

Orchestrating Reasoning and Reaction: An Asynchronous Hierarchical Framework for LLM-driven Traffic Signal Control

Fansheng Sun
Hong Kong University of Science and
Technology (Guangzhou)
Guangzhou, Guangdong, China
Xi'an Jiaotong University
Xi'an, Shanxi, China
sunfansheng@stu.xjtu.edu.cn

Jiyu Wang
Hong Kong University of Science and
Technology (Guangzhou)
Guangzhou, Guangdong, China
jwang738@connect.hkust-gz.edu.cn

Zhidan Liu*
Hong Kong University of Science and
Technology (Guangzhou)
Guangzhou, Guangdong, China
zhidanliu@hkust-gz.edu.cn

Abstract

Network-wide coordinated Traffic Signal Control (TSC) is critical for enhancing urban mobility. However, existing approaches face a fundamental trade-off: traditional Multi-Agent Reinforcement Learning (MARL) is often hindered by a myopic observational scope, while Large Language Model (LLM) agents are constrained by high inference costs and spatial-topological hallucinations. To address these limitations, we present *Astra*, an Asynchronous Synergistic Traffic Regulation Architecture that decouples high-level strategic reasoning from reactive execution through a three-layer hierarchy. Specifically, the Macro layer performs low-frequency strategic inference to identify global bottlenecks, which are then partitioned into regional congestion subgraphs. The Meso layer employs a topological semantic causal logic mechanism to ground LLM-based coordination in physical reality via semantic primitives. Simultaneously, the Micro layer governs the broader network using decentralized MARL agents optimized with spatial attention and auxiliary prediction for robust, high-frequency execution. To bridge the gap between reasoning depth and real-time constraints, *Astra* incorporates an asynchronous synergetic protocol featuring strategic locking for stability and proactive feedback for adaptive re-planning. Extensive evaluations on real-world datasets demonstrate that *Astra* consistently outperforms state-of-the-art baseline methods, offering superior efficiency, robustness, and generalization across diverse urban scenarios.

CCS Concepts

• **Computing methodologies** → **Artificial intelligence**.

Keywords

Traffic Signal Control; Multi-Agent System; Large Language Model; Reinforcement Learning

ACM Reference Format:

Fansheng Sun, Jiyu Wang, and Zhidan Liu. 2026. Orchestrating Reasoning and Reaction: An Asynchronous Hierarchical Framework for LLM-driven Traffic Signal Control. In *Proceedings of the 32nd ACM SIGKDD Conference*

*Corresponding author: Zhidan Liu (Email: zhidanliu@hkust-gz.edu.cn).



This work is licensed under a Creative Commons Attribution 4.0 International License. *KDD '26, Jeju Island, Republic of Korea*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2259-2/2026/08
<https://doi.org/10.1145/3770855.3818039>

on *Knowledge Discovery and Data Mining V.2 (KDD '26)*, August 09–13, 2026, Jeju Island, Republic of Korea. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3770855.3818039>

1 Introduction

Traffic Signal Control (TSC) is fundamental for modern urban management, governing the spatiotemporal allocation of limited road resources to optimize network throughput, efficiency, and safety [11, 22]. As urbanization accelerates, the complexity of traffic dynamics has surpassed the capabilities of traditional rule-based heuristics (e.g., FixedTime and MaxPressure) [19], necessitating intelligent, data-driven solutions [29, 32].

The advent of Reinforcement Learning (RL) has advanced adaptive control by enabling agents to learn optimal policies at individual intersections through continuous environmental interaction [15, 24, 30]. However, isolated intersection optimization is insufficient for network-wide efficiency. Multi-Agent Reinforcement Learning (MARL) addresses this by enabling coordination among intersections through graph attention mechanisms to integrate neighborhood information [16, 21, 26]. Yet, MARL methods are often constrained by a local observational scope, hindering their ability to capture macroscopic dynamics like tidal flows or cascading gridlocks. Furthermore, agents tend to greedily optimize local rewards at the expense of genuine global collaboration [2].

Recently, Large Language Models (LLMs) have emerged as a promising alternative, offering sophisticated reasoning capabilities and extensive world knowledge [8, 12, 25]. Unlike trial-and-error RL, LLMs enable zero-shot strategic planning and interpretable policies for complex scenarios [5, 10, 12, 33?]. However, directly applying LLMs to TSC faces significant hurdles. Despite their strategic insight, LLMs suffer from prohibitive inference overheads (e.g., latency and computational costs) and “spatial hallucinations” regarding graph structures. These issues are exacerbated when processing long contextual information for multi-intersection control, undermining the real-time responsiveness required in dynamic scenarios [31]. Consequently, a critical mismatch emerges: *LLMs offer high-level reasoning but lack real-time agility, whereas MARL excels at rapid execution but lacks long-term coordination depth*.

To bridge this gap, we propose to integrate the deliberative reasoning of LLMs with the reactive agility of MARL. However, achieving a seamless synergy in a hyper-dynamic environment entails three fundamental challenges:

(C1) *Computational burden vs. Real-time response*. Effective TSC demands rapid responsiveness to capture transient fluctuations.

However, LLM generation incurs substantial computational costs and inference latency. This creates a dual bottleneck: frequent invocations are resource-prohibitive, while the inherent latency renders high-level guidance obsolete by the time it is generated.

(C2) Spatial-topological mismatch. Road networks are defined by complex, non-Euclidean topological structures. While proficient in processing sequential text, LLMs lack a native inductive bias for graph-structured data. This often leads to “spatial hallucinations”, where LLMs propose collaborative strategies that violate physical propagation constraints, resulting in invalid instructions.

(C3) Asynchronous coordination instability. Integrating heterogeneous agents across disparate spatiotemporal scales requires robust orchestration. Naive coupling, where an LLM directly dictates MARL actions, frequently leads to system oscillation. Without stabilization, frequent strategy switching between high-level reasoning and low-level reaction can cause erratic signal behavior.

To systematically address these challenges, we present **Astra**, an **Asynchronous Synergistic Traffic Regulation Architecture**. Astra adopts a *hierarchical Macro-Meso-Micro design* to decouple decision-making logic from execution frequency. The Macro layer functions as a strategic monitor, leveraging the LLM for low-frequency global perception to identify overarching congestion patterns, thereby bypassing the bottleneck of computational costs and inference latency (**Addressing C1**) [6]. The Meso layer serves as a semantic bridge: it introduces a *topological semantic causal logic mechanism* to partition the network into topological primitives. By transforming graph connectivity into causal reasoning chains, it grounds strategic instructions in physical reality (**Addressing C2**). The Micro layer deploys *enhanced MARL agents* featuring spatial attention and auxiliary prediction tasks for robust, high-frequency execution. To unify these layers, we implement an *adaptive asynchronous synergetic protocol* with strategic locking and proactive feedback mechanisms. This bi-directional loop ensures that high-level coordination remains stable over a decision window while allowing bottom-up alerts to trigger re-planning during sudden traffic anomalies (**Addressing C3**).

Our main contributions are summarized as follows:

- We propose Astra, a novel hierarchical synergy framework that decouples strategic planning and reactive control across three layers, achieving a balance between reasoning depth and operational efficiency.
- We introduce a topological semantic causal logic mechanism that translates complex graph topologies into structured causal reasoning primitives, effectively bridging the gap between LLM semantic understanding and road network topology.
- We establish an asynchronous coordination protocol that features strategic locking and proactive feedback, providing a robust closed-loop solution to synchronize heterogeneous decision scales.
- Extensive experiments on diverse real-world datasets demonstrate that Astra significantly outperforms state-of-the-art baselines, offering superior generalization and robustness in complex urban scenarios.

2 Related Work

Traffic Signal Control. The traffic signal control (TSC) paradigm has shifted from rule-based strategies to data-driven approaches [22]. Early approaches such as FixedTime [11] rely on predefined cycle lengths and phase sequences designed by domain experts, while adaptive methods like MaxPressure [19] introduce dynamic control by minimizing queue pressure differences between adjacent intersections. Although efficient in constrained settings, these methods lack the flexibility to handle highly dynamic traffics, limiting their applicability in complex real-world networks. To enhance adaptability and local coordination, multi-agent reinforcement learning (MARL) [7] has been widely adopted. Representative works like IntelliLight [24] and MPLight [3] optimize individual intersection efficiency, while CoLight [21] utilizes graph attention mechanisms to aggregate neighborhood information for regional coordination. Subsequent efforts, including PressLight [20] and Advanced-MPLight [30], further improve throughput by explicitly encoding lane-level traffic states. Despite these advances, MARL-based methods often remain constrained by a limited observational scope, causing susceptibility to local optima when scaled to large networks [23].

The emergence of Large Language Models (LLMs) has introduced powerful reasoning capabilities to TSC tasks [12]. For instance, LLMLight [12] demonstrates the potential of pre-trained models in zero-shot TSC scenarios, while CoLLMLight [?] explores LLM-based multi-agent collaboration for network-level control. Traffic-R1 [33] further augments reasoning depth through reinforcement learning fine-tuning. However, these LLM-driven methods still face fundamental limitations in spatial understanding of road topologies and often incur high inference costs, especially when synchronously coordinating large numbers of intersections via long contextual interactions. These challenges constrain their practicality.

Distinct from existing LLM-based TSC solutions, Astra harmonizes LLM-driven reasoning with MARL-based reaction through a hierarchical synergy. It bridges the spatial modality gap via topological semantic causal grounding and mitigates inference latency via an asynchronous orchestration mechanism, ensuring physically valid coordination with minimal computational overhead.

Multi-Agent Systems and Coordination. Multi-agent systems (MAS) have been widely applied in domains such as smart grids [14], swarm robotics [18], and intelligent transportation [13], where coordinated behavior is essential for solving large-scale problems. Effective coordination is the cornerstone of MAS efficiency [8]. Previous research has explored the assignment of specialized roles to regulate system behavior [10]. Within transportation in particular, the focus has pivoted toward balancing decision precision with response speed [27]. Recent attempts to mitigate computational pressure include hierarchical processing and asynchronous protocols [31]. By combining advanced state encoding, such as the pressure representation in Advanced-XLight [30], with topology-aware resource allocation, systems can achieve robust coordination without sacrificing real-time performance. Our work contributes to this trajectory by proposing a hierarchical Macro-Meso-Micro architecture. Unlike existing synchronous or flat coordination models, our approach orchestrates LLMs and MARL asynchronously, providing a scalable blueprint for network-wide cooperative control that maintains both reasoning depth and operational agility.

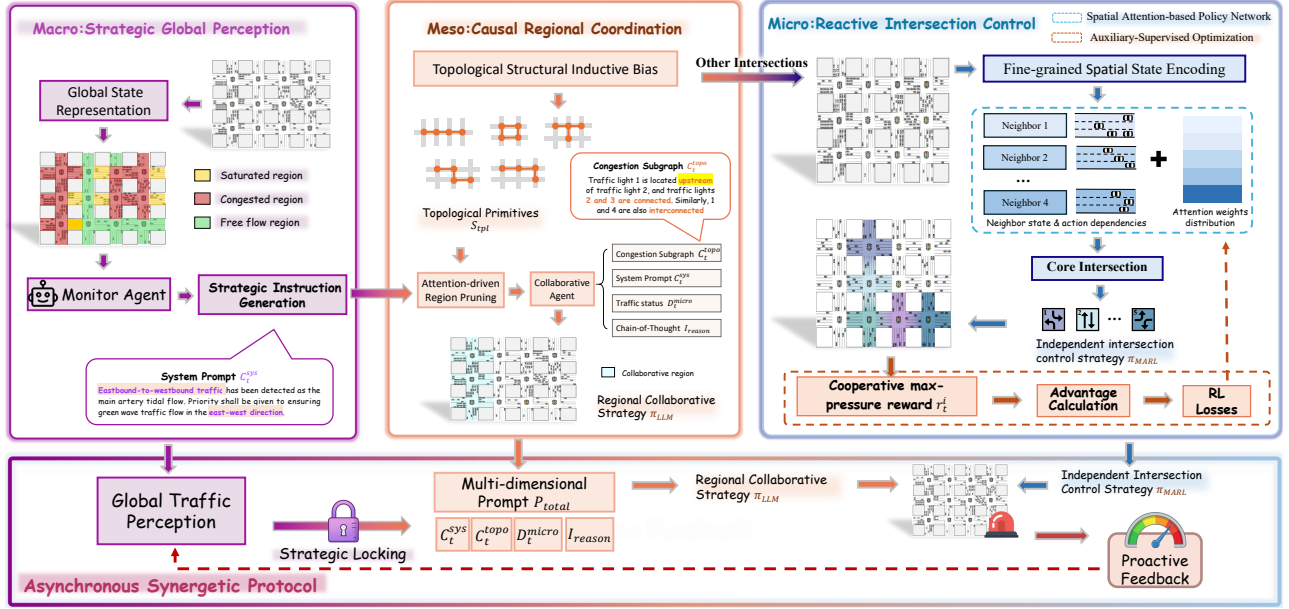


Figure 1: Architectural overview of Astra. Operating across a three-layer asynchronous hierarchy, the framework orchestrates strategic global perception (Macro), causal regional coordination (Meso), and reactive intersection control (Micro). An integrated asynchronous synergetic protocol synchronizes these layers to harmonize long-term planning with immediate execution.

3 Problem Statement

In this section, we formally define the core concepts and formulate the task of network-wide coordinated traffic signal control.

Definition 1 (ROAD NETWORK). The road network is modeled as a directed graph $G = (V, L)$, where V denotes the set of intersections (nodes) and L denotes the set of directed lanes (edges). Each lane $\ell \in L$ is characterized by its turning function, such that $L = \{L_{go} \cup L_{left} \cup L_{right}\}$, representing go-through, left-turn, and right-turn lanes, respectively.

Definition 2 (TRAFFIC STATE). For each intersection $i \in V$, the traffic state at a given signal-switching time step t is denoted by $h_t^i \in \mathbb{R}^d$, which captures localized features such as traffic pressure, queue lengths, or congestion indices. The global traffic state of the entire network is represented by the matrix $H_t = [h_t^1, h_t^2, \dots, h_t^{|V|}]^T$.

Definition 3 (CONGESTION SUBGRAPH). A congestion subgraph $\mathcal{R}_t^* \subseteq G$ is induced by a cluster of intersections whose aggregate congestion metrics exceed a predefined threshold at time t . Formally, \mathcal{R}_t^* identifies high-impact regional bottlenecks that require coordinated intervention to prevent network-wide gridlock.

Definition 4 (TRAFFIC SIGNAL). At each time step t , a control agent assigns a phase a_t^i to intersection i from a discrete action space $A = \{a_1, \dots, a_m\}$. A phase $a \in A$ is defined as a set of non-conflicting movements, denoted as $a = \text{set}(L_{allow})$, where permitted lanes receive a green signal while others are set to red.

Problem Statement (NETWORK-WIDE COORDINATED TRAFFIC SIGNAL CONTROL). Given a road network G , a multi-agent system (MAS) is deployed to control traffic signals under a policy Π . At each time step t , the MAS takes as input a user prompt P_{prompt} , current network-wide traffic state H_t , the topological structure of

G , and identified congestion subgraph \mathcal{R}_t^* . It then outputs coordinated signal actions a_t^i for each intersection $i \in V$, such that the network-wide traffic efficiency is maximized:

$$a_t^i = \Pi(P_{prompt}, H_t, G, \mathcal{R}_t^*). \quad (1)$$

The objective is thus to minimize the cumulative network-level congestion or, equivalently, minimize average vehicle delay.

4 Methodology

As illustrated in Figure 1, we propose Astra, a hierarchical framework that orchestrates a joint policy $\Pi = \{\pi_{LLM}, \pi_{MARR}\}$ across three heterogeneous scales. To resolve the conflict between reasoning depth and efficient response, Astra decouples the control process: the **Macro layer** performs low-frequency strategic inference to identify global bottlenecks; the **Meso layer** leverages a novel topological semantic causal logic to facilitate π_{LLM} in generating spatially-aware coordination instructions; and the **Micro layer** executes high-frequency reactive control via decentralized π_{MARR} agents. These layers are coordinated via an **asynchronous synergetic protocol**, which employs strategic locking for stability and proactive feedback for adaptive re-planning, ensuring scalable network-wide coordination.

4.1 Strategic Global Perception

We instantiate a monitor agent \mathcal{F}_{LLM}^{mon} at the Macro layer to infer global traffic patterns and pinpoint critical bottleneck regions.

4.1.1 Global State Representation. For each intersection $i \in V$, we construct a comprehensive traffic state index h_t^i by fusing the

normalized queue length q_{\max}^i and average speed s^i :

$$h_t^i = \alpha \cdot \frac{q_{\max}^i}{Q_{cap}} + (1 - \alpha) \cdot \left(1 - \frac{s^i}{s_{\max}}\right), \quad (2)$$

where Q_{cap} and s_{\max} denote lane capacity and speed limit, respectively, and α is a weighting factor. This hybrid index effectively characterizes the intersections' dynamic throughput and static accumulation, preventing the neglect of high-speed but saturated flows. Aggregating these indices yields the network-wide traffic state matrix H_t .

4.1.2 Strategic Instruction Generation. Leveraging the zero-shot reasoning capability of LLMs, the monitor agent \mathcal{F}_{LLM}^{mon} analyzes H_t to detect macroscopic patterns (e.g., tidal flows). It takes H_t and a domain-knowledge task descriptor P_{task} as input to generate a global strategic instruction C_t^{sys} :

$$C_t^{sys} = \mathcal{F}_{LLM}^{mon}(H_t, P_{task}). \quad (3)$$

Notably, C_t^{sys} encapsulates high-level directives and explicitly identifies candidate congestion regions, which are transmitted to the Meso layer to guide targeted regional intervention.

4.2 Causal Regional Coordination

The Meso layer serves as the semantic bridge, employing a collaborative agent \mathcal{F}_{LLM}^{co} to reason about intersections within critical bottlenecks. To address LLMs' spatial hallucinations, we introduce a **topological semantic causal logic mechanism**, which grounds reasoning in physical constraints via three steps: *subgraph extraction*, *region pruning*, and *causal decision-making*.

4.2.1 Topological Structural Inductive Bias. Traffic congestion typically manifests in specific geometric patterns that carry significant physical constraints. To capture these, we define a set of *topological primitives* $S_{tpl} = \{\tau_L, \tau_T, \tau_Z, \tau_O, \tau_I\}$, where each primitive $\tau \in S_{tpl}$ is a template graph consisting of exactly K intersections (i.e., $|V(\tau)| = K$). This parameter K serves as the granularity index, determining the receptive field of regional coordination. The set S_{tpl} functions as a minimal sufficient basis for synthesizing complex urban configurations:

- τ_L/τ_T (*couplings*): Model high-correlation turning or merging flows at K -intersection local junctions.
- τ_Z/τ_I (*corridors*): Represent staggered intersections or arterial segments prone to spillover effects across K consecutive intersections.
- τ_O (*cyclic loop*): Identifies closed-loop structures of size K that are inherently susceptible to gridlock deadlocks.

To facilitate real-time inference, we utilize a Breadth-First Search (BFS) based subgraph isomorphism algorithm [4, 9] to extract a candidate set \mathbb{C} from the global road network G :

$$\mathbb{C} = \{\mathcal{R} \subseteq G \mid |V(\mathcal{R})| = K \wedge \exists \tau \in S_{tpl}, \mathcal{R} \cong \tau\}, \quad (4)$$

where $\mathcal{R} \cong \tau$ denotes topological isomorphism. urban networks are typically sparse planar graphs (e.g., $d \leq 4$), fixing K strictly bounds the BFS search complexity to $\mathcal{O}(|V| \cdot d^K)$, effectively avoiding the combinatorial explosion associated with dynamic subgraph sizes. By pre-computing set \mathbb{C} offline, Astra achieves zero online search overhead, allowing the system to instantly map real-time congestion to its corresponding topological priors.

Furthermore, this extraction mechanism scales seamlessly to large networks with concurrent congestion. We prioritize subgraphs by congestion severity and resolve spatial overlaps through a strict exclusivity rule: shared intersections solely execute the strategy of the highest-scoring subgraph. This straightforward design efficiently eliminates command conflicts and erratic switching without requiring complex arbitration.

4.2.2 Attention-driven Region Pruning. To bridge the decision layers, the Meso layer adopts the global guidance C_t^{sys} as a spatial prior. Rather than conducting an exhaustive search across the entire network, we implement an attention-driven pruning strategy that dynamically narrows the candidate set \mathbb{C} based on the Macro-layer's directives. Specifically, at each time step t , we identify the primary bottleneck \mathcal{R}_t^* by maximizing the congestion intensity within the macro-identified Regions of Interest (RoIs). Let $V_{RoI} \subseteq V$ be the set of critical intersections extracted from the semantic instruction C_t^{sys} . The optimal congestion subgraph is formulated as:

$$\mathcal{R}_t^* = \arg \max_{\mathcal{R} \in \mathbb{C}_{ref}} \mathbb{I} \left(\sum_{i \in V(\mathcal{R})} h_t^i \geq \zeta \right) \cdot \sum_{i \in V(\mathcal{R})} h_t^i, \quad (5)$$

where $\mathbb{C}_{ref} = \{\mathcal{R} \in \mathbb{C} \mid V(\mathcal{R}) \cap V_{RoI} \neq \emptyset\}$ represents the subset of candidates spatially aligned with the macro-level directives, ζ is a severity threshold, and $\mathbb{I}(\cdot)$ is the indicator function.

This "focus-of-attention" design restricts the LLM's heavy computational footprint to the most critical subgraph, effectively pruning the reasoning complexity from the global network scale $\mathcal{O}(|V|)$ to a localized subgraph scale $\mathcal{O}(|V(\mathcal{R}^*)|)$. By selectively filtering non-essential areas, this on-demand pruning resolves the fundamental conflict between the depth of semantic reasoning and the timely responsiveness required for TSC tasks.

4.2.3 Causal-driven Collaborative Decision. To ensure physical validity, the Meso layer transforms the subgraph \mathcal{R}_t^* into a structured causal reasoning task. We formulate a multi-dimensional prompt:

$$P_{total} = (C_t^{sys} \oplus C_t^{topo} \oplus D_t^{micro} \oplus I_{reason}), \quad (6)$$

where \oplus denotes semantic concatenation. Specifically, C_t^{topo} maps the topological connectivity of \mathcal{R}_t^* into natural language as directed causal dependencies (e.g., "upstream $i \in V$ dictates inflow at $j \in V$ "), grounding spatial constraints in causal logic. D_t^{micro} provides real-time snapshots of intersection states (e.g., queues, active phases), while I_{reason} instantiates a *Causal-Chain-of-Thought* module. This module guides the agent to trace congestion to its root causes (e.g., arterial spillover) before generating interventions. Finally, the collaborative agent \mathcal{F}_{LLM}^{co} infers the regional strategy:

$$\pi_{LLM} = \mathcal{F}_{LLM}^{co}(P_{total}). \quad (7)$$

This strategy π_{LLM} serves as a synchronized coordination prior, providing high-level guidance for the subsequent micro-layer execution to resolve spatially-propagated traffic conflicts.

4.3 Reactive Intersection Control

While the Meso layer concentrate on critical bottlenecks, the Micro layer governs the remaining intersections to maintain network-wide stability. We model this distinct task as a Decentralized Partially Observable Markov Decision Process (Dec-POMDP) [1], formally defined in Appendix A.1. To ensure robustness against dynamic traffic flows, we enhance the standard Proximal Policy Optimization (PPO) algorithm [17] with *fine-grained state encoding*, *spatial attention mechanism*, and *auxiliary supervision*.

4.3.1 Fine-grained Spatial State Encoding. To capture precise vehicle distributions beyond scalar queue counts, we employ a zone-based pressure encoding. As shown in Equation (8), we discretize each incoming lane $\ell \in L_{in}^i$ into three functional zones, *i.e.*, $Z_i = \{\text{queuing, buffering, distant}\}$. The state of lane ℓ is formulated as a weighted pressure value x_ℓ^t , aggregating both vehicle density and kinematic status:

$$x_\ell^t = \sum_{v \in \mathbb{V}_\ell} \left(\sum_{j=1}^3 \mathbb{I}(v \in Z_j) \cdot \omega_j + \beta \cdot (1 - \tilde{s}_v) \right), \quad (8)$$

where \mathbb{V}_ℓ is the vehicle set of lane ℓ , ω_j assigns higher weights to vehicles closer to the stop line, and the speed factor β penalizes stagnation given the normalized speed \tilde{s}_v . The local observation o_ℓ^t concatenates these pressure values with the current phase, providing a high-resolution spatial footprint for decision-making.

4.3.2 Spatial Attention-based Policy Network. To achieve implicit collaboration among decentralized agents, we integrate a graph attention mechanism into the policy architecture. Instead of processing intersections in isolation, the encoder aggregates neighborhood information to generate a context-aware embedding z_i^t . By computing learnable attention weights $\varphi_{i,j}^t$ between intersection i and its neighbors \mathcal{N}_i , the feature representation is refined as:

$$z_i^t = \text{LayerNorm} \left(\mathbf{W}_e \cdot o_i^t + \sum_{j \in \mathcal{N}_i} \varphi_{i,j}^t \cdot \mathbf{W}_v \cdot o_j^t \right), \quad (9)$$

where $\mathbf{W}_e, \mathbf{W}_v$ are learnable projection matrices. This mechanism allows the agent to perceive arterial spillover risks from neighbors, enabling the Actor $\pi_{MARL}(a_t^i | z_i^t)$ to make cooperative phase selections without explicit communication overhead. More details about z_i^t and $\varphi_{i,j}^t$ are in Appendix A.2.

4.3.3 Auxiliary-Supervised Optimization. To accelerate convergence and enhance representation learning, we augment the PPO objective with an auxiliary supervision task. The total loss \mathcal{L}_{total} incorporates a predictive modeling term:

$$\mathcal{L}_{total} = \mathcal{L}_{clip} + c_1 \cdot \mathcal{L}_{value} + c_2 \cdot \mathcal{L}_{entropy} + c_3 \cdot |\mathcal{F}_{pred}(z_i^t) - q_{t+1}|^2, \quad (10)$$

where c_1, c_2 and c_3 are coefficients. \mathcal{L}_{clip} , \mathcal{L}_{value} , and $\mathcal{L}_{entropy}$ are the standard policy, value function, and entropy losses, respectively. While the term $|\mathcal{F}_{pred}(z_i^t) - q_{t+1}|^2$, serving as an auxiliary prediction loss, forces the model to comprehend environmental dynamics by predicting the next-step queue state q_{t+1} . These loss functions are detailed in Appendix A.3.

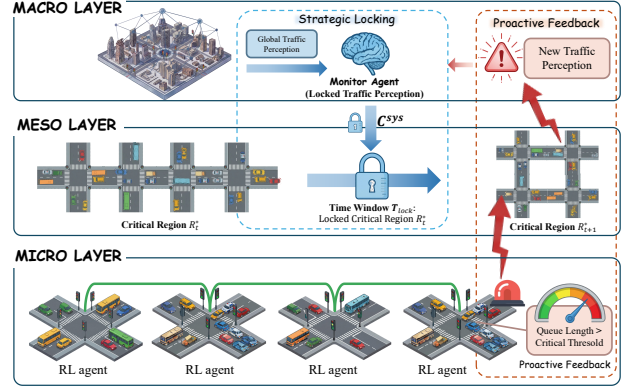


Figure 2: The asynchronous synergetic protocol orchestrates layer interactions via strategic locking and proactive feedback.

Furthermore, to align local actions with regional throughput, we design a cooperative max-pressure reward as:

$$r_t^i = -\gamma_{scale} \cdot \left(\sum_{\ell \in L_{in}^i} x_\ell^t + \lambda \cdot \frac{1}{|\mathcal{N}_i|} \cdot \sum_{j \in \mathcal{N}_i} \sum_{k \in \{L_{out}^i \cap L_{in}^j\}} x_k^t \right), \quad (11)$$

where L_{in}^i and L_{out}^i denote the upstream and downstream lanes of intersection i , respectively. By penalizing pressure accumulation in downstream lanes (weighted by λ) and scaling the total reward by γ_{scale} , this mechanism suppresses greedy behaviors and enforces a distributed logic of “upstream discharge, downstream absorption”.

4.4 Asynchronous Synergetic Protocol

To synchronize heterogeneous agents operating at varying temporal resolutions, Astra implements an adaptive **asynchronous synergetic protocol**. As illustrated in Figure 2, this protocol orchestrates layer interactions through a top-down *strategic locking* mechanism and a bottom-up *proactive feedback* loop.

4.4.1 Strategic Locking Mechanism. To maintain coordination stability, we grant the meso-level collaborative strategy π_{LLM} absolute priority within the identified bottleneck \mathcal{R}_i^* . Once the macro-meso reasoning is triggered, the system enforces a **strategic lock** that overrides local MARL actions for a fixed execution window T_{lock} :

$$a_t^i = \begin{cases} \pi_{LLM}(i), & \text{if } i \in V(\mathcal{R}_i^*) \text{ and } t \in [t_{start}, t_{start} + T_{lock}], \\ \pi_{MARL}(i), & \text{otherwise.} \end{cases} \quad (12)$$

By fixing the high-level coordination plan within T_{lock} , the protocol suppresses potential system oscillations caused by frequent strategy switching. This temporal decoupling allows the LLM to focus on long-term logical alignment while the MARL agents independently stabilize non-critical regions.

4.4.2 Proactive Feedback and Re-planning. To ensure the system remains responsive to sudden environmental shifts, we incorporate an event-driven **proactive feedback** mechanism. While the top-down guidance provides stability, underlying micro-agents continuously monitor local traffic volatility. If an agent detects that local congestion (*e.g.*, queue buildup) exceeds a predefined threshold, it proactively issues an interrupt to the Macro layer. This triggers an immediate re-planning cycle, forcing the monitor agent \mathcal{F}_{LLM}^{mon} to

Table 1: Performance comparison on network-wide traffic signal control tasks across five real-world datasets. Lower values for ATT and AWT indicate superior efficiency. The best results are highlighted in bold, and the second-best results are underlined.

Model	Jinan-1		Jinan-2		Jinan-3		Hangzhou-1		Hangzhou-2		Paradigm
	ATT	AWT	ATT	AWT	ATT	AWT	ATT	AWT	ATT	AWT	
FixedTime	284.41	439.58	386.98	206.99	404.09	245.11	513.86	315.09	427.98	279.92	Conventional Methods
Maxpressure	320.21	133.53	313.86	120.87	309.28	128.03	347.06	102.39	349.98	135.51	
Advanced Maxpressure	303.51	115.36	302.83	110.59	291.50	107.78	337.63	89.71	336.69	110.91	
MPLight	411.31±3.82	189.27±2.13	298.84±2.91	55.38±1.24	313.62±3.17	78.49±1.68	301.34±2.86	17.03±1.07	358.09±3.41	34.19±1.39	MARL-based Methods
Advanced MPLight	292.51±3.04	74.33±1.41	269.82±2.68	48.55±1.17	309.57±3.12	62.64±1.36	313.29±2.89	12.39±0.94	368.31±3.23	41.31±1.48	
Efficient MPLight	343.03±3.57	80.74±1.52	274.00±2.83	52.55±1.21	276.10±2.96	53.72±1.29	323.55±3.11	15.92±1.08	369.94±3.34	40.06±1.37	
CoLight	340.36±3.46	80.63±1.54	290.74±2.98	54.58±1.33	293.32±2.81	56.03±1.42	363.71±3.63	51.34±1.87	458.70±4.16	114.13±2.61	
Advanced-CoLight	294.29±3.17	66.25±1.46	276.51±2.78	47.32±1.22	281.01±2.84	51.90±1.31	315.05±2.31	23.69±1.17	357.20±3.14	77.19±1.83	
PressLight	455.35±4.61	147.43±2.87	319.36±3.47	62.59±1.58	311.42±3.38	58.07±1.44	407.34±4.12	58.71±1.93	463.12±4.46	112.00±2.52	
LLMLight	283.02±1.63	59.10±1.27	272.76±1.58	58.25±1.22	271.43±1.51	57.91±1.14	302.29±1.47	21.77±0.96	310.78±1.73	56.78±1.31	LLM-based Methods
Traffic-R1	273.94±1.49	54.13±1.16	261.45±1.34	40.16±0.92	258.06±1.42	49.51±1.07	301.38±1.17	30.94±0.68	315.54±1.52	56.93±1.24	
CoLLMLight	279.40±1.55	63.76±1.21	258.79±1.48	47.22±0.95	264.02±1.52	57.47±1.12	310.45±1.61	18.31±0.88	316.85±1.65	42.93±1.25	
Dec-POMDP	297.45±2.86	75.95±1.51	276.60±2.73	49.45±1.26	282.02±2.82	52.31±1.34	306.27±2.91	20.96±1.13	314.01±3.08	65.68±1.62	Our Methods
Astra (GLM-4.6)	270.33±0.68	56.84±0.53	246.73±0.59	39.21±0.47	254.99±0.64	47.05±0.51	303.78±0.77	13.22±0.44	308.93±0.82	35.12±0.56	
Astra (Qwen3-Max)	270.07±0.64	56.67±0.51	246.48±0.57	39.04±0.43	254.73±0.61	46.88±0.49	303.51±0.73	13.07±0.41	308.65±0.76	34.96±0.52	
Astra (Claude-4.5)	269.58±0.59	56.35±0.48	246.01±0.52	38.73±0.41	254.25±0.57	46.56±0.46	302.99±0.69	12.78±0.39	308.13±0.71	34.65±0.48	
Astra (DeepSeek-R1)	275.42±0.71	60.19±0.62	251.63±0.66	42.41±0.54	259.95±0.69	50.31±0.57	304.15±0.83	16.21±0.51	309.34±0.87	38.29±0.63	
Astra (GPT-5)	271.56±0.63	57.65±0.54	247.92±0.58	39.98±0.44	256.19±0.62	47.84±0.47	304.08±0.74	13.94±0.43	308.23±0.73	35.88±0.53	
Astra (Gemini-3-Flash)	266.17±0.53	54.11±0.45	242.73±0.47	36.59±0.39	250.93±0.52	44.38±0.43	299.40±0.62	10.78±0.37	304.51±0.64	32.53±0.46	

re-evaluate the global state. This bi-directional design achieves a closed-loop balance between the depth of high-level reasoning and the agility of low-level physical evolution.

5 Experiments

In this section, we evaluate *Astra* on real-world datasets to answer the following research questions:

- **RQ1 (Comparative Performance):** Does *Astra* outperform state-of-the-art (SOTA) traditional, MARL, and LLM-based methods in terms of efficiency and congestion relief?
- **RQ2 (Ablation Study):** How do the individual strategic mechanisms within the Macro and Meso layers contribute to the overall coordination efficiency of *Astra*?
- **RQ3 (Sensitivity):** How do key algorithmic hyperparameters—specifically the coordination granularity (K), the topological primitive set (S_{tp}), and the state balancing parameter (α)—influence the overall traffic efficiency and stability of *Astra*?
- **RQ4 (Scalability):** Can *Astra* effectively generalize to large-scale, complex urban road networks while sustaining its performance superiority over state-of-the-art baselines?
- **RQ5 (Interpretability):** How does *Astra* orchestrate its hierarchical components to diagnose and resolve complex coordination challenges in real-world scenarios?

5.1 Experimental Setup

5.1.1 Datasets. To assess robustness across diverse topologies, we utilize five real-world traffic datasets from two major Chinese cities, covering both grid-like and irregular road networks:

- **Jinan (3 datasets):** Collected from the Dongfeng sub-district. The road network comprises 12 intersections with a *regular grid topology*. Road segments are uniform (400m E-W, 800m N-S), representing standard arterial traffic.
- **Hangzhou (2 datasets):** Collected from the Gudang sub-district. This network includes 16 intersections with *complex, irregular topologies* and diverse connection patterns, posing greater challenges for spatial coordination.

5.1.2 Baseline Methods. We compare *Astra* against a comprehensive set of baselines spanning three paradigms:

- **Conventional Heuristics:** *FixedTime* [11] and *MaxPressure* [19] (including its advanced variant), representing rule-based control.
- **SOTA MARL Methods:** We select representative approaches including *MPLight* [3], *CoLight* [21], *PressLight* [20], and their efficient/advanced variants. These methods utilize varying observation shapes and attention mechanisms.
- **LLM-based Methods:** To benchmark against the latest trend, we include *LLMLight* [12], *Traffic-R1* [33] and *CoLLMLight* [28].

In addition, we introduce a concise internal baseline, **Dec-POMDP**, retaining only micro-layer MARL agents without upper-level coordination, to quantify the gain from the hierarchical architecture.

5.1.3 Evaluation Metrics. We adopt two standard metrics: (1) **Average Travel Time (ATT):** The average duration from vehicle entry to exit, reflecting overall network throughput. (2) **Average Waiting Time (AWT):** The average cumulative queuing time per vehicle, indicating the effectiveness of congestion mitigation.

5.1.4 Simulator Environment. Experiments are conducted on the open-source platform: CityFlow [29]. We define an 8-phase action space for each intersection, covering straight and left-turn movements for all directions. Crucially, to simulate asynchronous coordination, we configure differentiated decision cycles: (1) **Micro Layer (10 seconds):** High-frequency execution to capture instantaneous fluctuations; (2) **Macro/Meso Layer (30 seconds):** Low-frequency reasoning to filter noise and capture global trends. A 2-second yellow light buffer is enforced between phase transitions to ensure safety. The simulation duration is set to 3600 seconds.

5.1.5 Implementation Details. *Astra* is implemented using PyTorch and the Ray distributed framework to facilitate scalable and efficient execution. To evaluate the framework’s model-agnostic adaptability, we integrate a diverse suite of frontier LLMs as reasoning backbones, including GLM-4.6, Qwen3-Max, Claude-4.5, DeepSeek-R1, GPT-5, and Gemini-3-Flash. Interaction with these

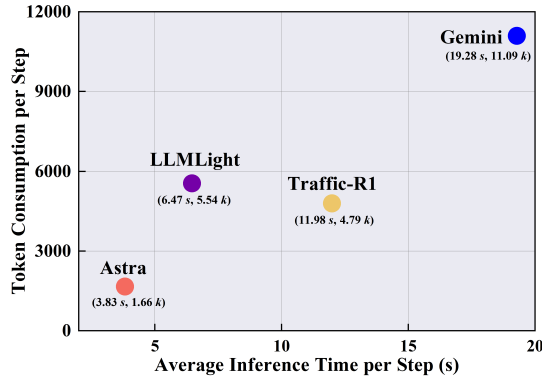


Figure 3: Efficiency analysis of average token consumption and inference latency per decision step.

models is by using their standardized official APIs to ensure consistent and reproducible inference across all evaluations. For a rigorous and equitable comparison, all baselines are optimized using their official open-source implementations to ensure peak performance. All experiments are conducted on a workstation equipped with 128 CPU cores and 8 NVIDIA GPUs. We report the mean performance metrics averaged over five independent runs. Further details regarding the experimental setups are provided in Appendix B.

5.2 Performance Comparison (RQ1)

Comparison on traffic performance. Table 1 presents the performance comparison of Astra against various baselines across five real-world datasets. The results consistently demonstrate that Astra (with Gemini-3-Flash as the backbone LLM) achieves the best performance on both ATT and AWT metrics, validating its effectiveness in network-wide traffic coordination. The key findings from Table 1 are summarized as follows:

- *Superiority across paradigms:* Astra (Gemini-3-Flash) consistently outperforms all traditional, MARL-based, and LLM-based methods. Notably, in the complex Hangzhou-1 dataset, Astra reduces AWT by 65.15% (10.78 vs. 30.94) compared to the strongest LLM baseline (Traffic-R1), highlighting its exceptional capability in alleviating congestion in irregular road networks.
- *Astra vs. SOTA LLM methods:* While LLM-based methods like LLMLight and Traffic-R1 show decent performance, they are often constrained by spatial reasoning limitations. Astra surpasses them by grounding strategic reasoning in topological causal chains. The consistent lead of Astra variants (e.g., Claude-4.5, GPT-5, and Gemini-3-Flash) over standalone LLM controllers proves that our asynchronous hierarchical architecture effectively unleashes the strategic potential of LLMs.
- *Astra vs. SOTA MARL methods:* Compared to leading MARL methods such as Advanced-CoLight and PressLight, Astra demonstrates superior global coordination. MARL agents, despite their rapid reaction, often suffer from coordination myopia. Astra’s macro-level guidance provides the necessary global foresight that prevents local agents from falling into greedy local optima.
- *Effectiveness of hierarchical synergy:* The performance gap between Astra and the internal baseline Dec-POMDP is significant. Dec-POMDP, which relies solely on Micro-layer MARL without upper-level coordination, yields substantially higher ATT and

Table 2: Ablation study evaluating the contribution of key hierarchical components and the asynchronous protocol. Performance is measured by ATT (lower is better), with the best results marked in bold.

Model	Jinan-1	Jinan-2	Jinan-3	Hangzhou-1	Hangzhou-2
<i>w/o Select</i> (Meso)	282.25	263.01	270.43	303.19	309.90
<i>w/o Topo</i> (Meso)	285.93	267.66	273.22	306.14	313.31
<i>w/o System</i> (Macro)	268.84	244.04	251.21	300.59	305.62
<i>Synchronous</i> (Protocol)	270.54	251.26	253.	300.02	307.02
<i>w/o Strat. Lock</i> (Protocol)	292.73	269.94	272.40	305.19	314.23
<i>w/o Pro. Feed</i> (Protocol)	273.86	257.59	262.37	301.14	310.52
Astra (Full)	266.17	242.73	250.93	299.40	304.51

AWT. This contrast quantitatively confirms that the synergistic integration of LLM-driven reasoning and MARL-driven reaction is the primary driver of the system’s overall efficiency.

- *Robustness and generalizability:* Astra maintains stable and superior performance across both regular grid-like topologies (Jinan) and irregular networks (Hangzhou). Even with different backbone LLMs (from GLM to Gemini), the Astra framework consistently achieves state-of-the-art results, showcasing its high compatibility and robustness against varying environmental structures and base model capabilities.

Comparison on computational overheads. Beyond traffic performance, we evaluate the computational viability of Astra against LLMLight, Traffic-R1, and a standalone Gemini-3-Flash intersection-level baseline. As illustrated in Figure 3, Astra occupies the optimal bottom-left quadrant of the efficiency-latency frontier, requiring only 1664 tokens and 3.83 seconds per decision step. While the comparative LLM methods like LLMLight and Traffic-R1 incur significantly higher computational footprints, the standalone Gemini model further reveals the pitfalls of direct intersection-level control, exhibiting prohibitive overhead (~11k tokens) and latency (~20 seconds) that risk rendering strategic guidance obsolete within the 30 seconds decision cycle. By decoupling strategic reasoning from reactive execution, Astra reduces token consumption and latency by approximately 67.68% and 54.42% respectively compared to the two SOTA LLM-based baselines, validating its scalability and readiness for real-world urban deployment.

5.3 Ablation Study (RQ2)

While the superior performance of the full Astra architecture over the decentralized Micro-layer (i.e., Dec-POMDP) has been established in Table 1, this section investigates the internal necessity and incremental value of the specific components within the Macro and Meso layers. To this end, we compare Astra against six variants: (1) *w/o System*, which omits the global observation instructions C^{sys} from the Macro layer; (2) *w/o Select*, which replaces attention-driven region pruning (to identify \mathcal{R}^*) with random region selection; (3) *w/o Topo*, which removes the topological semantic descriptions (C^{topo}) from Meso-level reasoning; (4) *Synchronous*, which forces LLM and MARL agents to operate at an identical, high-frequency cycle (10 seconds), eliminating the asynchronous frequency gap; (5) *w/o Strat. Lock*, which disables strategic locking, allowing local MARL agents to instantaneously override LLM plans via greedy pressure reduction; and (6) *w/o Pro. Feed*, which removes bottom-up proactive alerts, forcing the Macro layer to strictly wait for the 30s cycle even during sudden anomalies.

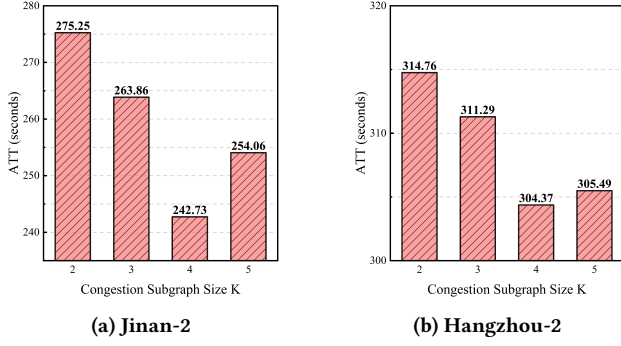


Figure 4: Impact of congestion subgraph size K on traffic performance across distinct road network topologies.

Table 2 summarizes the ATT performance across five datasets. The results demonstrate that removing any module leads to a performance drop, confirming the indispensability of each component. Our main findings are as follows:

- *Topological grounding is the cornerstone of LLM-driven spatial reasoning.* The *w/o Topo.* variant exhibits the most severe performance degradation in all scenarios. For instance, in Jinan-1, ATT surges to 285.93 seconds. This underscores that providing explicit geometric structures (e.g., corridors, loops) is vital for LLMs to map numerical metrics into physical space and formulate effective phase strategies.
- *Precise bottleneck identification is a prerequisite for efficient coordination.* The significant decline in the *w/o Select* variant (e.g., reaching 309.90 seconds in Hangzhou-2) confirms that random selection misallocates LLM reasoning resources to non-critical regions, failing to intervene in deadlocks timely.
- *Macro-level guidance prevents “local view traps”.* The performance gain of Astra over the *w/o System* variant (e.g., a 2.67 seconds improvement in Jinan-1) validates that global situational awareness helps Meso-level agents align regional optimizations with network-wide strategic goals.
- *The Asynchronous Protocol serves as the crucial “glue”.* Any structural deviation from the full asynchronous design strictly degrades overall performance. Despite its high update frequency, the *Synchronous* variant worsens delays (e.g., ATT increases from 266.17 to 270.54 seconds in Jinan-1), proving that forced synchronization disrupts effective orchestration. Furthermore, the severe degradation in *w/o Strat. Lock* (e.g., ATT surging to 292.73 seconds in Jinan-1) demonstrates that strategic locking is imperative to prevent erratic oscillations and greedy local overrides. Finally, the decline in *w/o Pro. Feed* confirms that bottom-up feedback is vital for rapid reaction to sudden anomalies.

5.4 Sensitivity Analysis (RQ3)

Topological primitives serve as the cognitive anchor for the LLM to comprehend complex road networks. Hence, we investigate the impact of the subgraph granularity (size K) and the semantic diversity of the primitive set (S_{tpl}) on system performance.

5.4.1 Impact of Congestion Subgraph Size K . To verify the robustness of our framework across diverse urban morphologies, we evaluate the sensitivity of coordination efficiency to the number of

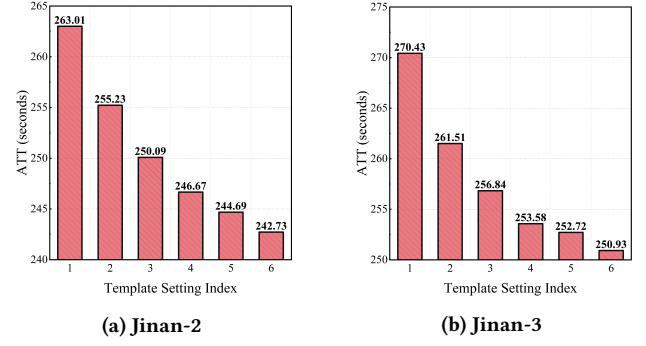


Figure 5: Incremental performance gains across different topological template sets.

intersections K within the critical region \mathcal{R}_i^* by varying K from 2 to 5 across two structurally distinct road networks: the regular grid of Jinan-2 and the more complex, irregular topology of Hangzhou-2.

As illustrated in Figure 4, the ATT result exhibits a consistent U-shaped trend across both scenarios, identifying $K = 4$ as the optimal configuration that generalizes across varying network structures. When $K < 4$, Astra suffers from *under-coordination* due to a constrained spatial scope. For instance, the ATT on Jinan-2 at $K = 2$ is 13.40% higher than at $K = 4$, indicating that sparse collaboration fails to encompass the full propagation of congestion. Conversely, increasing K to 5 results in performance degradation attributed to *topology mismatch*. Forcing oversized templates onto smaller physical bottlenecks introduces extraneous, non-critical nodes into the decision loop, which dilutes the LLM’s reasoning focus and imposes unwarranted constraints on free-flowing intersections. Consequently, we adopt $K = 4$ to strike a stable balance between coordination coverage and topological precision regardless of the underlying network morphology.

5.4.2 Impact of Topological Primitive Semantics. To validate the efficacy of our topological semantic logic, we perform an incremental ablation study on the template set S_{tpl} across six cumulative configurations: (1) Random selection (baseline); (2) $+\tau_L$ (Linear); (3) $+\tau_T$ (T-junction); (4) $+\tau_C$ (Corner); (5) $+\tau_Z$ (Offset); and (6) Full set ($+\tau_O$, Loop). Experiments are conducted on the Jinan datasets, chosen for their heavy traffic loads which necessitate robust structural reasoning to prevent congestion.

Figure 5 demonstrates that Astra achieves strictly monotonic performance gains as template diversity increases. We categorize this improvement into three logical phases:

- **Phase I: Arterial clearance (Linear & T-junction).** Introducing τ_L and τ_T yields the initial performance boost (e.g., ATT drops from 263.01 to 250.09 seconds in Jinan-2). This confirms that basic topological priors enable the LLM to effectively manage routine spillover effects along arterial corridors and merging points.
- **Phase II: Structural adaptation (Corner & Offset).** The inclusion of τ_C and τ_Z further optimizes control in irregular networks. Notably in Jinan-2, which features complex geometries, the ATT drops significantly. This proves that explicit semantic descriptions help the LLM comprehend flow dynamics at non-orthogonal intersections, preventing spatial hallucinations.
- **Phase III: Deadlock resolution (Cyclic Loop).** The final addition of the loop template τ_O triggers a decisive performance leap (e.g.,

Table 3: Sensitivity analysis of the parameter α on Jinan-1.

α	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
ATT	296.32	287.85	280.28	273.74	266.17	273.03	278.37	284.09	290.93
AWT	74.42	67.56	61.83	57.01	54.11	56.46	60.11	65.42	70.96

ATT reaches the optimal 242.73 seconds in Jinan-2). This validates our core hypothesis: cyclic templates are essential for guiding the LLM to identify and break *cyclic deadlocks*, a global coordination challenge that local-view RL agents struggle to resolve.

In summary, a comprehensive set S_{tpl} of topological semantic primitives maximizes the structural alignment between the identified subgraph \mathcal{R}_i^* and physical reality, ensuring that LLM reasoning is grounded in valid causal chains.

5.4.3 Impact of State Balancing Parameter α . The parameter $\alpha \in [0, 1]$ dictates the balance between queue length and vehicle speed in the congestion score (Equation (2)). As shown in Table 3, evaluating α on Jinan-1 reveals that extreme values degrade coordination. A low α (e.g., 0.1) overly prioritizes speed, failing to clear static accumulations (ATT: 296.32 seconds), whereas a high α (e.g., 0.9) overemphasizes queues at the expense of dynamic throughput (ATT: 290.93 seconds). Peak performance is achieved at $\alpha = 0.5$ (ATT: 266.17 seconds, AWT: 54.11 seconds), demonstrating that weighting static accumulation and dynamic flow equally yields the most accurate reflection of physical congestion.

5.5 Scalability to Large-scale Networks (RQ4)

To rigorously evaluate the scalability of Astra, we extend our experiments to two large-scale real-world New York networks (NY1 and NY2), each comprising 196 intersections. Astra is benchmarked against top-tier MARL and LLM baselines (Advanced-CoLight and Traffic-R1), alongside CoLLMLight, a recent cooperative LLM framework. To ensure absolute fairness, both Astra and CoLLMLight are deployed with the identical LLM backbone (Qwen3-30B).

Table 4 demonstrates that Astra sustains its superiority on massive road networks. Notably, the pure MARL baseline (Advanced-CoLight) struggles significantly with city-level coordination, suffering the highest delays (e.g., AWT reaching 261.37 seconds on NY2). While LLM-integrated methods generally perform better, Astra dominates all baselines, including the highly competitive Traffic-R1. Specifically, compared to CoLLMLight under the identical backbone, Astra consistently reduces delays, slashing ATT by 7.65% on NY2 and AWT by 24.65% on NY1. This substantial gap exposes a critical limitation of CoLLMLight: relying exclusively on localized communication (e.g., 1-hop text sharing) fails to orchestrate long-range traffic flows. Conversely, by strategically extracting critical regional subgraphs (\mathcal{R}^*), Astra’s hierarchical architecture equips the LLM with a precise topological context, effectively resolving large-scale coordination bottlenecks.

5.6 Case Study (RQ5)

Beyond the statistical evidence of Astra’s superiority in RQ1–RQ4, we address RQ5 by tracing the framework’s internal logic during a congestion event observed in the Hangzhou-2 dataset (see Appendix C for execution logs). This qualitative analysis illustrates the four-stage bottleneck resolution pipeline:

Table 4: Scalability performance (ATT / AWT) on city-level road networks (196 intersections).

Model	NewYork-1		NewYork-2	
	ATT	AWT	ATT	AWT
Advanced-CoLight	1180.56	141.46	1467.89	261.37
Traffic-R1	932.74	77.24	1194.08	197.04
CoLLMLight (Qwen3-30B)	971.88	93.38	1235.37	198.49
Astra (Qwen3-30B)	923.18	70.36	1140.82	154.84

1) Global situational awareness (Macro/LLM): The monitor agent at Macro layer detects an escalating congestion cluster in the western sector. It issues a strategic directive to pivot from “local queue minimization” to “corridor-level flux optimization”, preventing myopic local agents from exacerbating the arterial blockage.

2) Topological Region Grounding (Meso): Using the topological templates ($K = 4$), the Meso layer identifies a linear corridor (τ_I) spanning intersections 4_4 to 1_4. This region exhibits a critical pressure of 184.0, far exceeding the 45.0 baseline. The system locks this subgraph to provide a structural prior for reasoning.

3) Semantic Causal Reasoning (Meso/LLM): The collaborative agent diagnoses the corridor as a source-sink chain, with intersection 4_4 as the bottleneck (49 vehicles) and 1_4 as the sink. It prescribes a synchronized East-Through-West-Through (ETWT) phase strategy to facilitate rapid, coordinated discharge.

4) Synergetic Execution (Micro & Protocol): The asynchronous synergetic protocol activates strategic locking, overriding local MARL agents to establish a “green wave”. Once the pressure falls below the safety threshold, control is seamlessly returned to the Micro layer for adaptive fine-tuning.

6 Conclusion

In this paper, we present Astra, an asynchronous hierarchical framework for network-wide traffic signal control that decouples strategic LLM reasoning from reactive MARL execution. To align high-level planning with physical road structures, we introduce a topological semantic causal logic mechanism that identifies critical congestion subgraphs via semantic primitives. This design enables the LLM to ground its strategic reasoning in valid spatial constraints while maintaining real-time responsiveness. Furthermore, Astra incorporates an asynchronous synergetic protocol to effectively synchronize these heterogeneous layers for adaptive re-planning. Extensive experiments on five real-world datasets demonstrate that Astra consistently achieves state-of-the-art performance. Notably, it reduces token consumption by 67.68% and inference latency by 54.42% compared to leading LLM-based baselines.

Future work will focus on distilling Astra’s reasoning into lightweight open-source LLMs to facilitate edge deployment directly on local signal controllers.

Acknowledgement

This work was supported in part by National Natural Science Foundations of China under Grant No.62572416 and the Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things under Grant No.2023B1212010007.

References

- [1] Daniel S Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein. 2002. The complexity of decentralized control of Markov decision processes. *Mathematics of operations research* 27, 4 (2002), 819–840.
- [2] Wendelin Böhmer, Vitaly Kurin, and Shimon Whiteson. 2020. Deep coordination graphs. In *International Conference on Machine Learning*. PMLR, 980–991.
- [3] Chacha Chen, Hua Wei, Nan Xu, Guanjie Zheng, Ming Yang, Yuanhao Xiong, Kai Xu, and Zhenhui Li. 2020. Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 3414–3421.
- [4] Luigi P Cordella, Pasquale Foggia, Carlo Sansone, and Mario Vento. 2004. A (sub) graph isomorphism algorithm for matching large graphs. *IEEE transactions on pattern analysis and machine intelligence* 26, 10 (2004), 1367–1372.
- [5] Longchao Da, Minquan Gao, Hao Mei, and Hua Wei. 2024. Prompt to transfer: Sim-to-real transfer for traffic signal control with prompt learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 82–90.
- [6] Jie Feng, Jun Zhang, Junbo Yan, Xin Zhang, Tianjian Ouyang, Tianhui Liu, Yuwei Du, Siqi Guo, and Yong Li. 2025. CityBench: evaluating the Capabilities of Large Language Model as World Model. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5413–5424.
- [7] Sven Gronauer and Klaus Diepold. 2022. Multi-agent deep reinforcement learning: a survey. *Artificial Intelligence Review* 55, 2 (2022), 895–943.
- [8] Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruiqi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. 2024. Large Language Model Based Multi-agents: a Survey of Progress and Challenges. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*. 8048–8057.
- [9] Wook-Shin Han, Jinsoo Lee, and Jeong-Hoon Lee. 2013. Turboiso: towards ultrafast and robust subgraph isomorphism search in large graph databases. In *Proceedings of the 2013 ACM SIGMOD international conference on management of data*. 337–348.
- [10] Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiaowu Zheng, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, et al. 2023. MetaGPT: Meta programming for a multi-agent collaborative framework. In *The twelfth international conference on learning representations*.
- [11] Peter Koonce et al. 2008. *Traffic signal timing manual*. Technical Report. United States. Federal Highway Administration.
- [12] Siqi Lai, Zhao Xu, Weijia Zhang, Hao Liu, and Hui Xiong. 2025. LLMLight: large language models as traffic signal control agents. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2335–2346.
- [13] Kaixiang Lin, Renyu Zhao, Zhe Xu, and Jiayu Zhou. 2018. Efficient large-scale fleet management via multi-agent deep reinforcement learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1774–1783.
- [14] Om Prakash Mahela, Mahdi Khosravy, Neeraj Gupta, Baseem Khan, Hassan Haes Alhelou, Rajendra Mahla, Nilesh Patel, and Pierluigi Siano. 2020. Comprehensive overview of multi-agent systems for controlling smart grids. *CSEE Journal of Power and Energy Systems* 8, 1 (2020), 115–131.
- [15] Afshin Oroojlooy, Mohammadreza Nazari, Davood Hajinezhad, and Jorge Silva. 2020. Attendlight: universal attention-based reinforcement learning model for traffic signal control. In *Advances in Neural Information Processing Systems*, Vol. 33. 4079–4090.
- [16] KJ Prabhachandran, Hemanth Kumar AN, and Shalabh Bhatnagar. 2014. Multi-agent reinforcement learning for traffic signal control. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2529–2534.
- [17] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017).
- [18] Huaze Tang, Hengxi Zhang, Zhenpeng Shi, Xinlei Chen, Wenbo Ding, and Xiaoping Zhang. 2023. Autonomous swarm robot coordination via mean-field control embedding multi-agent reinforcement learning. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 8820–8826.
- [19] Pravin Varaiya. 2013. Max pressure control of a network of signalized intersections. *Transportation Research Part C: Emerging Technologies* 36 (2013), 177–195.
- [20] Hua Wei, Chacha Chen, Guanjie Zheng, Kan Wu, Vikash Gayah, Kai Xu, and Zhenhui Li. 2019. Presslight: learning max pressure control to coordinate traffic signals in arterial network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1290–1298.
- [21] Hua Wei, Nan Xu, Huichu Zhang, Guanjie Zheng, Xinshi Zang, Chacha Chen, Weinan Zhang, Yanmin Zhu, Kai Xu, and Zhenhui Li. 2019. CoLight: learning network-level cooperation for traffic signal control. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1913–1922.
- [22] Hua Wei, Guanjie Zheng, Vikash Gayah, and Zhenhui Li. 2019. A survey on traffic signal control methods. *arXiv preprint arXiv:1904.08117* (2019).
- [23] Hua Wei, Guanjie Zheng, Vikash Gayah, and Zhenhui Li. 2021. Recent advances in reinforcement learning for traffic signal control: a survey of models and evaluation. *ACM SIGKDD Explorations Newsletter* 22, 2 (2021), 12–18.
- [24] Hua Wei, Guanjie Zheng, Huaxiu Yao, and Zhenhui Li. 2018. Intellilight: a reinforcement learning approach for intelligent traffic light control. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2496–2505.
- [25] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, Vol. 35. 24824–24837.
- [26] Libing Wu, Min Wang, Dan Wu, and Jia Wu. 2021. Dynstgat: dynamic spatial-temporal graph attention network for traffic signal control. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2150–2159.
- [27] Zhengxu Yu, Shuxian Liang, Long Wei, Zhongming Jin, Jianqiang Huang, Deng Cai, Xiaofei He, and Xian-Sheng Hua. 2021. MaCAR: urban traffic light control via active multi-agent communication and action rectification. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*. 2491–2497.
- [28] Zirui Yuan, Siqi Lai, and Hao Liu. 2026. CoLLMLight: Cooperative Large Language Model Agents for Network-Wide Traffic Signal Control. In *The Fourteenth International Conference on Learning Representations*.
- [29] Huichu Zhang, Siyuan Feng, Chang Liu, Yaoyao Ding, Yichen Zhu, Zihan Zhou, Weinan Zhang, Yong Yu, Haiming Jin, and Zhenhui Li. 2019. Cityflow: a multi-agent reinforcement learning environment for large scale city traffic scenario. In *The World Wide Web Conference*. 3620–3624.
- [30] Liang Zhang, Qiang Wu, Jun Shen, Linyuan Lü, Bo Du, and Jianqing Wu. 2022. Expression might be enough: Representing pressure and demand for reinforcement learning based traffic signal control. In *International Conference on Machine Learning*. PMLR, 26645–26654.
- [31] Weijia Zhang, Le Zhang, Jindong Han, Hao Liu, Yanjie Fu, Jingbo Zhou, Yu Mei, and Hui Xiong. 2024. Irregular Traffic Time Series Forecasting Based on Asynchronous Spatio-Temporal Graph Convolutional Networks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4302–4313.
- [32] Guanjie Zheng, Xinshi Zang, Nan Xu, Hua Wei, Zhengyao Yu, Vikash Gayah, Kai Xu, and Zhenhui Li. 2019. Diagnosing reinforcement learning for traffic signal control. *arXiv preprint arXiv:1905.04716* (2019).
- [33] Xingchen Zou, Yuhao Yang, Zheng Chen, Xixuan Hao, Yiqi Chen, Chao Huang, and Yuxuan Liang. 2026. Traffic-R1: Reinforced LLMs Bring Human-Like Reasoning to Traffic Signal Control Systems. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*.

A Technical Details of MARL Agent

A.1 Dec-POMDP Formalization

The Micro-layer signal control is formulated as a Decentralized Partially Observable Markov Decision Process (Dec-POMDP) [1], defined by the tuple $\langle \mathcal{I}, \mathcal{S}, \mathcal{O}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$:

- **Agent set \mathcal{I}** : Represents the set of $N = |V|$ individual intersection controllers deployed within the target road network.
- **State \mathcal{S} and observation \mathcal{O}** : \mathcal{S} denotes the global traffic state. Given the partial observability of the environment, each agent i receives a local observation $o_t^i \in \mathcal{O}$ at time t , representing the vehicle distribution on its incoming lanes (see Section 4.3.1).
- **Action space \mathcal{A}** : For each agent i , an action $a_t^i \in \{1, \dots, n\}$ corresponds to the selection of one of n valid signal phases (e.g., North-South Straight) available at its intersection.
- **Transition probability \mathcal{P}** : $\mathcal{P}(s_{t+1}|s_t, \mathbf{a}_t)$ defines the environmental dynamics, where the joint action \mathbf{a}_t of all agents drives the traffic flow evolution from s_t to s_{t+1} .
- **Reward function \mathcal{R}** : We adopt the *cooperative max-pressure reward* (defined in Equation (11)) to maximize network-wide throughput by balancing local queue reduction and neighborhood pressure relief.
- **Discount factor γ** : $\gamma \in [0, 1)$ weighs the relative importance of immediate rewards versus future returns. We set $\gamma = 0.95$ in our implementation to prioritize long-term traffic stability.

A.2 Details of Spatial Attention

To effectively capture spatial dependencies among neighboring intersections, we employ a Multi-head Masked Attention mechanism for feature aggregation. For each intersection i at time t , the raw embedding \mathbf{e}_i^t is linearly projected into Query (Q), Key (K), and Value (V) spaces across H attention heads. The attention weight $\varphi_{i,j,h}^t$, representing the influence of neighbor j on intersection i for head h , is computed as:

$$\varphi_{i,j,h}^t = \text{softmax}_j \left(\frac{(\mathbf{W}_{Q,h}\mathbf{e}_i^t)^\top (\mathbf{W}_{K,h}\mathbf{e}_j^t)}{\sqrt{d_k}} \right), \quad (13)$$

where $\mathbf{W}_{Q,h}$ and $\mathbf{W}_{K,h}$ are head-specific weight matrices and d_k is the scaling factor. The final aggregated feature \mathbf{z}_i^t is generated by concatenating the outputs from all H heads, followed by a linear projection. To ensure training stability and mitigate the vanishing gradient problem, we further incorporate a residual connection and Layer Normalization, as defined in Equation (9).

A.3 PPO Loss Function Components

The total optimization objective \mathcal{L}_{total} is composed of four distinct terms, designed to balance policy stability, value estimation accuracy, and representation robustness:

- **Policy clipping loss (\mathcal{L}_{clip}):** To ensure stable policy updates and prevent destructive large steps, we employ the PPO-Clip objective. Let $\rho_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$ denote the probability ratio. The loss is defined as:

$$\mathcal{L}_{clip} = -\mathbb{E}_t \left[\min \left(\rho_t(\theta) \hat{A}_t, \text{clip}(\rho_t(\theta), 1 - \epsilon, 1 + \epsilon) \hat{A}_t \right) \right], \quad (14)$$

where $\epsilon = 0.2$ is the clipping hyperparameter, and \hat{A}_t represents the advantage term computed via generalized advantage estimation.

- **Value function loss (\mathcal{L}_{value}):** We utilize the mean squared error (MSE) to minimize the discrepancy between the Critic's value estimation $\mathcal{V}(z_i^t)$ and the actual sampled returns \hat{R}_t :

$$\mathcal{L}_{value} = 0.5 \cdot \mathbb{E}_t \left[(\mathcal{V}(z_i^t) - \hat{R}_t)^2 \right]. \quad (15)$$

- **Entropy loss ($\mathcal{L}_{entropy}$):** To encourage exploration and prevent premature convergence to suboptimal policies, we minimize the negative entropy of the policy distribution:

$$\mathcal{L}_{entropy} = -\eta \cdot \mathbb{E}_t \left[\mathbb{H}(\pi(\cdot|z_i^t)) \right], \quad (16)$$

where $\eta = 0.01$ serves as the entropy regularization coefficient.

- **Queue prediction loss (\mathcal{L}_{pred}):** As an auxiliary supervision signal, the system employs a predictor head \mathcal{F}_{pred} to regress the next-step queue length q_{t+1} :

$$\mathcal{L}_{pred} = |\mathcal{F}_{pred}(z_i^t) - q_{t+1}|^2. \quad (17)$$

This term acts as a regularizer, forcing the shared encoder to capture predictive traffic flow dynamics, thereby improving robustness under non-stationary conditions.

B Details of Experimental Setup

B.1 Baseline Methods

We compare Astra against a comprehensive set of baselines spanning rule-based, MARL-based, and LLM-based paradigms:

- **FixedTime** [11]: A traditional rule-based strategy that executes traffic signals according to predefined cycle lengths and phase sequences derived from historical data.
- **MaxPressure** [19]: An adaptive control method that greedily minimizes the pressure difference between upstream and downstream lanes to maximize network throughput.
- **Advanced MaxPressure** [19]: An enhanced variant of MaxPressure that incorporates refined pressure definitions and state features to improve throughput in complex scenarios.
- **MPLight** [3]: A decentralized reinforcement learning method that integrates the FRAP architecture with pressure-based rewards to optimize traffic flow at individual intersections.
- **Advanced MPLight** [3]: An optimized version of MPLight that employs improved state representations and reward shaping to enhance policy convergence and stability.
- **Efficient MPLight** [3]: A streamlined variant of MPLight designed to reduce computational complexity while maintaining competitive control performance.
- **PressLight** [20]: A reinforcement learning approach that combines MaxPressure concepts with Deep Q-learning, explicitly encoding lane-level traffic states to optimize signal phases.
- **CoLight** [21]: A multi-agent reinforcement learning model that employs graph attention networks to facilitate communication and coordinate traffic signals among neighboring intersections.
- **Advanced-CoLight** [21]: An extension of CoLight that further refines the graph attention mechanism to capture long-range spatial dependencies and complex traffic patterns more effectively.
- **LLMLight** [12]: A framework that utilizes Large Language Models (specifically the LightGPT-13B-Llama2 version) as decision-making agents, employing chain-of-thought reasoning to control traffic signals in a zero-shot manner.
- **Traffic-R1** [33]: An advanced LLM-driven method that augments the reasoning depth of language models through reinforcement learning fine-tuning to improve decision quality.
- **Dec-POMDP**: An internal baseline representing the standalone Micro-layer of Astra, where MARL agents operate decentrally without hierarchical guidance from the Macro or Meso layers.

B.2 Simulator Environment

To simulate complex real-world traffic demands and provide finer control granularity, we adopt an 8-phase signal control mechanism, defined as follows: (a) *straight phases*: north-south straight (NTST), east-west straight (ETWT); (b) *left-turn phases*: north-south left-turn (NLSL), east-west left-turn (ELWL); (c) *single-direction mixed phases*: north straight + left-turn, south straight + left-turn, east straight + left-turn, west straight + left-turn.

B.3 Configuration Details

The configurations of Astra are as follows: (1) *Optimization*: We use the Adam optimizer with a unified learning rate of 3×10^{-4} (decay rate 0.98/100 epochs) and a batch size of 4096. (2) *PPO hyperparameters*: The discount factor $\lambda = 0.95$, clip ratio $\epsilon = 0.2$, and update epochs as 4. The loss coefficients are set to $c_1 = 0.5$ (value), $c_2 = 0.01$ (entropy), and $c_3 = 0.5$ (auxiliary prediction). (3) *Network structure*: All neural components utilize a hidden dimension of 128.

Step 1: Global Situational Awareness
<p>>> Aggregating City-Wide Traffic State...</p> <p>Input: Network-wide Traffic State Matrix H_t (containing weighted pressure values for all intersections). Execution Window: $t \in [2820, 2850] / T_{\text{sim}} = 30\text{s}$. System Prompt Generation: Based on the network-wide state, the system generated a guiding System Prompt:</p> <p>[GENERATED MACRO SYSTEM PROMPT]</p> <p>[Macro System Directive] Current City-Wide Status: The network is experiencing a localized but severe "Arterial Paralysis" in the Westbound sector, while other regions remain stable. Strategic Goal: Shift from local optimization to "Corridor-Level Coordination". Priority: Immediate relief of the Westbound arterial blockage to prevent gridlock propagation. Constraint: Ensure downstream sinks are utilized to absorb the discharge.</p>
Step 2: Critical Region Locking
<p>>> Executing Region Selection Algorithm...</p> <p>Candidate Scanning: Region A: Pressure = 45.0 (Stable) Region B: Pressure = 184.0 (CRITICAL) </p> <p>Topological Matching: Identified Structure: Extended Linear Corridor (r_t). Nodes: [intersection_4_4, intersection_3_4, intersection_2_4, intersection_1_4]. Directive: Target Region \mathcal{R}_t^c (Region C) LOCKED.</p>
Step 3: Prompt Construction & LLM Collaboration
<p>>> Constructing Prompt P_{prompt}... >> Injecting Lane-Level Micro-State Data... >> Sending Request to LLM ...</p> <p>[INPUT STREAM TO AGENT]</p> <p>[SYSTEM CONFIGURATION] Role: Traffic Signal Control Expert Objective: Maximize Arterial Throughput & Minimize Corridor Travel Time. Constraint: Strategic Locking Protocol is active. Your decision will override local RL agents for the next T_{lock} seconds. [MACRO-LAYER STRATEGIC DIRECTIVE] >>> ALERT: GLOBAL STATE ANOMALY DETECTED <<< - Pattern Match: "Westbound Arterial Paralysis" (r_t, Corridor Template) - Scope: Group ID [4_4, 3_4, 2_4, 1_4] - Diagnosis: Severe upstream bottleneck at Intersection 4_4 is causing spillover risk. Downstream capacity at 1_4 is underutilized. - Command: INITIATE GREEN WAVE COORDINATION. Shift priority from local queue balancing to arterial discharge. [TOPOLOGICAL SEMANTIC CONTEXT] You are controlling a "Linear Arterial Corridor" consisting of 4 nodes. Flow Direction: East -> West (Westbound).</p> <p>[Node 1: Intersection 4_4] (UPSTREAM SOURCE) - Role: The "Bottle Neck". - Causal Link: If this node fails to discharge, the queue spills back to Road 5_4_2, blocking the entire eastern sector. - Dependency: Needs empty space at Node 3_4 to discharge.</p> <p>[Node 2: Intersection 3_4] (RELAY A) - Role: "Transmission Pipe". - Causal Link: Must synchronize with 4_4. If 3_4 is Red while 4_4 is Green, the "Green Wave" breaks, causing immediate gridlock.</p> <p>[Node 3: Intersection 2_4] (RELAY B) - Role: "Transmission Pipe". - Causal Link: Must synchronize with 3_4 to propagate the flow towards the sink.</p> <p>[Node 4: Intersection 1_4] (DOWNSTREAM SINK) - Role: "The Valve". - Causal Link: This is the only outlet. It MUST be open to absorb the shockwave from the upstream. Current capacity analysis shows it is capable of receiving flow.</p> <p>[REAL-TIME MICRO STATE MATRIX (QDSE ENHANCED)] *Data represents Lane-Level Queue Depth & Saturation Status*</p> <p>Causal Logic: A severe bottleneck at 4_4 is propagating downstream. If the downstream nodes (3_4, 2_4, 1_4) do not clear their queues, the discharge from 4_4 will be blocked (Spillover). You must prioritize the "Westbound Straight" phase across ALL nodes to form a Green Wave.</p> <p>[Real-Time Micro State (Lane-Level Detail)] > Intersection 4_4 Status (SOURCE): - East Approach (road 5_4_2): 80 vehicles total. * Lane 1 (Straight): 49 vehicles [CRITICAL SATURATION] * Lane 0 (Left): 8 vehicles * Lane 2 (Right): 23 vehicles - West Approach (road 3_4_0): 0/44 vehicles total. * Lane 1 (Straight): 26 vehicles > Intersection 3_4 Status (MID A): - East Approach (road 4_4_2): 44 vehicles total. * Lane 1 (Straight): 26 vehicles [High Queue from Upstream] - West Approach (road 2_4_0): 37 vehicles. > Intersection 2_4 Status (MID B): - East Approach (road 3_4_2): 37 vehicles total. * Lane 1 (Straight): 22 vehicles [Moderate Queue] > Intersection 1_4 Status (SINK): - East Approach (road 2_4_2): 23 vehicles total. * Lane 1 (Straight): 13 vehicles [Flow is relatively smooth] - West Approach (road 0_4_0): 40 vehicles (Counter-flow). * Lane 1 (Straight): 25 vehicles. </p> <p>[Instruction] 1. Analyze the lane-level saturation (specifically Lane 1 on the arterial). 2. Verify if the sink (1_4) has capacity to receive the shockwave. 3. Select the best phase ID for all 4 intersections (Phase 1 = East-West Straight). Output JSON.</p>
Step 4: Strategic Decision Reasoning & Parsing
<p>>> Receiving Response from LLM Agent... >> Parsing JSON Payload...</p> <p>[OUTPUT STREAM FROM AGENT]</p> <p>JSON { "reasoning": "Lane-level analysis confirms a severe blockage at the source (4_4) with 49 vehicles stuck in the straight lane. The macro directive requires 'Corridor-Level Coordination' to prevent gridlock. The sink node (1_4) is relatively open (only 13 vehicles in the receiving lane), making it a viable target for discharge. 'stratStrategy': Execute a synchronized 'Green Wave' by activating Phase 1 (East-West Straight) for the entire chain [4_4 -> 1_4]. This aligns with the strategic goal of clearing the arterial blockage.", "actions": { "intersection_4_4": "ETWT", "intersection_3_4": "ETWT", "intersection_2_4": "ETWT", "intersection_1_4": "ETWT" } }</p>
Step 5: Action Execution
<p>>> Parsing Actions:</p> <p>intersection_4_4: "ETWT" -> Phase 1 intersection_3_4: "ETWT" -> Phase 1 intersection_2_4: "ETWT" -> Phase 1 intersection_1_4: "ETWT" -> Phase 1</p> <p>>> System State Update:</p> <p>[LOCK ENGAGED] Extended Group (4_4 to 1_4) entering Strategic Locking Mode. Control Authority: \mathcal{R}_t^c (Group): LLM Instruction (Phase 1) overriding RL Policy. Other Nodes: Independent RL Agent (Continuing local optimization)</p> <p>>> Simulation Step Executing...</p> <p>Effect: Green wave established. High-pressure queue at road 5_4_2 begins discharging downstream.</p>

Figure 6: Detailed decision-making pipeline. (Top) Part 1; (Bottom) Part 2.

We set the weighting coefficient $\alpha = 0.5$ in Equation (2) to assign equal importance to the normalized queue length and the traffic speed ratio in the formulation of global traffic state h_t^i , ensuring a balanced perception of congestion accumulation and movement efficiency. Additionally, the speed penalty factor is set to $\beta = 0.45$ in Equation (8) for the fine-grained lane pressure encoding x_j^l . This configuration balances the contribution of the positional weight term ω_j and the kinematic stagnation penalty, enabling the model to effectively distinguish between high-density slow traffic and free-flow conditions.

C Case Study Details

This section provides exhaustive supplementary evidence for the Case Study presented in Section 5.6. To offer a transparent view of Astra's internal coordination, we detail the complete hierarchical decision-making pipeline (see Figures 6).

Specifically, we provide the *system prompts*, *input semantic mappings*, and *intermediate reasoning trajectories* that illustrate the precise transition from macro-level situational awareness to meso-level topological grounding. These materials, along with the raw execution logs omitted from the main text for brevity, serve to validate the interpretability and logic-driven nature of the proposed framework.