

Article

Multi-Agent Deep Deterministic Policy Gradient-Based Coordinated Control for Urban Expressway Entrance–Arterial Interfaces

Shunchao Wang ¹, Zhigang Wu ^{2,*} and Wangzi Yu ¹

¹ School of Transportation, Southeast University, Nanjing 210096, China; wangshunchao@seu.edu.cn (S.W.); 220224724@seu.edu.cn (W.Y.)

² School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen 518107, China

* Correspondence: wuzhg6@mail2.sysu.edu.cn

Abstract

Coordinated control of ramp metering, variable speed limits, and intersection signals is critical for mitigating congestion and enhancing efficiency at urban expressway–arterial interfaces. Existing strategies often operate in isolation, leading to fragmented responses and limited adaptability under heterogeneous traffic demands. This study develops a multi-agent reinforcement learning framework based on MADDPG to achieve cooperative decision-making across heterogeneous controllers. An asynchronous control cycle mechanism is designed to accommodate different temporal requirements of ramp meters, speed limits, and signal controllers, ensuring practical feasibility in real-time operations. A conflict-aware reward design further embeds density regulation, speed harmonization, and spillback prevention to stabilize flow dynamics. Simulation experiments on a calibrated urban network demonstrate that the proposed framework delays congestion onset, reduces shockwave propagation, and improves throughput compared with classical benchmarks. In particular, at the mainline merge, average travel time is reduced to 13.56 s (62.4% of VSL-only); at the ramp, occupancy is lowered to 6.4% (40.6% of ALINEA); and at the signalized approach, average delay decreases to 85.71 s (62.7% of actuated control). These results highlight the scalability and deployment potential of the proposed cooperative control approach for system-level traffic management in mixed traffic environments.

Keywords: multi-agent reinforcement learning; system-level traffic management; urban expressway–arterial interfaces; asynchronous control mechanisms; intelligent transportation systems

Received: 21 August 2025

Revised: 7 January 2026

Accepted: 12 February 2026

Published: 25 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In urban traffic systems, interface zones centered on expressway on-ramps and adjacent surface arterials mediate the transition of flows across facility classes [1–6]. These interfaces exhibit compounded heterogeneity along two axes: (i) speed gradients between road classes readily generate bottlenecks and capacity drops; and (ii) coupled effects of complex alignment and control elements constrain network performance. Consequently, developing coordinated control for the expressway–arterial interface is pivotal to stabilizing flows and unlocking network efficiency.

Research on facility-level control progressed from fixed-time and feedback schemes to optimization and learning. For ramp metering, fixed-time linear programs and ALINEA improved robustness but were limited under strong demand variability [7,8]. For variable speed Variable Speed Limits (VSL), rule-based designs evolved into Model Predictive Control (MPC) formulations grounded in traffic flow models, with applications to mixed traffic and multi-bottleneck corridors [9–21]. The development of reinforcement learning (RL) and multi-agent reinforcement learning (MARL) [22,23] further relaxes modeling fidelity by learning from interaction: on ramps, with VSL and at intersections, where RL-based controllers integrate macro–micro states and reduce delay while improving transit reliability [24,25].

At the expressway–arterial interface, studies increasingly coordinate multiple actuators. Coupled-node and cell transmission model (CTM) models synchronize ramp releases with adjacent signals to suppress spillback, while bi-level formulations co-optimize signal plans and ramp throughput [26,27]. A complementary stream integrates RM with VSL to align inflows with near-critical densities: deep RL adapts to traffic heterogeneity [28]; macroscopic fundamental diagram (MFD)-based coordinated ramp metering (CRM) fuses VSL and ramp metering (RM) to balance network flow and avoid bottleneck constraints [29]; MPC with METANET minimizes system travel time while controlling multiple bottlenecks [30]. These advances motivate unified, data-driven coordination.

Despite notable progress, several gaps constrain the effectiveness and deployability of interface-level coordination:

- Control-unit scope remains narrow: Most studies use single strategies or RM-signal RM-VSL pairs, overlook dual upstream influences, and lack a unified, dynamic scheme jointly optimizing VSL, RM, and intersection signals for the interface.
- Coupling mechanisms are suboptimal: Coordination often relies on fixed rules or one-way feedback: MPC/GA need accurate models, heavy calibration, and nontrivial computation-hindering, real-time, and bidirectional coupling under stochastic demand, mixed traffic, and partial observability.
- Robustness to phase transitions is insufficient: Single-point or weakly coordinated control struggles with capacity drops and instability; multi-agent decisions conflict without explicit safety/stability mediation, degrading throughput and elevating risk.

To fill these gaps, we develop a multi-agent framework based on MADDPG for the expressway–arterial interface. Three agents—VSL, RM, and signals—act on local observations with shared parameters and asynchronous control cycles. We design compact state/action spaces and multi-objective rewards (near-critical density tracking, speed-difference moderation, spillback safeguards), build a finely calibrated microscopic platform, and train cooperative policies. Comparative experiments against classical baselines use trajectory visualization, action-sequence diagnostics, and standard efficiency/safety metrics to demonstrate real-time feasibility, stability, and network-level benefits. The main contributions are as follows. In this study, “advancement” is defined in terms of system-level gains enabled by interface-level coordination under realistic operational constraints, rather than solely algorithm-level comparisons.

- A multi-agent architecture is formulated to jointly optimize variable speed limits, ramp metering, and intersection signals via MADDPG while representing dual upstream influences. The design resolves local conflicts, strengthens density regulation, and enhances control effectiveness under time-varying mixed conditions.
- A bidirectional coupling mechanism is proposed that consolidates local observations into a joint state and coordinates asynchronous control cycles through time-slice alignment and action holding. Unlike prior rule-based or MPC/GA approaches that

rely on heavy calibration, this mechanism enables threshold-free real-time coordination while preserving feasibility.

- A multi-region coordination mechanism is proposed to achieve cooperative optimization across different spatial and temporal domains, ensuring globally optimal traffic performance and overcoming the local optima, uncoordinated timings, and cascading breakdowns inherent in isolated control.

Section 2 reviews related work; Section 3 states the problem; Section 4 details the MADDPG-based coordinated control strategy; Section 5 presents experiments and results; Section 6 concludes and outlines future work.

2. Literature Review

2.1. Applications of Reinforcement Learning at Intersections

The application of reinforcement learning has become a dominant trend across various domains of traffic control, including ramp metering, VSL, and intersection signal timing. Early ramp metering methods were dominated by rule-based and feedback strategies, such as fixed-time metering [7] and the feedback-based ALINEA algorithm [8]. While these methods improved robustness in dynamic environments, they remained limited in adaptability. With the advent of artificial intelligence, RL-based ramp metering has emerged as a promising direction. For instance, Deng et al. [31] proposed a MARL framework, enabling adaptive and cooperative ramp metering with significant improvements in expressway efficiency.

Similarly, VSL control evolved from rule-based systems with limited resilience [11–13,32] to MPC-based optimization methods, which use traffic flow models to derive optimal strategies [15–20]. Although MPC approaches demonstrated effectiveness in shock-wave dissipation and performance optimization under mixed traffic, they rely on accurate modeling [18–20]. Reinforcement learning provides a model-free alternative, directly learning from traffic environment interactions. Recent works have applied multi-objective deep reinforcement learning frameworks for tunnel VSL control, enhancing traffic safety and stability in complex scenarios [32].

At urban intersections, RL has also been increasingly adopted to enhance adaptive signal control. Traditional adaptive methods struggled to capture stochastic demand variations and multimodal priorities, whereas RL-based controllers dynamically adjust signal phases by interacting with traffic states. Chow et al., for example, developed an RL-driven adaptive controller that integrates both macroscopic traffic flow variables and microscopic vehicle conditions, significantly reducing delays while improving public transit reliability [25].

Collectively, these studies demonstrate a clear trajectory: from traditional rule-based and feedback strategies to MPC optimization and ultimately to reinforcement learning approaches. RL has proven effective in ramp metering, VSL, and intersection signal control, providing superior adaptability, responsiveness, and robustness in heterogeneous and dynamic traffic environments.

2.2. Coordinated Control at Expressway–Urban Road Interfaces

The integration of expressway ramps and adjacent urban intersections has become a focal point in recent years, as this interface directly influences congestion spillback and overall network efficiency. One research stream focuses on ramp–intersection coordination, where traffic dynamics are jointly modeled and optimized. For example, Pang & Yang [26] developed a coupled-node model using an improved cell transmission model (CTM) to propose a synchronized fixed-time coordination method, effectively suppressing shockwave propagation and mitigating spillback. Deng et al. [27] further introduced

a bi-level optimization model that jointly optimizes signal timing and ramp metering, enhancing ramp throughput and minimizing average delays and stops, thereby alleviating upstream mainline congestion.

Another line of work emphasizes ramp–VSL coordination, aiming to balance ramp inflows with mainline stability through integrated control. Cheng et al. [28] proposed a deep reinforcement learning-based framework that dynamically adjusts ramp release rates and mainline speeds to maintain near-critical density conditions. He et al. [29] developed a coordinated ramp metering (CRM) strategy within the macroscopic fundamental diagram (MFD) framework, integrating VSL and RM to achieve “network-level flow balancing, bottleneck-level constraint avoidance, and flexible control extension,” effectively improving stability under heterogeneous density conditions. Similarly, Hegyi et al. [30] introduced an MPC-based coordination framework with the METANET model, integrating VSL and ramp metering to minimize system-wide travel times while suppressing congestion.

In summary, coordinated control at expressway–urban road interfaces reflects a methodological shift from local, isolated interventions to integrated, network-oriented solutions. By coupling ramp metering, VSL, and signal control, these approaches highlight the potential for establishing a more resilient and self-organizing traffic management paradigm. Table 1 summarizes the comparison between our work and existing studies.

Table 1. Comparison of recent studies.

Ref	Scope	Paradigm	Hetero. Actuators	Async Cycles	Key Takeaway vs. Ours
[32]	VSL (tunnels)	Model-based RL +safety	No	No	Safety-aware VSL; not cross-actuator or asynchronous.
[33]	VSL	Distributed RL (single actuator)	No	No	Strong VSL; single actuator, no RM/ISC coupling.
[34]	VSL (lane-level)	DRL (TD3)	No	No	Lane-level refinement; still no multi-actuator coordination.
[35]	RM	Optimization +safety constraints	No	No	SOTA RM with safety; single actuator, no VSL/ISC.
[36]	Signals (network)	GraphRL/distributed	No	No	Scales signals via graphs; no VSL/RM; not heterogeneous.
[37]	Signals	Spatiotemporal graph attention MARL	No	No	Strong ATSC with ST-GAT; no VSL/RM integration.
[38]	Signals	Counterfactual MARL (credit assignment)	No	No	Coordination via credit assignment; not heterogeneous, no async cycles.
[39]	RM (multi-ramp)	Game/MARL hybrid	No	No	Multi-ramp coordination; no VSL/ISC or explicit async cycles.
Ours	VSL + RM + ISC	MADDPG	Yes	Yes	Conflict-aware reward + asynchronous scheduling at the expressway–arterial interface; system-level performance beyond isolated controllers.

3. Problem Description

3.1. Interface Types and Study Area

Urban expressway–arterial interfaces can be categorized as direct access and indirect access. Direct access connects facilities via ramps, enabling efficient network utilization while requiring fine-grained management of multiple conflict points; it is prevalent in expressway systems. Indirect access relies on frontage/collector roads with parallel links or shallow-angle merges, which attenuate speed discontinuities and suit surface expressway–arterial

connections. This study focuses on a direct-access interface comprising the expressway merge zone, an entrance ramp, and an adjacent signalized intersection on the arterial.

In practice, an expressway entrance ramp most commonly connects to one downstream signalized intersection on the adjacent arterial, which represents the prevailing and operationally most relevant configuration in many urban expressway systems. While multi-intersection corridor connections may exist in some areas, their coordination requirements are highly site-dependent due to heterogeneous spacing, access forms, and control logics, making a unified “one-size-fits-all” interface formulation difficult to establish without introducing additional assumptions or control layers.

Therefore, this study deliberately focuses on the expressway–single-intersection interface, which enables a clear and deployable characterization of heterogeneous controllers (VSL, RM, and ISC) and their strong coupling mechanisms under realistic operational constraints. Extensions to more complex multi-intersection interface corridors are identified as future work.

The real-world site and topology are illustrated in Figures 1 and 2.



Figure 1. Satellite map of the study area.

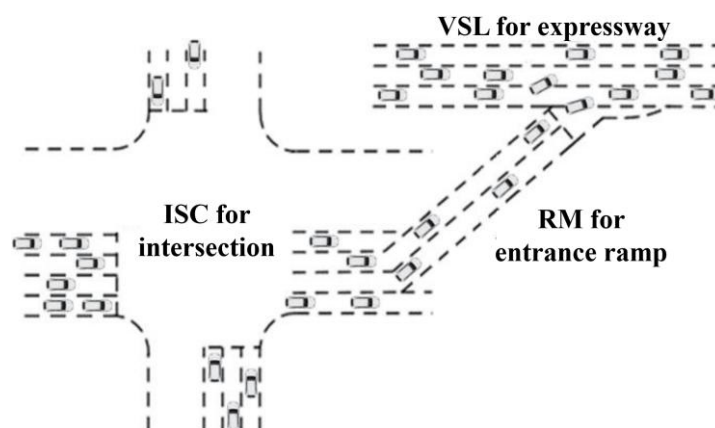


Figure 2. Road network topology of the study area.

3.2. Control Problem Characterization

Coordinated control at the expressway entrance–arterial interface is formulated as a fully cooperative multi-agent problem, where the objective is to optimize global traffic flow under distributed decision-making. Control agents, such as VSL, RM, and ISC, operate on local observations while sharing a global reward, ensuring that individual actions are aligned with system-level objectives. This structure naturally fits the Centralized Training with Decentralized Execution (CTDE) paradigm and motivates the adoption of

a Multi-Agent Deep Deterministic Policy Gradient (MADDPG) approach. The framework is shown in Figure 3

Within this multi-agent reinforcement learning (MARL) framework, each actuator is treated as an autonomous decision-making agent. Specifically, the VSL agent outputs speed limits, the RM agent determines red–green splits, and the signal agent adjusts phase and cycle plans. All agents receive real-time traffic states from the environment, with their policies trained centrally to optimize a shared objective but executed in a decentralized manner to enable real-time responsiveness.

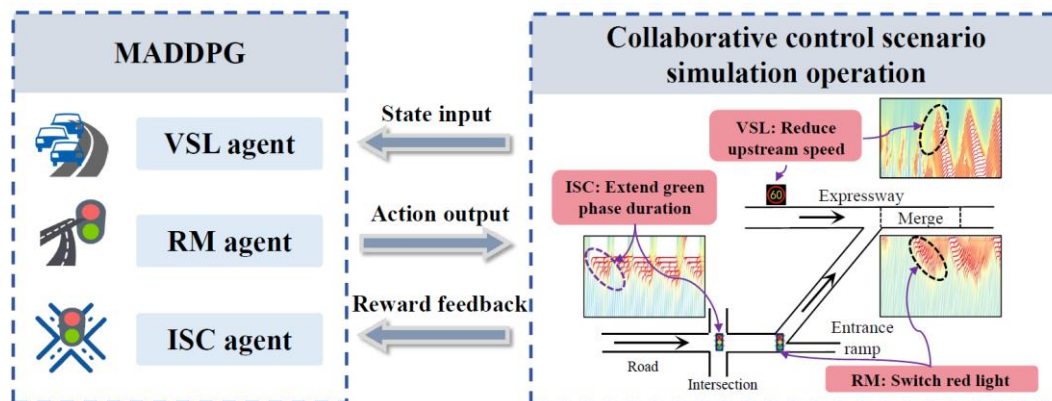


Figure 3. Framework for cooperative control strategies.

3.3. Generalization of the Localized Modeling Approach

The proposed framework adopts a localized modeling strategy centered on the interface region jointly controlled by VSL, RM, and ISC. Although optimization is executed locally, two properties promote generalization to more complex traffic. First, the interface is a structural bottleneck where freeway, ramp, and arterial interactions concentrate. By regulating density near critical values, capping ramp inflow to respect storage, and coordinating junction discharge, the framework suppresses the primary transmission path of congestion and reduces spillovers to upstream and adjacent segments. Second, the asynchronous coordination mechanism aligns heterogeneous control cycles on a common decision grid. This alignment enables modular replication of the VSL + RM + ISC triad at additional interfaces without heavy synchronization overhead. These features make the localized approach composable and transferable to larger networks. Nevertheless, the present study focuses on a single freeway–arterial interface with one adjacent signalized intersection, and network-wide coordination across multiple intersections is beyond the scope of this paper.

4. MADDPG-Based Coordinated Control Strategy

The interface is a multi-factor, multi-scale dynamical system in which heterogeneous agents interact and may conflict. To address decision conflicts, state dimensionality, and mixed action spaces, this study adopts a hierarchical design under a cooperative multi-agent framework. It defines complementary agent roles, builds a unified joint-state representation, and specifies a joint action space with decomposed agent rewards, yielding a coordinated policy via MADDPG, as is shown in Figure 4.

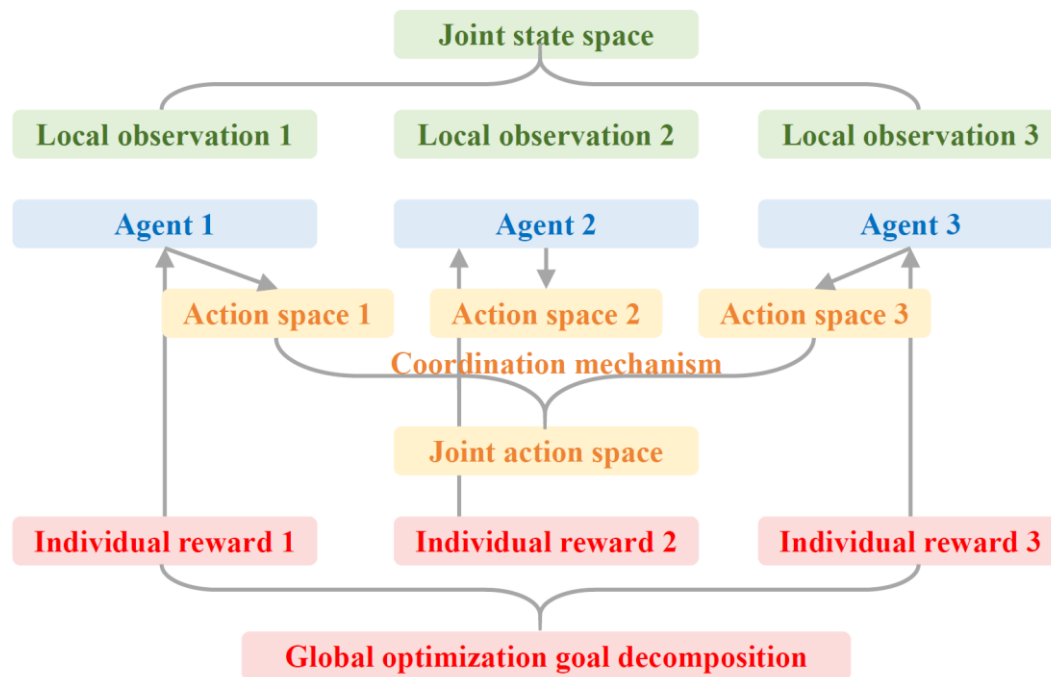


Figure 4. MADDPG-based learning framework.

4.1. Agent Design

VSL agent: Operates on the mainline upstream of the merge. It computes dynamic speed setpoints from density, speed, and flow, harmonizes speeds with ramp inflows, and cooperates with the ramp controller to prevent sharp speed drops and shockwave formation.

RM agent: Acts on the entrance ramp. It adjusts the metering rate based on ramp queue length and local mainline conditions, co-optimizing mainline capacity and ramp delay, and coordinating with the intersection controller to avoid downstream saturation.

Intersection signal control (ISC) agent: Controls the signalized junction on the arterial. It adapts phase splits and cycle parameters using approach queues, aligns the entry cadence to the ramp, and allocates right-of-way to balance arterial inflow and merge stability.

4.2. State Space

The joint state S concatenates local observations of the three agents:

$$S = [s^{VSL}, s^{RM}, s^{ISC}] \quad (1)$$

(1) VSL state:

$$s^{VSL} = [\rho_{up}, q_{up}, \rho_{down}, q_{down}] \quad (2)$$

where ρ_{up} and ρ_{down} denote upstream/downstream densities (veh/km), and q_{up} and q_{down} denote upstream/downstream flows.

(2) RM state:

$$s^{RM} = [w_{queue}, r_{in}, r_{out}] \quad (3)$$

where w_{queue} is the ramp queue length (veh), r_{in} is the on-ramp arrival rate (veh/s), and r_{out} is the metered discharge rate to the merge area (veh/s).

(3) ISC state:

$$s^{ISC} = [phase_{curr}, W_{through}, W_{left}] \quad (4)$$

where $phase_{curr}$ is the current-phase identifier, while $W_{through}$ and W_{left} are the aggregated through and left-turn queue lengths at the entry approaches (veh); the junction is operated with a four-phase plan.

All state variables are measured via detectors placed in SUMO at predefined cross-sections. To remove scale/unit bias and facilitate training stability, each dimension is min-max normalized to $[0, 1]$. During CTDE, the centralized learner can access other agents' observations and actions to stabilize value estimation; execution remains decentralized in real time.

Actions consist of (i) the VSL setpoint for the upstream mainline, (ii) the RM rate (or red-green split) at the entrance ramp, and (iii) the ISC phase/split update at the junction. A global objective (stability and efficiency near the merge) is decomposed into agent-wise rewards that penalize density overshoot, speed variance, spillback risk, excessive queues, and infeasible actuation, while enforcing operational constraints (legal speed bounds, minimum greens, storage limits). This yields a coordinated policy that aligns local decisions with interface-level performance.

4.3. Action Space

The joint action space is composed of the three actuators—VSL, RM, and ISC—each issuing an action for its control domain:

$$\mathcal{A} = [A^{VSL}, A^{ISC}, A^{RM}] \quad (5)$$

(1) VSL action space:

To achieve graded speed control while balancing precision and stability for urban expressways with a baseline limit of 80 km/h, and following evidence that 10 km/h steps effectively smooth shockwaves, the VSL agent selects a discrete speed setpoint from

$$\mathcal{A}^{VSL} = \{30, 40, 50, 60, 70, 80\} \text{ [km/h]} \quad (6)$$

which is broadcast to the upstream mainline.

(2) RM action space:

The RM agent enforces supply-demand balance at the entrance ramp by two-state gating of the metering signal, avoiding ramp spillback and mainline saturation. Because ramp discharge in practice is realized by adjusting the gap between successive green intervals, the instantaneous actuation remains binary (green/red). We therefore model ramp metering as binary logic executed at a fine control period, so that the duty cycle over short windows yields a continuous, demand-responsive effective release rate. Its action is defined as

$$a^{RM} \in \{0, 1\}, a^{RM} = 1 \Rightarrow \text{green(mergepermitted)}, a^{RM} = 0 \Rightarrow \text{red(mergeprohibited)} \quad (7)$$

(3) ISC action space:

Signal optimization requires both phase selection and phase holding time. Let the junction operate a four-phase plan, as shown in Figure 5. The ISC agent chooses a mixed discrete-continuous action:

$$\mathcal{A}^{ISC} = \{(\text{PhaseA}, t_p^A), (\text{PhaseB}, t_p^B), (\text{PhaseC}, t_p^C), (\text{PhaseD}, t_p^D)\} \quad (8)$$

where the phase holding time t_p satisfies operational bounds $t_p \in [10, 50]$ s to ensure minimum green. A smooth squashing map converts a normalized policy output $\mu \in [-1, 1]$ into a feasible t_p :

$$t_p = 30 + 20 \tanh(\mu) \quad (9)$$

which preserves continuity for learning while respecting timing constraints.

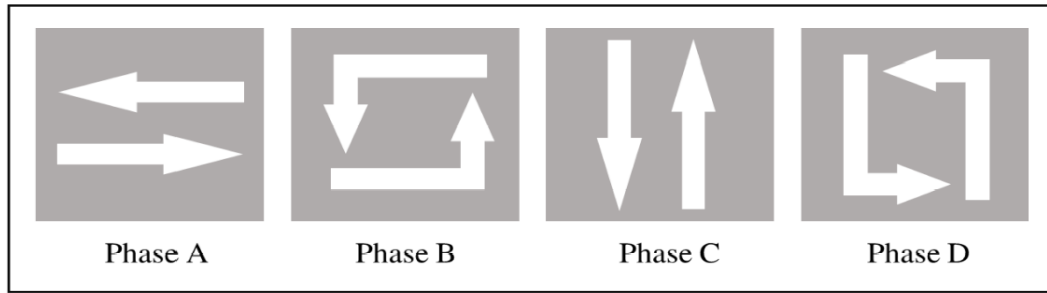


Figure 5. Traffic light phase action space.

During each decision epoch, the selected phase is maintained for t_p seconds and the next decision is then taken. The above action design is compatible with asynchronous update cycles (e.g., VSL every 30 s, RM every 5 s, and ISC decision is quantized to the 5 s scheduling grid, while the effective signal cycle length remains within 10–50 s through phase holding between consecutive decisions), enabling coordinated yet feasible real-time control.

4.4. Asynchronous Control Cycle Coordination Mechanism

In the coordinated setting, the three actuators operate with heterogeneous cycles, which requires a lightweight synchronization rule that preserves feasibility while avoiding excessive switching. To accommodate the heterogeneous temporal requirements of different control agents, an asynchronous control cycle coordination mechanism is developed, as illustrated in Figure 6. Let the minimal time slice be $\Delta = 5$ s and the global decision timeline be $t_k = k\Delta$. The cycles are set as $C_{VSL} = 30$ s, $C_{RM} = 5$ s, and $C_{ISC} = t_p^*$, where t_p is produced by (9) and quantized to the 5 s grid by

$$t_p^* = \text{clip}_{[10,50]}(5 \cdot \text{round}(\frac{t_p}{5})) \quad (10)$$

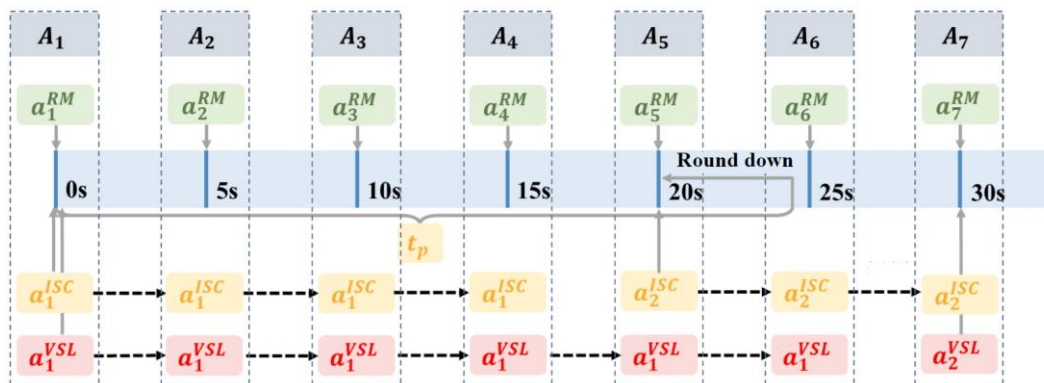


Figure 6. Illustration of asynchronous control cycle coordination mechanism.

At each boundary t_k , the scheduler evaluates three gating conditions and updates only the agents that are due, while the others hold their last action over (t_k, t_{k+1}) :

$$g_{VSL}(k) = \mathbb{I}(k \bmod \frac{C_{VSL}}{\Delta} = 0) \quad (11)$$

$$g_{RM}(k) = 1 \quad (12)$$

$$g_{ISC}(k) = \mathbb{I}(\tau_k \geq t_p^*) \quad (13)$$

where τ_k is the elapsed time since the last phase change. If $g_{VSL}(k) = 1$, a new speed setpoint is issued; otherwise, the previous setpoint persists. If $g_{RM}(k) = 1$, the metering state is

recomputed every 5 s. If $g_{ISC}(k) = 1$, the controller selects the next phase and a new holding time via Equation (9) followed by Equation (10), then resets $\tau_k \leftarrow 0$; otherwise, the current phase continues.

The above rule realizes asynchronous yet coordinated execution on a shared 5 s grid: VSL is reconsidered every 6 slices; RM is updated each slice; and ISC changes only when its held phase expires, with $t_p^* \in [10, 50]$ s ensuring the minimum green. This design aligns updates across actuators without forcing simultaneous changes, preserves legal speed bounds and signal timing constraints, and reproduces the timeline (e.g., VSL at 0/30 s; RM at every 5 s; ISC at phase-expiry instants). As a result, the interface benefits from stable, feasible real-time coordination with minimal overhead.

4.5. Reward Design

In cooperative multi-agent reinforcement learning, the reward must align local actions with the network-wide objective while enabling credit assignment. For the expressway–arterial interface, the global objective is defined as the net boundary flow (outflow minus inflow), which proxies total travel time reduction over the study area. Let f_{in}^t and f_{out}^t denote the aggregated boundary inflow and outflow at time t (veh/s or veh/h, consistent with the sampling step). The global reward is

$$R_{global}^t = f_{out}^t - f_{in}^t \quad (14)$$

To quantify each agent's contribution, the global target is decomposed into agent-specific rewards based on flows that each actuator directly regulates.

Let $f_{ramp-in}^t$ be the arrival flow to the ramp (upstream of the signal) and $f_{ramp-out}^t$ be the discharge to the merge area (downstream of the signal). The RM reward maximizes admissible merge inflow without causing spillback:

$$r_{RM}^t = f_{ramp-out}^t - f_{ramp-in}^t \quad (15)$$

Let $f_{ramp-in}^t$ and $f_{ramp-out}^t$ be the upstream and downstream mainline flows bracketing the VSL zone. To isolate mainline regulation and avoid double-counting the ramp discharge, the VSL reward is

$$r_{VSL}^t = f_{main-out}^t - f_{main-in}^t - f_{ramp-out}^t \quad (16)$$

Let $f_{intersection-in}^t$ and $f_{intersection-out}^t$ be the aggregated inflow and outflow at the connecting junction. The ISC reward promotes safe discharge and queue clearance:

$$r_{ISC}^t = f_{intersection-out}^t - f_{intersection-in}^t \quad (17)$$

To improve learning stability while avoiding structural bias toward any single subsystem, we adopt a shaped reward that combines the flow-based credit assignment with lightweight, physically interpretable regularizers. Specifically, in addition to the global boundary flow objective (Equation (14)) and agent-wise flow contributions (Equations (15)–(17)), we introduce four penalty terms: (i) density overshoot relative to a calibrated critical density $(\rho - \rho^*)$, (ii) speed variance within the VSL zone $\text{Var}(v)$, (iii) queue overflow beyond the available storage at the ramp and intersection approaches $(w - w_{max})$, and (iv) switching costs that penalize excessive actuation changes Δu .

Importantly, these terms are designed to be complementary rather than redundant: the flow terms quantify throughput-oriented contributions attributable to each actuator, whereas the penalties act as feasibility and stability safeguards that prevent “performance gains” from being achieved via undesirable side effects (e.g., suppressing ramp inflow by inducing spillback, or stabilizing mainline flow by creating excessive stop-and-go variability). Moreover, we avoid double-counting by (i) defining each flow-based reward on

the boundary regulated by the corresponding actuator and (ii) applying overflow penalties only when storage constraints are violated, ensuring that the reward does not systematically favor the mainline over the ramp (or vice versa) under normal conditions.

Accordingly, the shaped reward for agent $i \in \{\text{VSL}, \text{RM}, \text{ISC}\}$ is defined as

$$\tilde{r}_i^t = \alpha_i r_i^t + \beta R_{global}^t - \lambda_\rho(\rho - \rho^*) - \lambda_v \text{Var}(v) - \lambda_w(w - w_{max}) - \lambda_s \Delta u \quad (18)$$

where Δu measures control switching, and $\alpha, \beta, \lambda_\rho, \lambda_v, \lambda_w, \lambda_s$ are non-negative weights. This formulation preserves the flow-based credit assignment in Equations (15)–(17) while embedding stability and feasibility considerations that improve convergence and practical deployability.

The shaped reward maps traffic-engineering stability requirements to learnable objectives so that VSL, RM, and ISC produce physically meaningful actions:

- i. Critical-density tracking: Penalizing $(\rho - \rho^*)$ keeps the mainline near the efficient operating point on the fundamental diagram; when density approaches/exceeds ρ^* , the policy lowers the VSL and throttles ramp inflow to avoid breakdown.
- ii. Speed homogenization: Penalizing $\text{Var}(v)$ promotes smooth speed profiles, reducing disturbance gain and preventing stop-go amplification in the VSL zone.
- iii. No spillback: Penalizing $(w - w_{max})$ enforces storage limits at ramps/junctions; when queues rise, ISC prioritizes discharge (longer greens) and/or RM tightens inflow (shorter greens) to avert spillback.
- iv. Smooth actuation: Penalizing Δu discourages myopic high-frequency toggling, thereby stabilizing both speed fields and queue dynamics.

To reduce hyperparameter tuning costs while maintaining stability, we design a lightweight, dynamic scaling mechanism for the regularization weights in the shaped reward. Each penalty weight λ_c^t for constraint $c \in \{\rho, v, w, s\}$ is updated at the end of every 60 control intervals according to

$$\lambda_c^{t+1} = \begin{cases} \min(\lambda_c^t(1 + \eta), \lambda_c^{max}), & \widehat{\mathcal{C}}_c > \epsilon_c \\ \max(\lambda_c^t(1 - \eta), 0), & \widehat{\mathcal{C}}_c \leq \epsilon_c \end{cases} \quad (19)$$

where $\widehat{\mathcal{C}}_c$ denotes the exponentially smoothed violation measure, ϵ_c is the tolerance threshold, and η is a fixed updating rate for both upward and downward scaling. A maximum λ_c^{max} prevents unbounded escalation, while the zero lower bound allows penalties to naturally decay when the system consistently satisfies the constraint.

To prevent weights from remaining permanently inactive, a simple revival mechanism is added: when $\lambda_c = 0$ and the violation $\widehat{\mathcal{C}}_c$ exceeds ϵ_c for 3 consecutive checks, the weight is reactivated to a small seed value λ_c^{seed} (e.g., 10^{-3}). This ensures that critical terms are automatically restored if violations re-emerge. According to traffic flow theory and engineering practice, the thresholds are set as $\epsilon_\rho = 0.003$ veh/m/lane, $\epsilon_v = 4$ (m/s)², $\epsilon_w = 2$ vehicles, and $\epsilon_s = 10$ km/h for VSL. Based on preliminary experiments, the updating rate is set to $\eta = 0.02$.

4.6. Collaborative Control Algorithm

As shown in Figure 7, the collaborative control algorithm architecture involves three key intelligent agents: the VSL agent, the RM agent, and the ISC agent. Each agent is equipped with independent actor networks, while sharing a common critic network (both online and target networks) with multi-head outputs to provide agent-specific value estimates under a centralized training paradigm.

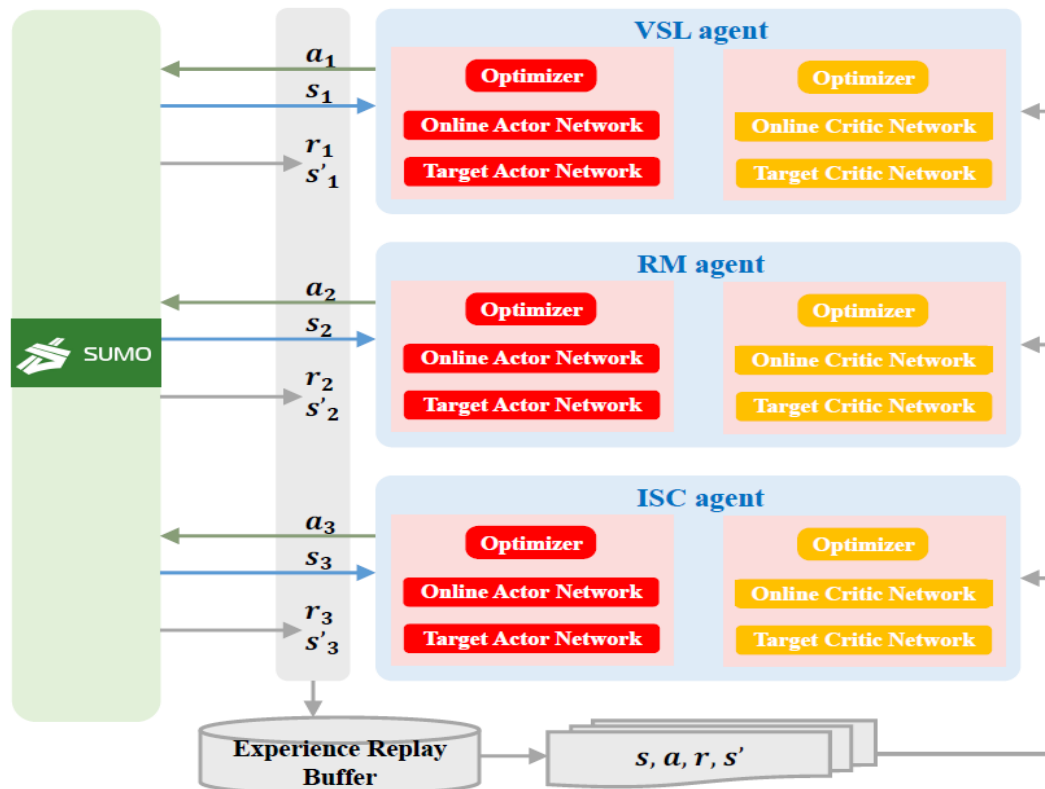


Figure 7. The algorithmic architecture of the collaborative control algorithm.

In terms of the algorithm architecture, the actor networks of each agent maintain an independent structure. The online and target networks of the VSL control agent consist of three fully connected layers (input, hidden, and output layers). The output layer utilizes a Softmax activation function to generate a 6-dimensional discrete action probability distribution, corresponding to six speed limit levels: {30, 40, 50, 60, 70, 80} km/h. The ramp metering control agent shares a similar network structure with the VSL agent, with its output layer using a Sigmoid activation function to produce continuous values between 0 and 1. These values are thresholds (e.g., >0.5 indicates a green light) to convert them into binary control actions.

The intersection signal control agent employs a dual-branch output structure. The phase selection branch outputs a 4-dimensional Softmax distribution, representing the probability of selecting one of the four phase-switching options. The time parameter branch uses a Tanh activation function to generate continuous values in the range $[-1, 1]$, which are then mapped to the 10–60 s interval through the linear transformation in Equation (9).

The critic network, which is based on a global traffic state-sharing mechanism, receives a concatenated vector containing global traffic state information and actions of all agents as input. This network consists of three fully connected layers, with the output layer using the Softmax function to generate a distribution of state–action value evaluations. Both online and target network parameters are synchronized using a delayed update mechanism.

The training process is summarized in Algorithm 1. After system initialization, the agents begin interacting with the traffic simulation platform to acquire the current state s . The current state is fed into the respective actor networks, which select action strategies asynchronously based on the control cycle. The global scheduling tick is 5 s. Controllers are executed asynchronously on this unified 5 s grid: the VSL agent is updated every 30 s, the RM agent is updated every 5 s, and the signal controller (ISC) is updated only when

its phase-holding time expires, where the holding time $t_p \in [10, 50]$ s is quantized to integer multiples of 5 s. Within each 5 s tick, agents that are not due keep their most recent actions (action holding). If an agent does not update its action during a time step, the most recent action strategy is retained. The traffic control actions are then implemented on the simulation platform based on these strategies.

Algorithm 1 Cooperative Traffic Control Strategy

Input: The location of the agents, the traffic simulation environment, and the action exploration parameters

Output: Updated agent actions and network parameters

```

1: Initialize Actor and Critic network parameters, randomly initialize  $Q_\pi$ ,  $\theta_\pi$ ,  $\varphi_\pi$ , and assign corresponding target networks
2: Clear the experience replay pool, set training hyperparameters
3: for each training cycle  $m = 1, 2, \dots, M$  do
4:   Initialize traffic simulation environment, randomly initialize action exploration process  $M$ , and obtain initial state  $s_0$ 
5:   for each time interval do
6:     Get the interaction between the agent and the simulation environment, based on current observation, select action  $a_t^{\{VSL, RM, SC\}}$  according to the cooperative control strategy
7:     Combine the environment state with action  $a_t^{\{VSL, RM, SC\}}$ , calculate the reward, and return new state  $s_t$ 
8:     Store the experience  $(s_t, a_t, r_t, s'_t)$  for each time step in the experience replay pool, regardless of whether all agents update actions
9:     if the experience pool reaches capacity then
10:      for each agent  $i = 1, 2, 3$  do
11:        Sample a mini-batch of experience data from the experience pool
12:        Compute the Critic network loss value  $L(\varphi_i)$ :
          
$$L(\varphi_i) = \mathbb{E}_{(s, a, r, s') \sim \mathcal{D}} [(Q_{\varphi_i}(s, a_1, \dots, a_N) - y_i)^2],$$

          
$$y_i = r_i + \gamma Q_{\tilde{\varphi}_i}(s', \mu_{\tilde{\theta}_1}(o'_1), \dots, \mu_{\tilde{\theta}_N}(o'_N))$$

13:        Compute the Actor network's policy gradient  $\nabla_{\theta_i} J(\theta_i)$ :
          
$$\nabla_{\theta_i} J(\theta_i) = \mathbb{E}_{s \sim \mathcal{D}, a_j \sim \mu_{\theta}} [\nabla_{\theta_i} \mu_{\theta_i}(o_i) \nabla_{a_i} Q_{\varphi_i}(s, a_1, \dots, a_N) |_{a_i = \mu_{\theta_i}(o_i)}]$$

14:        Update the target network parameters:  $\tilde{\varphi} \leftarrow \tau \varphi + (1 - \tau) \tilde{\varphi}$ ,  $\tilde{\theta} \leftarrow \tau \theta + (1 - \tau) \tilde{\theta}$ ,  $\tau \in (0, 1)$ 
15:      end for
16:    end if
17:  end for
18: end for
  
```

During the offline training phase in the digital-twin environment, the simulator provides immediate rewards r and the next state s . Subsequently, the simulator provides immediate rewards r and the next state s' . At the end of each time step, the experience tuple (s, a, r, s') is stored in the experience replay pool, regardless of whether all agents have updated their actions. Once the training samples in the experience replay pool reach the preset capacity, the actor–critic neural networks of all agents are trained based on the algorithm architecture. A small batch of data is randomly sampled, with each sample containing global state, actions of all agents, immediate reward, and the next state. The critic network evaluates the entire state transition process and updates the network parameters, while the actor network synchronously updates the action strategy parameters. Both networks use the Adam algorithm for parameter optimization, forming a complete policy gradient optimization process. It is emphasized that no replay buffer, critic update, or

parameter learning is performed during online field deployment, where only the trained actor networks are executed.

Note that each roadside agent runs its actor locally at the field controller. No peer-to-peer data exchange occurs among agents. Instead, a central controller connected via fiber provides time synchronization and state sharing: every 5 s it aggregates a compact shared state vector from locally reported measurements and broadcasts this vector to all agents. Raw high-rate sensor streams are not redistributed. Each agent consumes its own local sensors plus the shared state vector and computes its command locally (metering state, speed setpoint, or phase/timing). The resulting star topology avoids peer links, keeps bandwidth bounded, and remains compatible with existing roadside infrastructure.

It should be emphasized that the central controller does not execute any policy inference, nor does it relay actions among agents; its role is limited to time synchronization and broadcasting the compact shared state every 5 s, while all policy inference is executed locally by roadside actors.

4.7. Real-Time Feasibility and Computational Complexity

Ensuring practical feasibility is essential for real-time deployment of reinforcement learning-based traffic controllers. In the proposed framework, this requirement is addressed through three design choices:

- (1) Asynchronous control cycle coordination. VSL, RM, and ISC agents operate on heterogeneous cycles (30 s, 5 s, and 10–50 s, respectively), aligned on a unified 5 s grid with an action-holding mechanism. Here, the 5 s grid refers to the global scheduling/communication tick used for state aggregation and due-check, not the signal cycle length; the ISC cycle remains 10–50 s through phase holding, while decisions are synchronized to the 5 s grid by action holding and time quantization. This avoids unnecessary updates while preserving feasibility constraints such as minimum green times and legal speed limits.
- (2) Lightweight execution. Only actor networks are invoked online, while critic networks are used solely during training. Each actor is a compact fully connected model, ensuring constant-time inference.
- (3) Complexity analysis. Let C_{act} denote the cost of one actor forward pass and $l_i(k)$ indicate whether agent i is due at tick k . The per-step cost is

$$T_{exec}(k) = \sum_{i \in \{VSL, RM, ISC\}} l_i(k) C_{act} + O(1) \quad (20)$$

- (4) Online execution pipeline (software mechanism). The system deployment strictly follows CTDE: critics are used only for offline training, and the real-time side does not execute critics; online operation involves only actor policies. The real-time system is driven by a 5 s scheduler and executes: (i) collect local detector summaries and the broadcast compact shared state; (ii) normalize the state; (iii) check due flags of each controller; (iv) perform one actor forward inference only for controllers that are due; otherwise hold the previous action; (v) apply feasibility constraints (legal speed limits, minimum green time, queue safety thresholds, etc.); and (vi) dispatch commands to RM/VSL/ISC. This pipeline uses constant memory, has bounded computation per tick, and is consistent with the complexity analysis in Equation (20).

Relationship to real-time/near-real-time RL literature. Existing real-time reinforcement learning studies can generally be categorized into: (i) online-learning paradigms that continually update policy parameters during operation; and (ii) deployment paradigms that train offline and execute online under strict time budgets. This paper belongs to the latter: the policy is trained in an offline digital-twin environment and only lightweight

actor inference is executed in the field. Unlike synchronous decision or online-learning methods, this work emphasizes deployable real-time coordination under heterogeneous actuator constraints (RM, VSL, signal control) via a unified scheduling grid and an action-holding mechanism. This design better fits practical traffic systems' requirements for safety, stability, and communication feasibility.

It should be emphasized that the proposed framework is designed for online inference rather than online learning. All policy training is conducted offline, while real-time deployment only requires forward inference of pre-trained actor networks. Consequently, the runtime burden during operation is dominated by neural network inference, rather than iterative optimization or gradient-based updates.

The actor networks employed in this study are compact feedforward models with fixed architecture, resulting in constant-time inference complexity for each control decision. Given the heterogeneous control cycles (5 s for ramp metering, 30 s for VSL, and 10–50 s for signal control), the required inference and coordination operations can be completed well within the corresponding control intervals under typical traffic control hardware configurations. This deployment-oriented runtime property is also consistent with the online complexity analysis in Equation (20). This design ensures that the proposed coordination framework is compatible with real-time deployment in practical traffic management systems.

While hardware-specific latency measurements may vary across platforms, the focus of this study is to demonstrate algorithmic and architectural feasibility for real-time operation, rather than platform-dependent performance benchmarking.

5. Experimental Setup and Results

5.1. Scenario and Parameter Settings

5.1.1. Scenario Design

A high-fidelity network model in SUMO is built for the expressway–arterial interface to support experimental validation of coordinated control among VSL, RM, and ISC. The scenario reproduces the Wanjiali Elevated Road and the Renmin East Road–Wanjiali Middle Road junction in Changsha, Hunan. As illustrated in Figure 8, the elevated expressway has six lanes (two-way); the entrance ramp provides two lanes; and the merge section expands to four lanes due to an acceleration lane of 160 m. The ramp length is 250 m. For visual clarity, the expressway alignment is vertically offset in the schematic while preserving geometry and control distances. The ramp origin lies 90 m upstream of the junction on Wanjiali Middle Road. Both Wanjiali Middle Road and Renmin East Road include lane additions 120–150 m from the intersection.

5.1.2. Parameter Settings

(1) Simulation calibration.

The Intelligent Driver Model (IDM) for car-following and the LC2013 model are adopted in SUMO and calibrated for this site to ensure credible dynamics under mixed conditions. Calibrated parameters are reported in Tables 2 and 3.

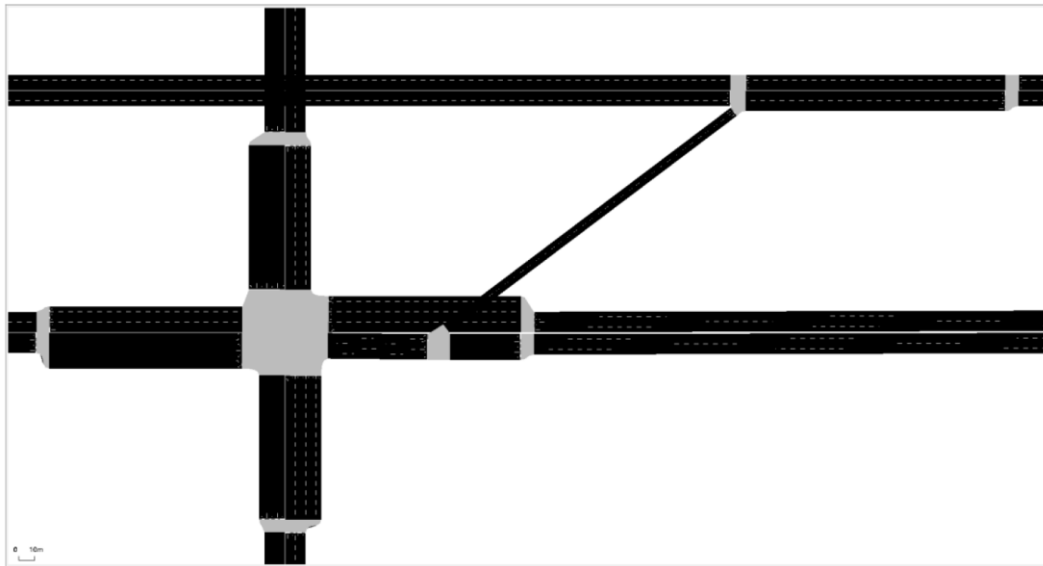


Figure 8. SUMO model of the road network in the study area.

Table 2. IDM Specifications.

Parameter	Normal Road	Ramp	Expressway
Desired Speed	45 km/h	39 km/h	71 km/h
Maximum Acceleration	2.4 m/s ²	1.6 m/s ²	3.5 m/s ²
Comfortable Deceleration	1.9 m/s ²	2.0 m/s ²	1.8 m/s ²
Minimum Car Distance	1.1 m	1.4 m	2.1 m
Desired Headway Time	1.6 s	1.7 s	2.2 s

Table 3. LC2013 Model Specifications.

Parameter	Normal Road	Ramp	Expressway
Strategy Traffic Control	2.0	5.0	0.6
Cooperative Traffic Inclination	1.0	1.5	0.5
Speed Increase Benefit	1	0.1	4
Right-Turn Inclination	0.1	0	1.0

(2) Training configuration.

Training runs for 1500 episodes, each lasting 2 h of simulated time. The first 150 episodes warm up the replay buffer; the remaining 1350 episodes optimize agent policies toward the shared objective. Demands are shown in Table 4.

To assess robustness beyond a single operating condition, the training/evaluation scenarios are configured with multiple demand levels (Table 4), which span moderate to heavily congested periods. These demand levels are designed to cover both moderate and heavily congested regimes with recurrent bottleneck activation, rather than a single operating condition. This design allows the relative performance of different strategies to be examined across distinct congestion regimes, rather than relying on a single fixed demand realization.

Experiments use an Intel i5-14500 CPU, Windows 10, Python 3.8, PyTorch 1.12.1 with CUDA 11.6, and an NVIDIA RTX 3060 (24 GB). This configuration supports SUMO multi-

processing and large replay buffers. Hyperparameter settings are summarized in Table 5. It should be noted that, in practical implementation, the forward inference of each actor network requires only a few milliseconds on a standard CPU, which is negligible relative to the 5 s scheduling interval.

Table 4. Traffic demand of the training scenario.

Time Interval (min)	Expressway Flow (veh/h/lane)	Ramp Flow (veh/h/lane)	Arterial Road Flow (veh/h/lane)
0–20	1400	400	700
20–40	1500	500	800
40–60	1600	600	900
60–80	1700	700	1000
80–100	1800	800	1100
100–120	1900	900	1200

Table 5. Hyperparameter settings of the MADDPG algorithm.

Hyperparameter	Value
Training episodes	1500
Learning rate (Actor)	0.0001
Learning rate (Critic)	0.003
Discount factor	0.99
Replay buffer size	100,000
Batch size	256
Random noise	0.2
Target network update parameter	0.005

5.2. Baseline Controllers

(1) ISC-only: In the actuated plan, this computes the initial green from historical demand,

$$G_{initial} = \alpha \bar{Q}_{min} + \beta \quad (21)$$

and updates the current green with detector-triggered extensions capped by G_{max} :

$$G_{current}(t) = \min(G_{initial} + \sum_{k=1}^n \Delta G_k, G_{max}) \quad (22)$$

If no call is detected, the phase terminates; otherwise, it holds with ΔG_k while respecting the minimum green.

(2) RM-only (ALINEA): Occupancy-feedback metering adjusts the rate each cycle:

$$r_k = r_{k-1} + K_r(o^* - o_{k-1}) \quad (23)$$

with saturation to feasible bounds and queue spillback protection.

(3) VSL-only (PID): Speed setpoints follow a PID law on the density error $e(k) = \rho_{target} - \rho(k)$:

$$v_{lim}(k) = v_{lim}(k-1) + K_p e(k) + K_i \sum_{j=0}^k e(j) \Delta t + K_d \frac{e(k) - e(k-1)}{\Delta t} \quad (24)$$

followed by projection to $[v_{min}, v_{max}]$. Gains are tuned once for the site and held fixed across tests.

5.3. Training Procedure and Results

A total of 1500 training episodes are conducted, and Figure 9 shows the per-episode average reward. Since the global objective is defined by the net boundary flow within the study area, the reward is negative by construction, and values closer to zero indicate higher network efficiency. The learning process can be divided into three stages. During episodes 0–450, exploratory actions with stochastic noise led to large and frequent fluctuations. During episodes 450–1050, the reward improves steadily as the decisions of RM, VSL, and ISC become progressively aligned. During episodes 1050–1500, the policy converges and stabilizes above -100 with only small oscillations. These results demonstrate that the proposed MADDPG-based coordinated control achieves stable convergence while enhancing interface-level efficiency under mixed traffic conditions.

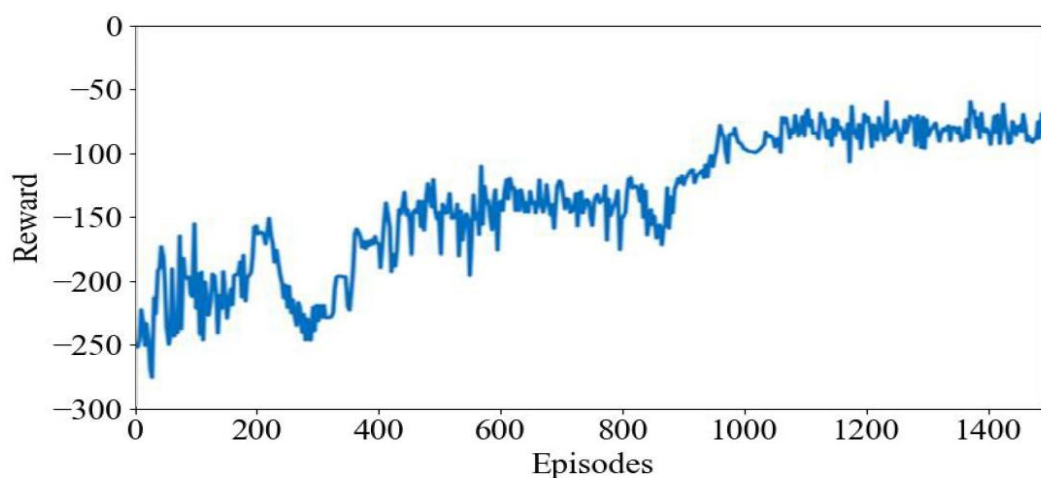


Figure 9. Convergence curve.

5.4. Comparative Analysis of Control Performance

Based on the convergence curve in Figure 9, we select the policy after 1300 training iterations as the optimal model; it exhibits stable convergence and strong generalization and is benchmarked against classical traffic-control methods. Using the simulation outputs, we extract for each strategy (i) vehicle trajectories, (ii) control actions, and (iii) traffic performance indicators to visualize and compare spatiotemporal trajectories on the on-ramp, expressway, and arterial, to contrast control actions, and to compare efficiency metrics, thereby enabling a comprehensive evaluation of control effectiveness.

From a mechanism perspective, the performance advantage of the proposed coordinated control does not come from a more aggressive regulation by any single actuator. Instead, it is primarily attributed to cross-controller coordination that aligns (i) mainline speed harmonization (VSL), (ii) ramp inflow regulation (RM), and (iii) junction discharge organization (ISC) toward a consistent interface-level objective. Such alignment mitigates the typical conflicts observed under isolated control, where a locally optimal action (e.g., overly restrictive ramp metering or delayed signal discharge) can unintentionally activate bottlenecks, amplify shockwaves, or trigger queue spillback across subsystems.

5.4.1. Comparison of Vehicle Spatiotemporal Trajectories

Figure 10 illustrates the spatiotemporal trajectories on the on-ramp across the four control strategies. As flow transitions toward congestion, all strategies exhibit sharp speed drops at the ramp bottleneck with backward-propagating shockwaves. In the no-control case, ramp congestion is the most severe, emerging after approximately 4200 s and spreading to the ramp entrance. Ramp-only control (ALINEA) does not alleviate ramp conges-

tion and even triggers an earlier breakdown (around 4000 s), because heavy mainline demand activates ALINEA and over-shortens ramp green time, intensifying queues. Under VSL-only, congestion onset is deferred to approximately 5200 s; slower mainline speeds enable smoother merges and markedly improve control, yet once demand exceeds a critical threshold, congestion still reaches the ramp entrance. With MADDPG global coordination, onset is not materially postponed, but shockwave propagation along the ramp decelerates notably, yielding performance superior to either single-control scheme.

Figure 11 presents the spatiotemporal patterns on the expressway mainline under the same strategies. Throughout operations, both the beginning and end of the merge area show pronounced speed drops with backward shockwaves. In no control, congestion starts earliest and persists longest (from approximately 3600 s to the end). As conditions deteriorate, the high-speed region contracts, and a low-speed band propagates upstream from the merge. VSL-only delays the onset of shockwave spread by about 300 s, but once demand surpasses the threshold, the wave still advances upstream. With ramp-only (ALINEA), compressed ramp green time under high mainline density postpones mainline shockwave diffusion until after approximately 5000 s—an evident improvement that comes at the expense of heavier ramp queues. Under MADDPG coordination, the diffusion onset is comparable to ramp-only, yet alternating release/suppression creates a dynamic buffer, fragmenting the upstream congestion band and further enhancing overall performance.

Figure 12 depicts the spatiotemporal distribution on the arterial. Two dominant bottlenecks appear: the signalized intersection (approximately 150 m) and the arterial–on-ramp junction (approximately 250 m), each accompanied by upstream-moving low-speed bands. Strategy effects differ markedly. In no control, the intersection congests earliest and over the widest extent; after approximately 2500 s, a strong speed gradient emerges at the junction, and shockwaves spread from the on-ramp entrance toward the intersection, forming a continuous low-speed band. Around 3000 s, junction congestion dissipates, but high demand at the (previously) unsignalized intersection drives upstream propagation. With actuated signal control, real-time cycle adjustments defer congestion onset at both locations to around 5000 s; thereafter, junction congestion extends to the intersection, and intersection congestion propagates upstream. Under MADDPG coordination, onset is pushed further back—to approximately 5600 s at the junction and after 6000 s at the intersection—demonstrating a clear advantage over single-control approaches.

Collectively, these panels indicate that MADDPG-based global coordination improves the spatiotemporal organization of traffic, attenuates the speed and extent of shockwave propagation, and substantially increases throughput in the on-ramp/arterial interface region. Specifically, the spatiotemporal trajectory comparisons in Figures 10–12 suggest that the proposed coordination forms a dynamic buffer at the interface: VSL moderates upstream speed and reduces the disturbance gain of mainline traffic, RM regulates the ramp discharge to prevent abrupt merging-induced breakdown, and ISC adjusts arterial release to avoid excessive queue accumulation that could spill back to the ramp and further destabilize the merge. This coordinated buffering mechanism explains why the proposed method can simultaneously slow down shockwave propagation on the ramp and delay congestion onset on the arterial, rather than merely shifting congestion from one subsystem to another.

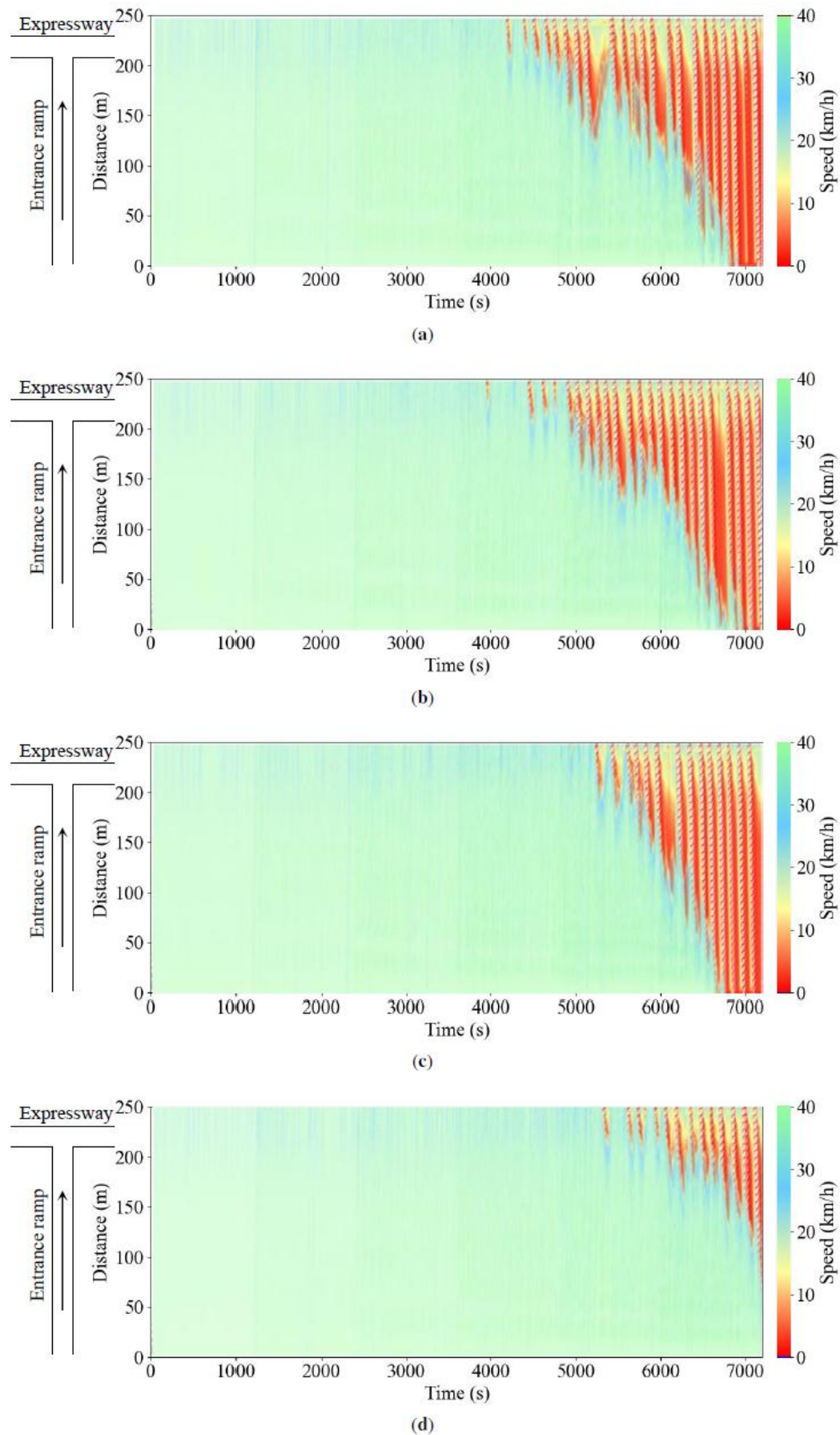


Figure 10. Spatial-temporal trajectories of vehicles on ramp under different control methods: (a) No control. (b) RM-only. (c) VSL-only. (d) MADDPG.

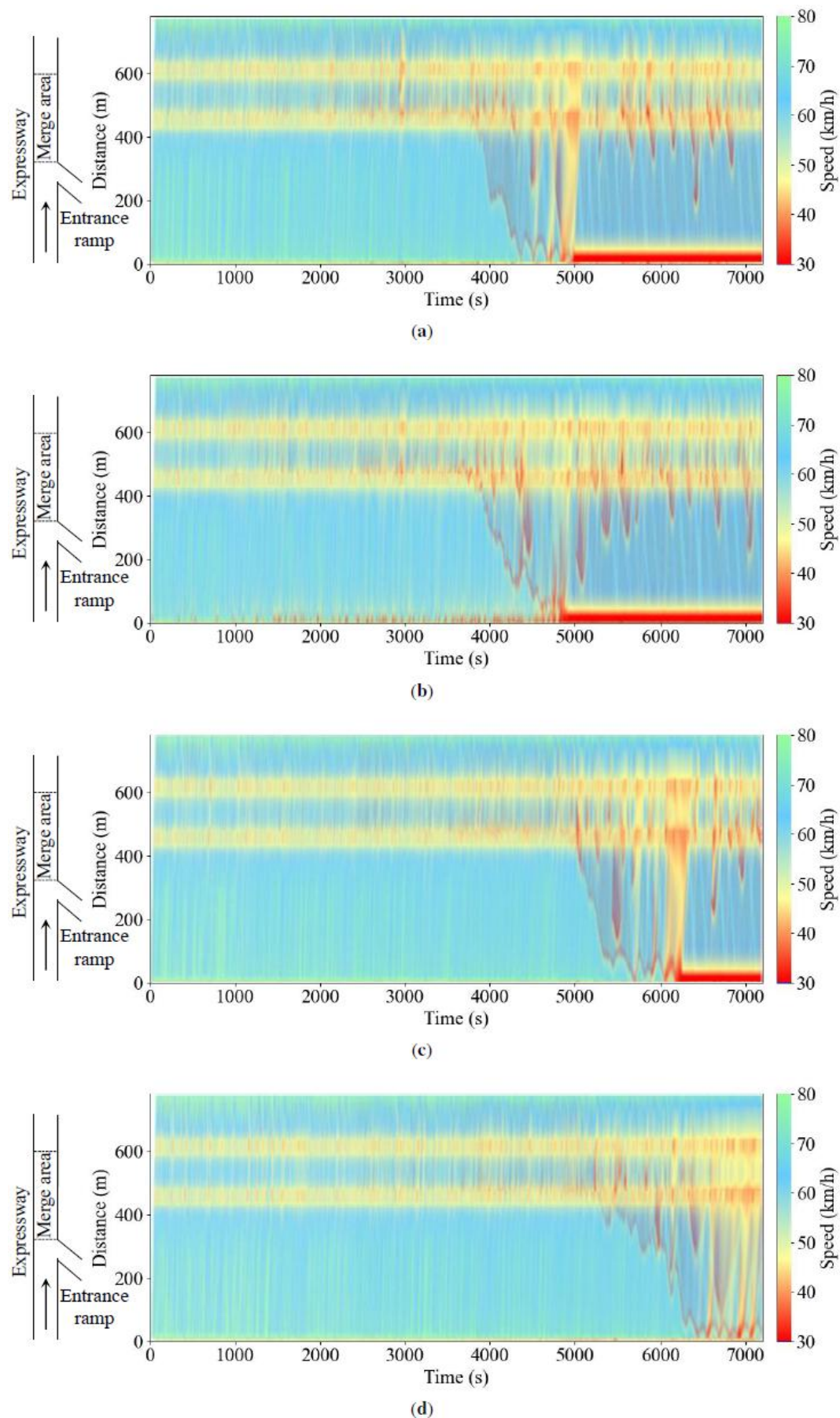


Figure 11. Spatial-temporal trajectories of vehicles on expressway under different control methods: (a) No control. (b) VSL-only. (c) RM-only. (d) MADDPG.

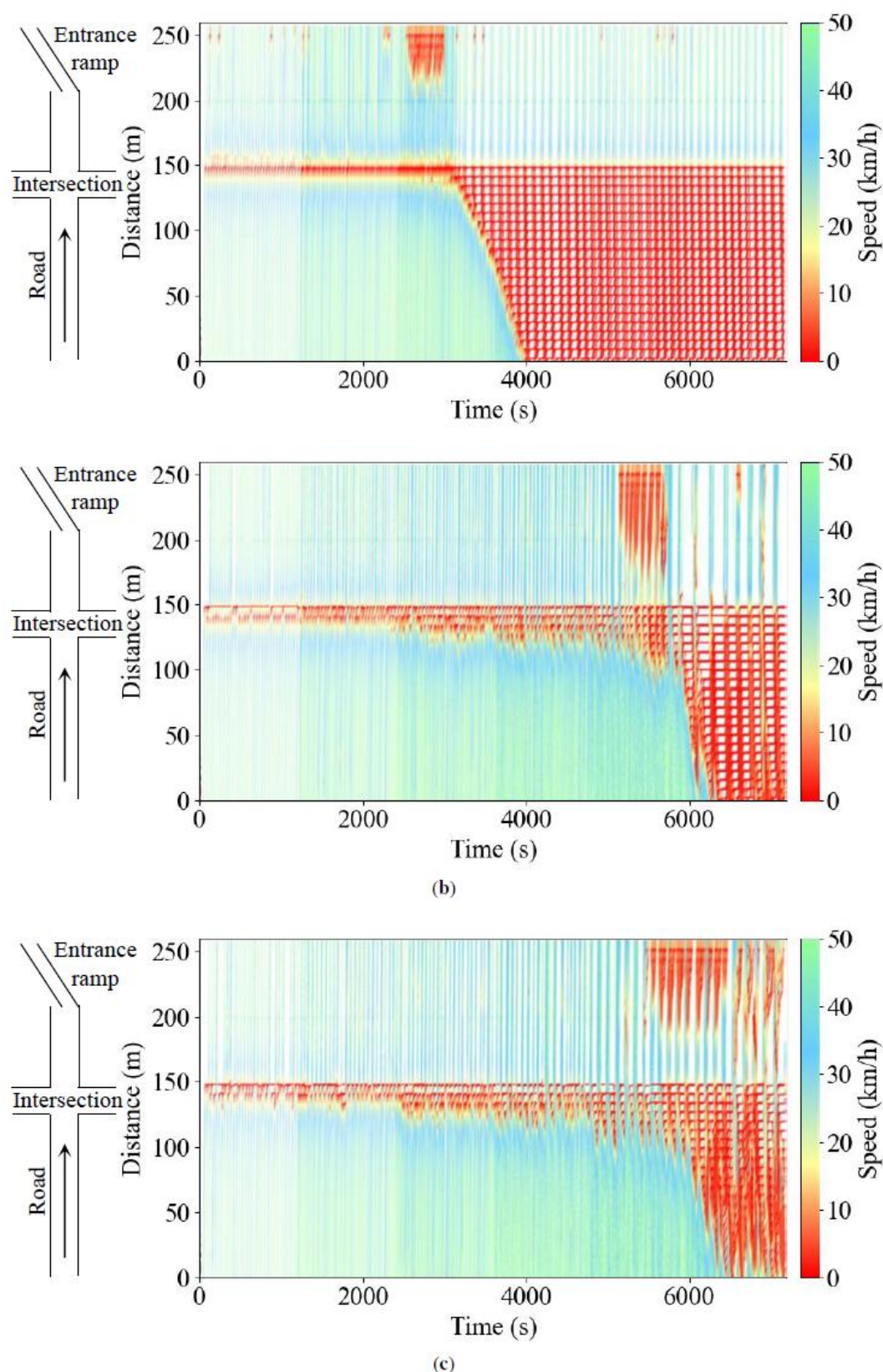


Figure 12. Spatial-temporal trajectories of vehicles on regular road under different control methods: (a) No control. (b) ISC-only. (c) MADDPG.

5.4.2. Comparative Analysis of Control Actions

Figure 13 contrasts the time series of control actions across strategies, allowing a direct, like-for-like reading. In the VSL dimension, relative to MADDPG coordination, the

VSL-only scheme exhibits frequent, large swings and multiple plateaus at 70 km/h or higher, indicating limited sensitivity and weak regulation. For instance, from approximately 4950 s to 4980 s it drops abruptly from 80 km/h to 60 km/h, whereas MADDPG over the same interval adjusts more moderately from 50 km/h to 60 km/h. As demand rises, VSL-only continues to elevate setpoints without anticipating ramp demand; from about 5070 s to 5220 s it maintains speeds at 70 km/h or above, enabling the congestion band to propagate upstream on the expressway mainline (consistent with the red band in Figure 11b).

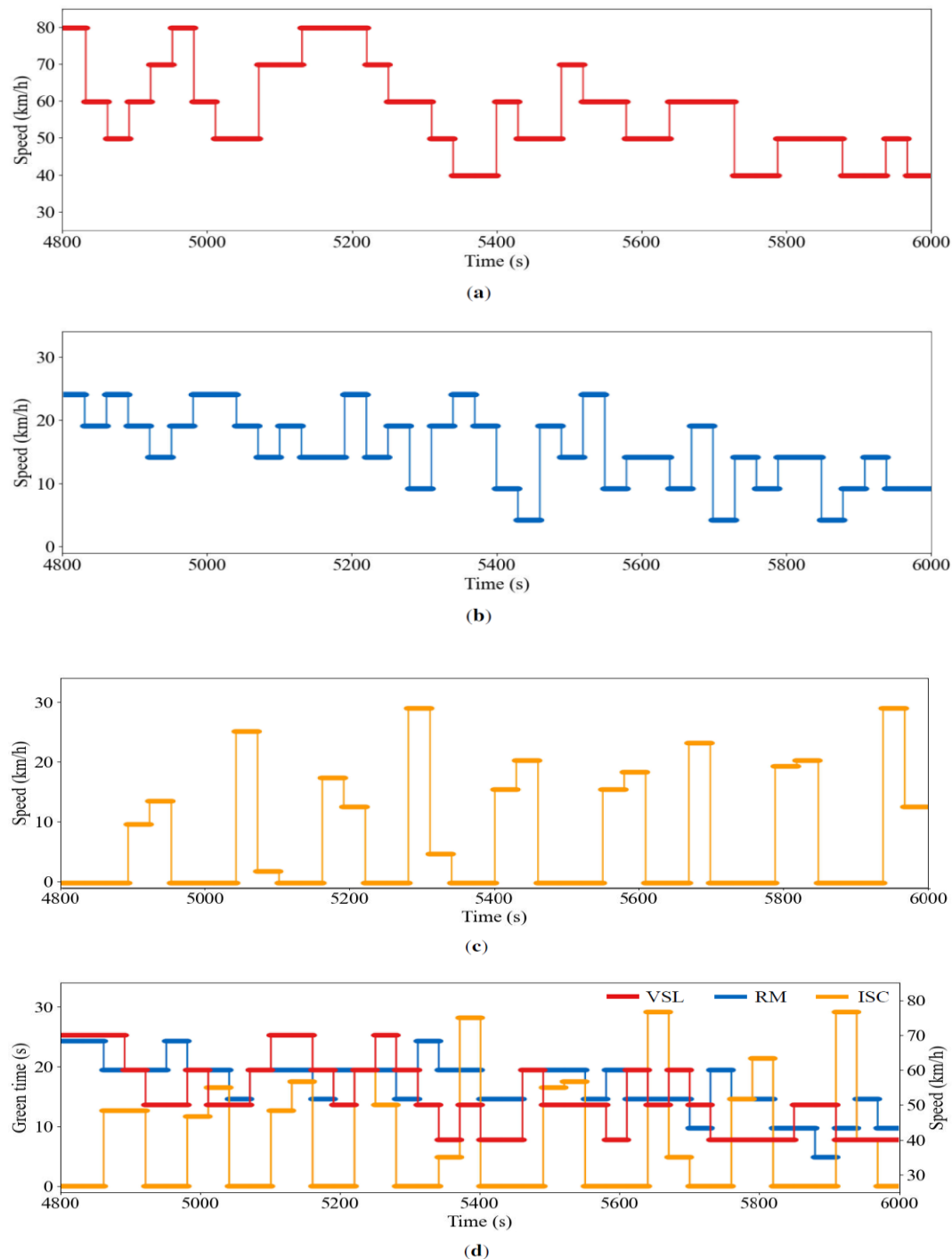


Figure 13. Sequence diagram of control actions under different control methods: (a) VSL-only. (b) RM-only. (c) ISC-only. (d) MADDPG.

Turning to ramp metering, the ramp-only (ALINEA) sequence fluctuates more than MADDPG and tends to over-compress ramp greens when mainline demand is high. At

5430 s, 5700 s, and 5850 s, it issues three very short green intervals of 5 s, while MADDPG in the same periods keeps green time at 10 s or longer and, over the full horizon, produces only one 5 s minimum. Summary statistics reinforce this pattern: intervals with green time at or below 10 s account for 27.5% under ramp-only, versus 15% under MADDPG. A representative coordinated action occurs at 5310 s, where MADDPG lowers the mainline limit to 40 km/h while maintaining a 25 s ramp green, thereby avoiding over-queuing.

At the intersection, actuated signal control is less stable than MADDPG: individual green-phase durations vary more, and the gaps between greens are longer. The standard deviation of green-phase duration is 6.49 s for actuated control compared with 3.64 s under MADDPG. Prolonged green-phase gaps exceeding 90 s occur during 4800–4890 s, 4950–5040 s, 5460–5550 s, 5700–5790 s, and 6050–6140 s under actuated control, whereas MADDPG exhibits only three such instances. A notable coordinated maneuver appears at 5910 s: MADDPG reduces the VSL from 50 km/h to 40 km/h, simultaneously opens the intersection green for 38 s, and—30 s later—extends the ramp green from 10 s to 15 s. This coupled, multi-agent adjustment materially improves control effectiveness at the expressway–arterial interface.

Figure 14 provides a spatiotemporal visualization of the control actions summarized in Figure 13. Around 4800 s, where Figure 13 indicates a coordinated maneuver: reducing the mainline VSL, maintaining a 25 s ramp green, and shortening the intersection phase. The corresponding traffic states evolve markedly differently between the uncontrolled and MADDPG cases. In the uncontrolled scenario (left column), congestion originating at the merge area propagates upstream, restricting outflow and triggering cascading effects that eventually spill back to the intersection. In contrast, under MADDPG coordination (right column), the VSL agent enforces a smoother outflow at 70 km/h, the RM agent restricts inflow with a stable 25 s green, and the ISC agent temporarily sets the green phase to 0 s to relieve downstream pressure. These coupled actions jointly stabilize the traffic stream, eliminate shockwave amplification, and prevent spillback. Hence, the figure not only validates the effectiveness of the selected actions in Figure 13 but also demonstrates the key advantage of multi-region coordination: overcoming the locally optimal yet globally inefficient outcomes of isolated control and achieving system-wide stability.

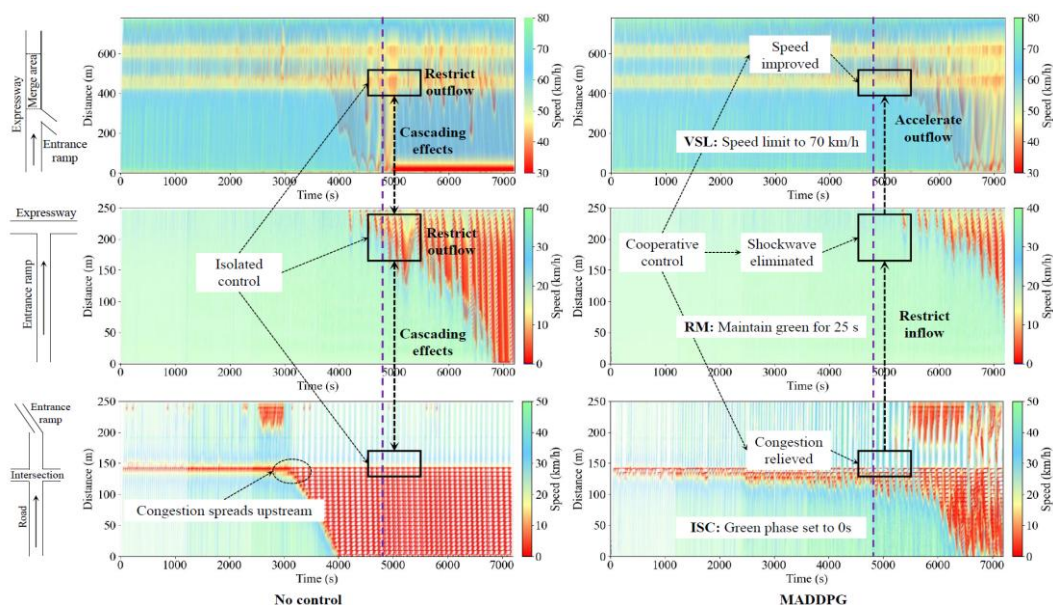


Figure 14. Case analysis of coordinated actions: spatiotemporal evolution under no control and MADDPG.

5.4.3. Comparative Analysis of Traffic Efficiency

The representative section of the expressway is selected at the merging area (marked as No.1), the representative section of the ramp corresponds to the entire ramp segment (marked as No.2), and the representative section of the arterial road is chosen at the approach of the Wanjiali Road intersection (marked as No.3). These areas best reflect the traffic flow conditions and the control effectiveness at the interface between urban expressway entrances and arterial roads.

Tables 6–8 summarize detector outputs at the three locations over the 2 h simulation, with averages computed every 20 min for travel time, speed, and occupancy. In Tables 4 and 5, Method 1 denotes ramp-only control, Method 2 denotes VSL-only control, and Method 3 denotes MADDPG coordination. In Table 6, Method 1 denotes no control, Method 2 denotes actuated signal control, and Method 3 again denotes MADDPG coordination.

From Table 6, performance under low demand (0–60 min) is similar for Methods 1 and 2, whereas Method 3 already shows an advantage: occupancy remains consistently lower than in the single-strategy baselines. Under medium–high demand (60–120 min), Method 1 outperforms Method 2, but Method 3 exhibits stronger anti-saturation capability and widens the gap to both baselines. During 60–80 min, the travel time of Method 3 (13.56 s) is 62.4% of Method 2 (21.74 s), and the average speed increases by 58.6%, aligning with the sparser, less saturated red bands observed on the mainline—evidence that coordinated control delays the formation of merge-area congestion. During 100–120 min, the travel times of Methods 1 and 2 are both around 30 s, while Method 3 remains below 25 s.

The above comparisons also provide empirical evidence that the shaped reward does not bias the control toward a single subsystem. As summarized in Tables 6–8, improvements in mainline efficiency are not achieved by sacrificing ramp or intersection operations; instead, multiple indicators (e.g., travel time, speed, and occupancy) improve concurrently across demand periods. This consistency supports the intended role of the penalty terms in Equation (18) as stability/feasibility safeguards rather than dominant objectives that distort subsystem priorities.

Overall, the experimental results indicate that, compared with representative strategies such as ALINEA-RM, PID-VSL, and actuated signal control, the proposed interface-level coordinated control framework achieves consistent and significant advantages across multiple performance metrics, including mainline traffic stability, ramp queue safety, and overall interface efficiency (see Tables 6–8). These improvements stem from cross-controller information sharing and coordinated decision-making, thereby validating the effectiveness of the interface-level coordination mechanism.

Turning to Table 7, differences between Methods 1 and 2 are minor during low demand (0–60 min). Between 40 and 60 min, Method 3 reduces occupancy by more than 10 percentage points relative to Methods 1 and 2, with only a slight increase in travel time—indicating less vehicle accumulation on the ramp without sacrificing efficiency. Under medium–high demand (60–120 min), once the flow reaches 700 veh/h/lane, Method 1's occupancy surpasses Method 2 and surges toward 40%, confirming the ALINEA-induced “over-control” loop whereby excessive green-time compression aggravates queuing (consistent with the severe red band at the ramp entrance in Figure 11d). In contrast, Method 3 dynamically coordinates ramp and mainline: during 80–100 min, occupancy is held at 6.4% (40.6% of Method 1), and travel time is 56% lower.

Table 8 indicates that, under low demand (0–60 min), all strategies perform comparably; Method 3 is slightly worse than Method 2 on several metrics, reflecting a modest coordination overhead when volumes are light. Under medium–high demand (60–120 min), Method 1's travel time spikes sharply, matching the continuous red band from the intersection to upstream segments in Figure 12a. Method 3 yields substantially shorter

travel times than Methods 1 and 2, implying a more favorable spatial distribution of congestion. During 100–120 min, the travel time of Method 3 (85.71 s) is 62.7% of actuated control (136.72 s), consistent with the fragmented red bands in Figure 12c and confirming that coordination dampens shockwave propagation.

Table 6. Comparison of traffic efficiency indicators at expressway location no. 1.

Time (min)	Flow (veh/h/lane)	Travel Time (s)/Speed (m/s)/Occupancy (%)		
		Method 1	Method 2	Method 3
0–20	1300	11.82/13.57/7.08	12.06/13.31/8.17	11.52/13.92/6.18
20–40	1400	12.27/13.08/8.58	12.76/12.58/9.98	11.88/13.51/7.30
40–60	1500	13.02/12.34/10.23	15.22/10.58/13.17	12.37/12.98/8.91
60–80	1700	16.02/10.06/13.95	21.74/7.47/19.78	13.56/11.85/10.96
80–100	1800	23.39/6.95/21.17	25.01/6.51/22.63	20.54/7.90/17.90
100–120	1900	29.30/5.53/22.91	34.62/4.68/21.87	24.88/6.51/23.35

Table 7. Comparison of traffic efficiency indicators at ramp entrance location no. 2.

Time (min)	Flow (veh/h/lane)	Travel Time (s)/Speed (m/s)/Occupancy (%)		
		Method 1	Method 2	Method 3
0–20	400	23.32/10.72/3.32	23.29/10.73/3.47	23.27/10.74/2.84
20–40	500	23.48/10.64/4.02	23.53/10.62/4.18	23.37/10.69/3.5
40–60	600	23.65/10.57/4.68	23.66/10.56/4.84	23.35/10.67/4.17
60–80	700	26.21/9.15/5.92	23.86/10.47/6.15	23.70/10.55/4.85
80–100	800	62.72/3.99/15.76	40.85/6.12/10.5	27.55/9.05/6.4
100–120	900	135.13/1.85/38.92	91.42/2.69/28.85	63.65/3.93/16.9

Table 8. Comparison of traffic efficiency indicators at conventional road location no. 3.

Time (min)	Flow (veh/h/lane)	Travel Time (s)/Speed (m/s)/Occupancy (%)		
		Method 1	Method 2	Method 3
0–20	700	33.38/3.07/1.15	26.28/3.84/1.57	31.92/3.18/1.86
20–40	800	35.22/2.71/1.5	26.02/3.87/1.56	31.83/3.2/2.02
40–60	900	40.32/2.57/1.84	29.01/3.67/2.01	32.33/3.09/2.2
60–80	1000	189.47/0.54/8.35	37.56/2.83/2.55	34.16/3.08/2.16
80–100	1100	225.83/0.45/30.85	44.92/2.36/15.4	38.26/2.74/10.75
100–120	1200	334.96/0.31/39.81	136.72/0.75/35.49	85.71/1.21/25.49

Across demand periods, although absolute metrics vary with traffic intensity, the relative performance ordering remains consistent in Tables 6–8. Therefore, the reported improvements are systematic across demand scenarios instead of being driven by stochastic variation in a single run. In particular, the coordinated control strategy exhibits systematic advantages under medium–high demand periods, suggesting that the reported improvements reflect stable coordination effects rather than stochastic variation from a single simulation instance.

While contemporary RL/MPC controllers could be considered under unified modeling and timing assumptions, the selected baselines (PID-VSL, ALINEA-RM, and actuated signal control) represent mature and widely deployed practice. Therefore, the comparisons in Tables 6–8 quantify the practical advancement of interface-level coordination over long-standing isolated control paradigms, which is the primary focus of this study.

5.4.4. Sensitive Analysis

To further examine robustness against imperfect sensing, we injected random observation errors of 5%, 10%, and 15% into the state inputs and evaluated performance across three representative locations, as shown in Figure 15. The results indicate that the proposed framework degrades gracefully with increasing noise. At all locations, larger observation errors lead to slightly longer travel times, lower average speeds, and higher occupancies, particularly under congested periods (Periods 5–6). For example, at Location 3 in Period 6, the average travel time increases from approximately 120 s under 5% noise to about 160 s under 15% noise. Nevertheless, even under 15% error, the overall system remains stable without collapse, and the control strategy continues to suppress shockwave amplification and prevent spillback. These findings confirm that the proposed coordination mechanism is tolerant to moderate sensing inaccuracies, thereby enhancing its applicability in real-world environments where measurement noise is unavoidable.

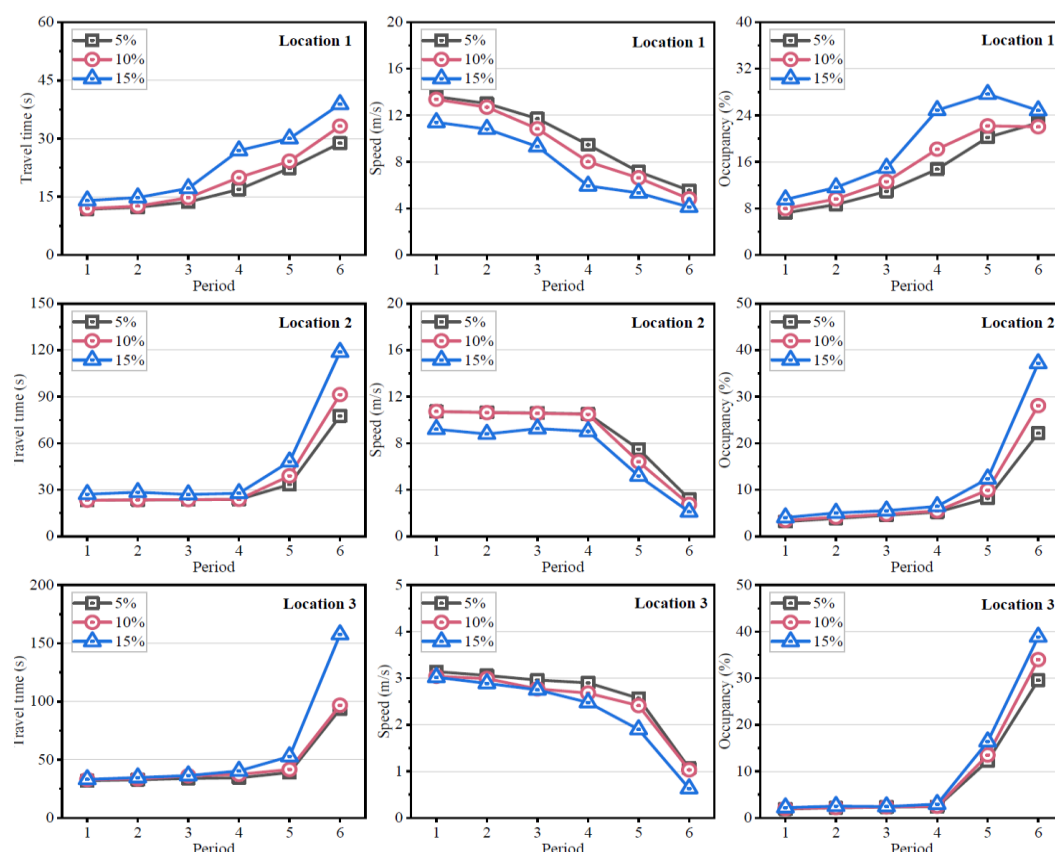


Figure 15. Effect of sensor observation errors on traffic performance.

We assess the lightweight dynamic scaling of regularization weights by varying the updating rate $\eta \in \{0.01, 0.02, 0.03, 0.05\}$. A clear U-shaped pattern emerges in Figure 16; with $\eta = 0.01$, weight adaptation is conservative and shockwave suppression is delayed under rising demand; with $\eta = 0.05$, penalties are over-amplified and gradients become aggressive, inducing action chattering and, in peak Periods 5–6, higher travel time, lower speed, and higher occupancy. Mechanistically, η governs the convergence rate of the relative importance of penalties and the primary task terms—being too small under-emphasizes these constraints, while being too large mis-emphasizes them and distorts the action space.

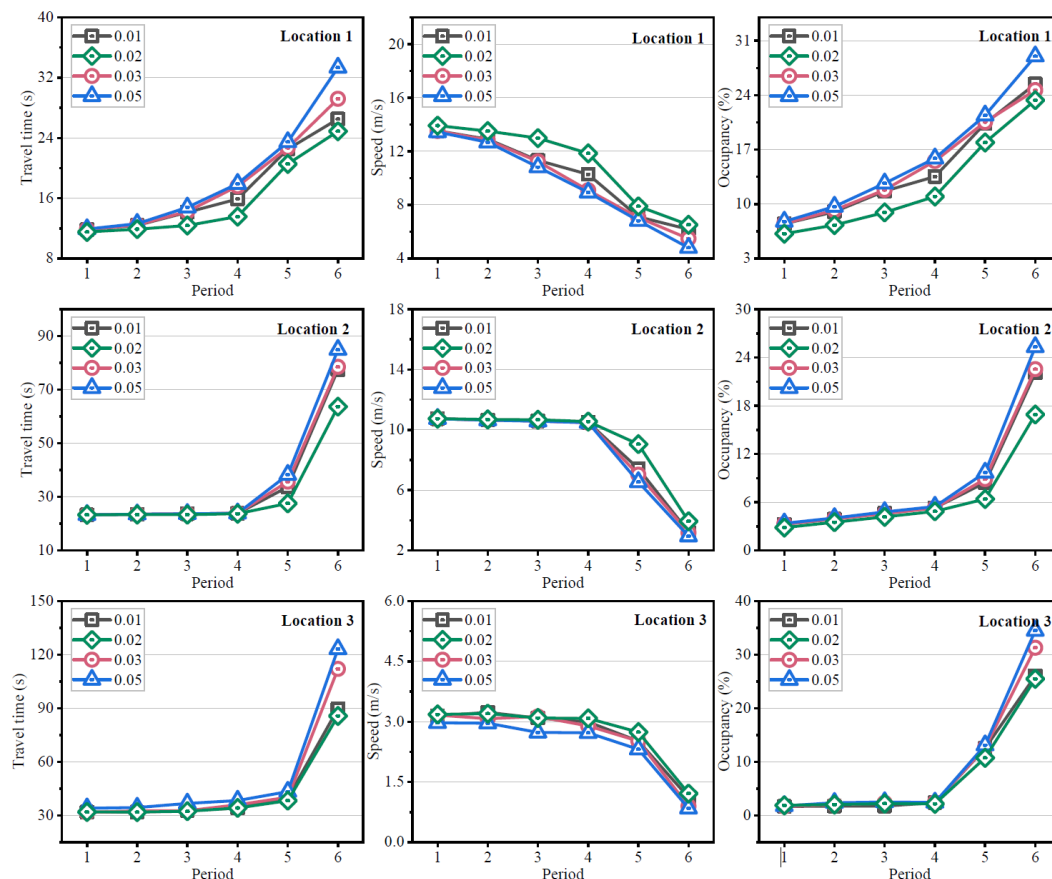


Figure 16. Effect of reward scaling rate on traffic performance.

As shown in Figure 17, we further probe asynchronous scheduling by lengthening control cycles from the default RM 5 s and VSL 30 s to RM {8, 10, 12} s and VSL {35, 40, 45} s. Results exhibit consistent monotonic degradation across locations and periods: longer cycles reduce feedback bandwidth and delay disturbance rejection, yielding higher travel time, lower speed, and higher occupancy, most notably in Periods 5–6. The degradation is more pronounced for RM than for VSL—lengthening the RM cycle directly lowers the frequency of inflow regulation, allowing transient merge-area disturbances to accumulate and increasing spillback risk. By contrast, VSL is more tolerant to cycle length yet still shows slower recovery and reduced steady-state efficiency under high demand. Therefore, fast, localized inflow perturbations require higher-frequency closed-loop regulation to avoid queue growth and shockwave amplification.

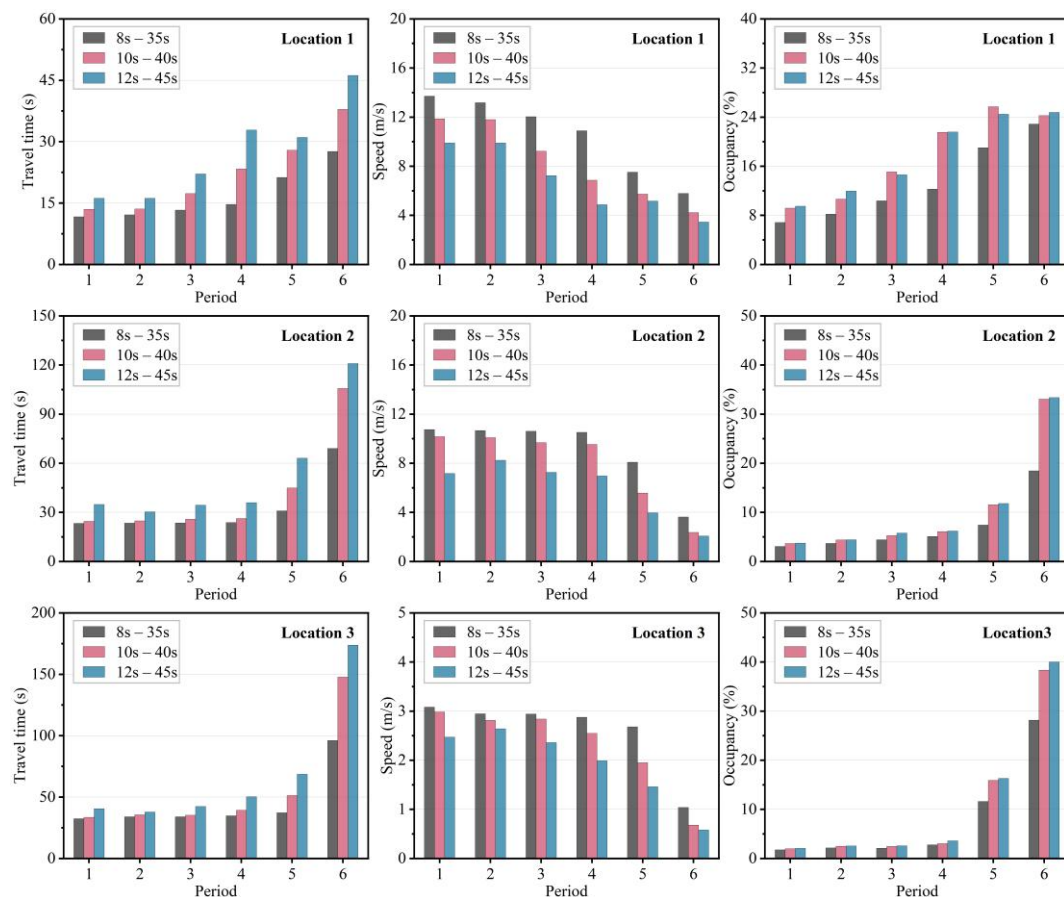


Figure 17. Effect of asynchronous control cycles on traffic performance.

6. Conclusions

Urban expressway entrance–arterial interfaces are critical bottlenecks in metropolitan traffic networks, where variable speed limits, ramp metering, and signal control interact across different spatial and temporal scales. Their uncoordinated operation often leads to congestion spillback, shockwave propagation, and inefficient capacity utilization. To address these challenges, this study develops a cooperative control strategy based on the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm. The proposed framework formulates the interface as a fully cooperative multi-agent problem; integrates unified state–action representation, decomposed rewards, and asynchronous scheduling; and is implemented in a simulation-driven environment to evaluate its effectiveness and real-time feasibility.

Compared with conventional single-strategy baselines, the learned policy consistently improves system-level performance. It smooths mainline traffic flow, alleviates ramp spillback, and stabilizes intersection operations. Mechanistic analysis reveals that coordinated adjustments—linking upstream speed regulation, ramp metering, and adaptive signal timing—create a dynamic buffer that weakens congestion propagation and enhances throughput across the interface.

This study can be enhanced by addressing the following research tasks. Future research will extend the proposed framework to multi-intersection interface corridors by incorporating intersection-to-intersection interactions and additional coordination layers under heterogeneous spacing and control settings. More systematic ablation and weight-sensitivity analyses of the shaped reward components can be explored as future work to further quantify the relative impacts of different penalties under broader operating con-

ditions. More comprehensive statistical robustness evaluations based on repeated simulations with different random seeds and demand realizations can be investigated as future work to quantify performance variability. Future work may benchmark the proposed framework against contemporary RL/MPC-based controllers under unified state/action design and timing assumptions to further examine algorithm-level differences. The lightweight inference structure and asynchronous execution design further indicate that the proposed method is suitable for real-time deployment under realistic traffic management system constraints.

Author Contributions: Conceptualization, S.W. and Z.W.; methodology, S.W., Z.W. and W.Y.; software, Z.W. and W.Y.; validation, S.W. and W.Y.; formal analysis, S.W. and W.Y.; investigation, Z.W.; resources, S.W.; data curation, W.Y.; writing—original draft preparation, S.W. and Z.W.; writing—review and editing, S.W. and Z.W.; visualization, S.W. and W.Y.; supervision, S.W.; project administration, S.W.; funding acquisition, S.W. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the National Natural Science Foundation of China (grant no. 52402401), China Postdoctoral Science Foundation (grant no. 2024M750440), Jiangsu Funding Program for Excellent Postdoctoral Talent (grant no. 2024ZB073).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Yuan, T.; Ioannou, P.A. Integrated freeway traffic control using Q-learning with adjacent arterial traffic considerations. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 7655–7666.
2. Chen, S.; Mao, B.; Liu, S.; Sun, Q.; Wei, W.; Zhan, L. Computer-aided analysis and evaluation on ramp spacing along urban expressways. *Transp. Res. Part C Emerg. Technol.* **2013**, *36*, 381–393.
3. Wang, L.; Abdel-Aty, M.; Lee, J.; Shi, Q. Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors. *Accid. Anal. Prev.* **2019**, *122*, 378–384.
4. Peng, T.; Xu, X.; Li, Y.; Wu, J.; Li, T.; Dong, X.; Cai, Y.; Wu, P.; Ullah, S. Enhancing expressway ramp merge safety and efficiency via spatiotemporal cooperative control. *IEEE Access* **2025**, *13*, 25664–25682.
5. Xu, Z.; Zheng, Y.; Li, Y. Coordinated control of urban expressways and connecting intersection based on genetic algorithm. In *Proceedings of the 2024 9th International Conference on Computer and Communication Systems (ICCCS), Xi'an, China, 19–22 April 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 1132–1138.
6. Ma, J.; Zeng, Y.; Chen, D. Ramp spacing evaluation of expressway based on entropy-weighted TOPSIS estimation method. *Systems* **2023**, *11*, 139.
7. Papageorgiou, M. A new approach to time-of-day control based on a dynamic freeway traffic model. *Transp. Res. Part B Methodol.* **1980**, *14*, 181–196.
8. Papageorgiou, M.; Habib, H.S.; Blosseville, J.M. ALINEA: A local feedback control law for on-ramp metering. *Transp. Res. Rec.* **1991**, *1320*, 58–64.
9. Yue, W.; Yang, H.; Li, M.; Wang, Y.; Zhou, Y.; Zheng, P. Hierarchical control based on ramp metering and variable speed limit for port motorway. *Systems* **2025**, *13*, 446.
10. Cheng, R.; Lou, H.; Wei, Q. Analysis of the impact for mixed traffic flow based on the time-varying model predictive control. *Systems* **2025**, *13*, 481.
11. Smulders, S. Control of freeway traffic flow by variable speed signs. *Transp. Res. Part B Methodol.* **1990**, *24*, 111–132.
12. Ma, C.; Guo, J.; Zhao, Y. Variable speed limit control strategy at freeway tunnel entrance based on cooperative lane changing. *Phys. A Stat. Mech. Its Appl.* **2023**, *620*, 128768.
13. Guo, H.; Jia, H.; Wu, R.; Huang, Q.; Tian, J.; Liu, C.; Wang, X. Variable speed limits for mixed traffic flow with connected autonomous vehicles: A reinforcement learning approach. *J. Transp. Eng. Part A Syst.* **2024**, *150*, 04024024.

14. Jin, Z.; Ma, M.; Liang, S.; Yao, H. Differential variable speed limit control strategy consider lane assignment at the freeway lane drop bottleneck. *Phys. A Stat. Mech. Its Appl.* **2024**, *633*, 129366.
15. Sun, R.; Hu, J.; Xie, X.; Zhang, Z. Variable speed limit design to relieve traffic congestion based on cooperative vehicle infrastructure system. *Procedia Soc. Behav. Sci.* **2014**, *138*, 427–438.
16. Yu, R.; Abdel-Aty, M. An optimal variable speed limits system to ameliorate traffic safety risk. *Transp. Res. Part C Emerg. Technol.* **2014**, *46*, 235–246.
17. Khondaker, B.; Kattan, L. Variable speed limit: A microscopic analysis in a connected vehicle environment. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 146–159.
18. Han, Y.; Wang, M.; He, Z.; Li, Z.; Wang, H.; Liu, P. A linear Lagrangian model predictive controller of macro- and micro-variable speed limits to eliminate freeway jam waves. *Transp. Res. Part C Emerg. Technol.* **2021**, *128*, 103121.
19. Ding, H.; Zhang, L.; Chen, J.; Zheng, X.; Pan, H.; Zhang, W. MPC-based dynamic speed control of CAVs in multiple sections upstream of the bottleneck area within a mixed vehicular environment. *Phys. A Stat. Mech. Its Appl.* **2023**, *613*, 128542.
20. Zhang, L.; Ding, H.; Feng, Z.; Wang, L.; Di, Y.; Zheng, X.; Wang, S. Variable speed limit control strategy considering traffic flow lane assignment in mixed-vehicle driving environment. *Phys. A Stat. Mech. Its Appl.* **2024**, *656*, 130216.
21. Iordanidou, G.-R.; Roncoli, C.; Papamichail, I.; Papageorgiou, M. Feedback-based mainstream traffic flow control for multiple bottlenecks on motorways. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 610–621.
22. Qiu, S.; Li, Z.; Pang, Z.; Li, Z.; Tao, Y. Multi-agent optimal control for central chiller plants using reinforcement learning and game theory. *Systems* **2023**, *11*, 136.
23. Karalakou, A.; Troullinos, D.; Chalkiadakis, G.; Papageorgiou, M. Deep reinforcement learning reward function design for autonomous driving in lane-free traffic. *Systems* **2023**, *11*, 134.
24. Jin, J.; Huang, H.; Li, Y.; Dong, Y.; Zhang, G.; Chen, J. Variable speed limit control strategy for freeway tunnels based on a multi-objective deep reinforcement learning framework with safety perception. *Expert Syst. Appl.* **2025**, *267*, 126277.
25. Chow, A.H.F.; Su, Z.C.; Liang, E.M.; Zhong, R.X. Adaptive signal control for bus service reliability with connected vehicle technology via reinforcement learning. *Transp. Res. Part C Emerg. Technol.* **2021**, *129*, 103264.
26. Pang, M.; Yang, M. Coordinated control of urban expressway integrating adjacent signalized intersections based on pinning synchronization of complex networks. *Transp. Res. Part C Emerg. Technol.* **2020**, *116*, 102645.
27. Deng, M.; Chen, F.; Gong, Y.; Li, X.; Li, S. Optimization of signal timing for urban expressway exit ramp connecting intersection. *Sensors* **2023**, *23*, 6884.
28. Cheng, M.; Zhang, C.; Jin, H.; Wang, Z.; Yang, X. Adaptive coordinated variable speed limit between highway mainline and on-ramp with deep reinforcement learning. *J. Adv. Transp.* **2022**, *2022*, 2435643.
29. He, Z.; Han, Y.; Yu, H.; Bai, L.; Guo, W.; Liu, P. Integrated feedback perimeter control-based ramp metering and variable speed limits for multibottleneck freeways. *J. Transp. Eng. Part A Syst.* **2024**, *150*, 04024054.
30. Hegyi, A.; De Schutter, B.; Hellendoorn, H. Model predictive control for optimal coordination of ramp metering and variable speed limits. *Transp. Res. Part C Emerg. Technol.* **2005**, *13*, 185–209.
31. Deng, F.; Jin, J.; Shen, Y.; Du, Y. A dynamic self-improving ramp metering algorithm based on multi-agent deep reinforcement learning. *Transp. Lett.* **2024**, *16*, 649–657.
32. Jin, J.; Li, Y.; Huang, H.; Dong, Y.; Liu, P. A variable speed limit control approach for freeway tunnels based on the model-based reinforcement learning framework with safety perception. *Accid. Anal. Prev.* **2024**, *201*, 107570.
33. Han, Y.; Hegyi, A.; Zhang, L.; He, Z.; Chung, E.; Liu, P. A new reinforcement learning-based variable speed limit control approach to improve traffic efficiency against freeway jam waves. *Transp. Res. Part C Emerg. Technol.* **2022**, *144*, 103900.
34. Lu, W.; Yi, Z.; Gu, Y.; Rui, Y.; Ran, B. TD3LVSL: A lane-level variable speed limit approach based on twin delayed deep deterministic policy gradient in a connected automated vehicle environment. *Transp. Res. Part C Emerg. Technol.* **2023**, *153*, 104221.
35. Pooladsanj, M.; Savla, K.; Ioannou, P.A. Ramp metering to maximize freeway throughput under vehicle safety constraints. *Transp. Res. Part C Emerg. Technol.* **2023**, *154*, 104267.
36. Wang, T.; Zhu, Z.; Zhang, J.; Tian, J.; Zhang, W. A large-scale traffic signal control algorithm based on multi-layer graph deep reinforcement learning. *Transp. Res. Part C Emerg. Technol.* **2024**, *162*, 104582.
37. Bie, Y.; Ji, Y.; Ma, D. Multi-agent deep reinforcement learning collaborative traffic signal control method considering intersection heterogeneity. *Transp. Res. Part C Emerg. Technol.* **2024**, *164*, 104663.

38. Song, X.B.; Zhou, B.; Ma, D. Cooperative traffic signal control through a counterfactual multi-agent deep actor critic approach. *Transp. Res. Part C Emerg. Technol.* **2024**, *160*, 104528.
39. Lin, Q.; Huang, W.; Wu, Z.; Zhang, M.; He, Z. Multi-Agent Game Theory-Based Coordinated Ramp Metering Method for Urban Expressways With Multi-Bottleneck. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 3643–3658.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.