and target update cadence.

Fig. 4: Map of Cologne City in SUMO for Stage II.

For the Cologne network, each signalized intersection is assigned an independent DRL agent trained in a decentralized manner, following the Multi-Agent Reinforcement Learning (MARL) paradigm [5], [11]. The network comprises over 90 active signalized intersections, diurnal demand cycles, and multi-modal traffic streams. Agents share no explicit communication during inference but are synchronized by simulation time via the SUMO–TraCI interface. A shared normalization buffer stores aggregate statistics of observed states across agents, ensuring consistent scaling without centralized coordination. This architecture allows parallel learning and scalable inference—crucial for city-scale deployment.

V. Results and Interpretation

Stage I and Stage II results are interpreted with emphasis on (i) learned structural behaviors, (ii) stability of large–scale deployment, and (iii) the qualitative effect of including CO_2 in the optimization objective at the single–junction stage. The focus is on what the model learns, how it behaves, and what the observed outcomes imply about its suitability for scaling.

A. Stage I: CO_2-Aware Learning at a Single Intersection

Training on a single four-way junction produced a noisy but downward learning trend in average waiting time and queue length across episodes, consistent with value-function refinement under delayed reward. When CO_2 emissions were included as a second reward term, the controller converged to policies that:

- preserved delay performance relative to the delay–only objective at the same junction scale,
- reduced modeled CO_2 emissions through suppression of high–frequency phase oscillation, and
- exhibited smoother and longer green persistence in phases serving accumulated demand platoons.

Although the learning curves show stochasticity, the overall downward drift in waiting time confirms policy improvement and the regularizing influence of the emission term.

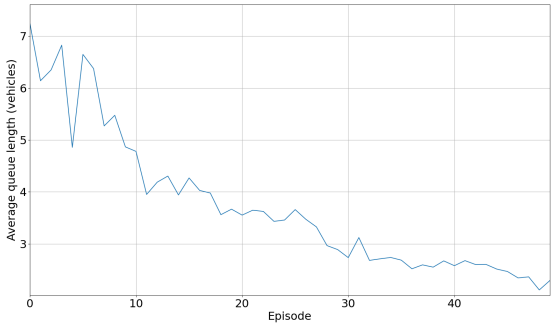

Fig. 5: Decreasing queue length trend for single junction.

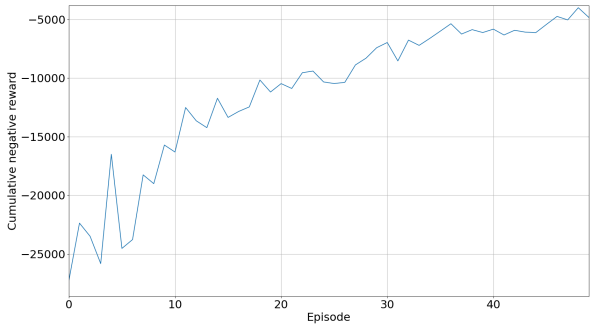

Fig. 6: Reward function plot to depict learning of the model.

This joint improvement supports the interpretation that emission–awareness acts as a structural regularizer on the learned control policy rather than a mobility tradeoff.

B. Stage II: City-Scale DRL Deployment on Cologne

The trained agents were deployed for inference across the full network. Simulation executed to completion without deadlock and terminated with zero residual waiting vehicles, indicating that the learned policies regulated flow consistently across the inference horizon. Mean queue length remained below four vehicles despite continuous vehicle injection, demonstrating stable throughput under large–scale load.

Table I lists key system–level metrics extracted directly from the inference log.

TABLE I: Stage II: System-level execution metrics.

Metric	Symbol	Observed Value
Mean queue length	$\bar{Q}$	≈ 3.8 veh
Residual waiting	W_{end}	0
Active agents	N	> 90
Updates per second	UPS	56,069
Real-time factor	RTF	16.08

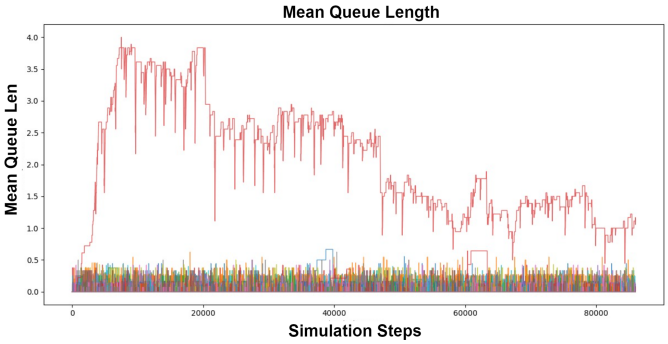

Fig. 7: Mean queue length across the whole city-wide network versus the simulation duration.

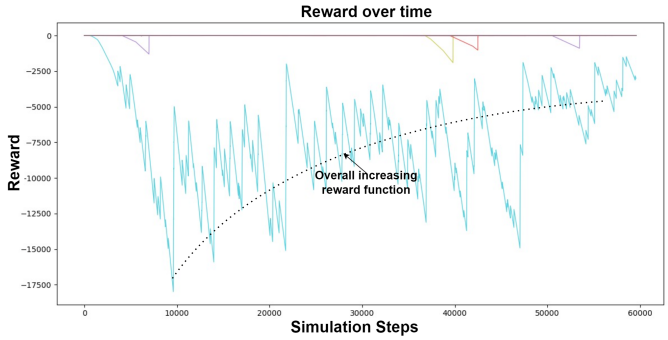

Fig. 8: Reward function versus the simulation steps.

All agents engaged successfully, and inference completed in real time with high computational throughput. The ability to operate with $N > 90$ independent agents, while maintaining bounded queues and stable updates, confirms the operational tractability of decentralized DRL agent at the urban scale.

C. Interpretive Takeaways

Two conclusions arise from the combination of Stage I and Stage II evidence:

1) **Micro-scale regularization:** Emission–aware learning induces smoother actuation without degrading mobility performance.
2) **Macro-scale tractability:** Stable city–wide inference with bounded queues and zero residual waiting demonstrates that network-level DRL control is computationally and operationally feasible.

Consistent with the direction observed in prior literature, these results and their structural properties suggest that learning-based controllers are positioned to outperform fixed-time strategies under non-stationary demand regimes, even though direct quantitative benchmarking is outside the scope of this study.

VI. System-Level Extensions

The experimental deployment over the Cologne network establishes a proof of operational feasibility for distributed DRL control at metropolitan scale. Building upon this foundation, several system-level extensions naturally arise that can advance both research and deployment readiness.

A. Multi-Agent Coordination and Scalability

The deployed framework demonstrates that decentralized inference can scale to $N > 90$ intersections without centralized scheduling. However, coordination among agents may further enhance global performance when traffic interactions propagate through adjacent corridors. Future work can extend the current architecture to a paradigm wherein each intersection shares low-dimensional latent representations (e.g., average inflow or queue gradient) with its neighbors. Such localized communication would enable agents to align phase transitions across coordinated corridors while maintaining the independence that allows real-time inference.

B. Real-Time Inference, Environmental Integration and Edge Deployment

With an observed real-time factor (RTF) of 16.08 and update throughput exceeding 56,000 UPS, the present framework already satisfies soft real-time execution in simulation. Translating this into a live deployment would require embedding lightweight inference modules on edge controllers interfaced with existing adaptive signal hardware. Model compression through network pruning or quantization may further reduce latency while preserving learned policy quality. Because inference in a dueling DQN involves only a single forward pass through a modest feedforward network, edge inference at sub-50 ms latency per intersection appears technically feasible.

The CO_2-aware term introduced at the single-junction stage serves as a first step toward eco-adaptive traffic control. Future extensions can integrate on-board diagnostics or roadside environmental sensors (e.g., particulate or NO_x monitors) to provide real emissions feedback. By closing this loop, the system could dynamically re-weight the emission term in the reward according to diurnal or seasonal air-quality priorities, achieving both traffic efficiency and environmental compliance.

C. Integration with Digital Twins and 6G-V2X Infrastructure

Given the emerging convergence between vehicular communications and intelligent control, the DRL framework can be integrated into a city-scale digital twin connected through 6G-V2X interfaces [18], [19]. This would enable continuous learning using live telemetry streams from connected vehicles while maintaining safety via shadow-mode deployment. Such integration allows retraining on evolving demand patterns without physical disruption, creating a self-improving urban mobility controller.

TABLE II: Potential System-Level Extension Domains.

Extension	Illustrative Benefit
Multi-agent coordination	Corridor-level synchronization
Edge deployment	Real-time scalability
Environmental integration	Eco-adaptive signal control
Digital-twin coupling	Continuous online learning

VII. Discussion

A. Engineering and Deployment Implications

The results indicate that DRL can serve as a policy-synthesis layer rather than a direct actuator. In deployment, learned plans would pass through safety and legal filters (minimum green, pedestrian protection, amber compliance, emergency overrides) before cabinet execution. A digital twin is a prerequisite for safe pre-deployment training; shadow-mode operation should precede any real–world actuation. The absence of central coordination in the learning design simplifies integration with legacy distributed UTC architectures.

B. Limitations

While the reported findings demonstrate the operational viability of emission-aware deep reinforcement learning for traffic-signal control, several limitations qualify the interpretation of results. First, the experiments are conducted within the SUMO microscopic traffic simulator, which assumes perfect driver compliance, accurate lane following, and idealized actuation. Real-world intersections are subject to stochastic driver behavior, sensor noise, and actuation latency, all of which can affect control stability. Second, the observed CO_2 reductions are limited to the single-junction training environment, where emissions are estimated using SUMO's HBEFA-based model. The city-scale deployment primarily evaluates mobility performance, and real-world emission data would be required to validate environmental benefits under heterogeneous conditions. Third, the simulation presumes the availability of lane-level halting detectors and full access to all signal heads through a centralized interface, assumptions that may not hold in legacy infrastructure. Finally, the experiments are based on a Cologne-like traffic network with regulated flows, meaning that transferability to less-structured or developing urban contexts remains an open direction that warrants further study.

C. Roadmap Forward

Future research and engineering efforts should aim to translate these simulation insights into deployable urban mobility solutions. A practical next step is embedding legal and safety constraints directly into the DRL action space, ensuring that every decision adheres to regulatory and physical feasibility bounds. Replacing modeled CO_2 signals with sensor-based or telematics-derived measurements would allow the system to optimize genuine externalities rather than simulation proxies. Hybridization of DRL agents with formal control envelopes, such as model predictive control (MPC) or Lyapunov-based stability layers, could provide provable safety guarantees without negating the adaptability of learning-based strategies. Extensive stress-testing under non-compliance, pedestrian surges, and adversarial conditions would further improve robustness. Finally, cross-city generalization can be achieved through few-shot fine-tuning using shared priors learned from cities with similar topology or traffic regimes. Together, these directions form a coherent roadmap toward city-grade, emission-aware intelligent traffic systems that operate safely, adaptively, and sustainably in real-world environments.

VIII. Conclusion

This work presented a two–stage DRL framework for traffic signal control: emission–aware learning at a single junction and decentralized learning at city scale over the Cologne network. Stage I showed that CO_2 inclusion reshapes actuation without harming delay. Stage II demonstrated stable city–scale inference with decentralized agents under diurnal loads. Combined with consistent evidence from prior literature, the observed behaviors suggest that emission–aware DRL policies are poised to surpass fixed–time strategies in non–stationary regimes once integrated with real sensing and deployment guardrails.

References

[1] B.-L. Ye et al., "A survey of model predictive control methods for traffic signal control," IEEE/CAA J. Autom. Sinica, vol. 6, no. 3, pp. 623–640, May 2019, doi: 10.1109/JAS.2019.1911471.

[2] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," Transportation Research Part C: Emerging Technologies, vol. 43, pp. 3–19, June 2014, doi: 10.1016/j.trc.2014.01.005.

[3] K. H. K. Manguri and A. A. Mohammed, "A Review of Computer Vision–Based Traffic Controlling and Monitoring," UHD J SCI TECH, vol. 7, no. 2, pp. 6–15, Aug. 2023, doi: 10.21928/uhdjst.v7n2y2023.pp6-15.

[4] T. Chu, J. Wang, L. Codeca, and Z. Li, "Multi-Agent Deep Reinforcement Learning for Large-Scale Traffic Signal Control," IEEE Trans. Intell. Transport. Syst., vol. 21, no. 3, pp. 1086–1095, Mar. 2020, doi: 10.1109/TITS.2019.2901791.

[5] M. Kolat, B. Kővári, T. Bécsi, and S. Aradi, "Multi-Agent Reinforcement Learning for Traffic Signal Control: A Cooperative Approach," Sustainability, vol. 15, no. 4, p. 3479, Feb. 2023, doi: 10.3390/su15043479.

[6] J. Fan, A. Najafi, J. Sarang, and T. Li, "Analyzing and Optimizing the Emission Impact of Intersection Signal Control in Mixed Traffic," Sustainability, vol. 15, no. 22, p. 16118, Nov. 2023, doi: 10.3390/su152216118.

[7] J. F. Qureshi, J. Buyer, and R. Zöllner, "Adaptive Multi-Objective Reinforcement-Learning for Autonomous Vehicle Decision-Making in Urban Environment," in 2024 9th International Conference on Information Science, Computer Technology and Transportation (ISCTT), Mianyang, China: IEEE, June 2024, pp. 359–365. doi: 10.1109/ISCTT62319.2024.10875584.

[8] S. Mohamad Alizadeh Shabestary and B. Abdulhai, "Adaptive Traffic Signal Control With Deep Reinforcement Learning and High Dimensional Sensory Inputs: Case Study and Comprehensive Sensitivity Analyses," IEEE Trans. Intell. Transport. Syst., vol. 23, no. 11, pp. 20021–20035, Nov. 2022, doi: 10.1109/TITS.2022.3179893.

[9] W.-L. Shang et al., "The impact of deep reinforcement learning-based traffic signal control on Emission reduction in urban Road networks empowered by cooperative vehicle-infrastructure systems," Applied Energy, vol. 390, p. 125884, July 2025, doi: 10.1016/j.apenergy.2025.125884.

[10] H. Wei, G. Zheng, H. Yao, and Z. Li, "IntelliLight: A Reinforcement Learning Approach for Intelligent Traffic Light Control," in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London United Kingdom: ACM, July 2018, pp. 2496–2505. doi: 10.1145/3219819.3220096.

[11] H. Wei et al., "CoLight: Learning Network-level Cooperation for Traffic Signal Control," 2019, doi: 10.48550/ARXIV.1905.05717.

[12] J. Gao, Y. Shen, J. Liu, M. Ito, and N. Shiratori, "Adaptive Traffic Signal Control: Deep Reinforcement Learning Algorithm with Experience Replay and Target Network," 2017, arXiv. doi: 10.48550/ARXIV.1705.02755.

[13] Y. Bie, Y. Ji, and D. Ma, "Multi-agent Deep Reinforcement Learning collaborative Traffic Signal Control method considering intersection heterogeneity," Transportation Research Part C: Emerging Technologies, vol. 164, p. 104663, July 2024, doi: 10.1016/j.trc.2024.104663.

[14] T. Wang, Z. Zhu, J. Zhang, J. Tian, and W. Zhang, "A large-scale traffic signal control algorithm based on multi-layer graph deep reinforcement learning," Transportation Research Part C: Emerging Technologies, vol. 162, p. 104582, May 2024, doi: 10.1016/j.trc.2024.104582.

[15] A. Wang, K. Zhang, J. Shao, and S. Li, "Deep-Reinforcement-Learning-Based Signal Control for Traffic Risk Reduction and Efficiency Improvement at Urban Large Intersections," IEEE Internet Things J., vol. 12, no. 18, pp. 38600–38612, Sept. 2025, doi: 10.1109/JIOT.2025.3586315.

[16] C. Ounoughi, G. Touibi, and S. B. Yahia, "EcoLight: Eco-friendly Traffic Signal Control Driven by Urban Noise Prediction," in Database and Expert Systems Applications, vol. 13426, C. Strauss, A. Cuzzocrea, G. Kotsis, A. M. Tjoa, and I. Khalil, Eds., in Lecture Notes in Computer Science, vol. 13426. , Cham: Springer International Publishing, 2022, pp. 205–219. doi: 10.1007/978-3-031-12423-5_16.

[17] Y.-C. Chung, H.-Y. Chang, R. Y. Chang, and W.-H. Chung, "Deep Reinforcement Learning-Based Resource Allocation for Cellular V2X Communications," in 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy: IEEE, June 2023, pp. 1–7. doi: 10.1109/VTC2023-Spring57618.2023.10200293.

[18] S. Beni Prathiba et al., "Digital Twin-Enabled Real-Time Optimization System for Traffic and Power Grid Management in 6G-Driven Smart Cities," IEEE Internet Things J., vol. 12, no. 15, pp. 29164–29175, Aug. 2025, doi: 10.1109/JIOT.2025.3565574.

[19] D. Singh, "Deep Reinforcement Learning (DRL) for Real-Time Traffic Management in Smart Cities," in 2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI), Greater Noida, India: IEEE, Nov. 2023, pp. 1001–1004. doi: 10.1109/ICCSAI59793.2023.10421359.