Safety-Critical Decision Making and Coordination for Autonomous Vehicles in Mixed Traffic

Zhihao Lin

Supervisor: Dr. Jianglin Lan

SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

SCHOOL OF ENGINEERING

COLLEGE OF SCIENCE & ENGINEERING

11/11/2025

To my supervisor and life-mentor,

Dr. Jianglin Lan,

my parents,

Mr. Cunhai Lin and Ms. Haimei Zhao,

my pal,

Mr. Zhen Tian,

my academic-guiders,
Dr. Christos Anagnostopoulos and Prof. David Flynn, *whose guidance, support, and encouragement made this work possible.*

Abstract

Coordinating multiple autonomous vehicles at unsignalized intersections remains a fundamental challenge in multi-agent systems. The exponential growth of joint action spaces, coordination ambiguity under symmetric configurations, and stringent real-time constraints render centralized approaches intractable. This thesis introduces a framework that reformulates Level-k cognitive hierarchy for safety-critical coordination, integrating with Monte Carlo Tree Search (MCTS) to achieve scalable planning with emergent safety properties.

The central contribution redefines Level-0 as a universal safety initialization generating conservative baseline trajectories, rather than modeling naive random behaviors. This transforms Level-k reasoning from a descriptive cognitive model into a constructive planning framework where safety emerges structurally through cascading conservative margins: Level-1 agents inherit Level-0 safety anchors while optimizing efficiency, and Level-2 agents amplify these margins by anticipating Level-1 strategic responses. The framework decomposes multi-agent coordination via dual-filtered interaction graphs combining spatial conflict detection with strategic reasoning, reducing computational complexity from exponential to linear in agent count.

The MCTS integration enables efficient exploration of the action space through selective sampling guided by Upper Confidence Bounds. Safety-aware pruning eliminates approximately 70% of unsafe actions during tree expansion, reducing the effective branching factor from 15 to approximately 4–5 actions per node. Trajectory caching exploits the deterministic nature of Level-k rollouts to achieve 35% cache hit rates, avoiding redundant computation. For mixed traffic scenarios involving human-driven vehicles, the framework incorporates style-aware behavior prediction based on the Intelligent Driver Model, time-varying uncertainty quantification, and adaptive safety thresholds that respond to interaction-specific risks. The complete framework reduces computational complexity over 20 orders of magnitude compared to joint optimization, enabling sub-100-millisecond planning cycles suitable for real-time deployment.

Extensive experimental validation across challenging scenarios demonstrates the framework's effectiveness. In symmetric eight-agent intersection coordination where baseline methods exhibit 15–35% collision rates, the proposed approach achieves zero collisions with 95–98% arrival rates. Mixed traffic experiments at 50% autonomous vehicle penetration maintain collision rates below 2% despite diverse human driving behaviors, with consistent performance across penetration rates from 20% to 100%. The framework's interpretability through explicit reasoning traces, modularity enabling component-wise validation, and fully decentralized architecture requiring no inter-vehicle communication provide practical advantages for real-world deployment. Beyond autonomous driving, the theoretical contributions—reconstructed Level-k reasoning with emergent safety and dual-filtered interaction decomposition—offer broader insights applicable to multi-agent coordination challenges across robotics, game theory, and artificial intelligence.

Acknowledgements

I wish to extend my heartfelt gratitude to everyone who supported me throughout the completion of this dissertation. Above all, I am thankful to God for bestowing upon me the strength, perseverance, and grace necessary to reach this milestone.

My profound gratitude goes to Dr. Jianglin Lan, my supervisor and mentor in both academia and life. Dr. Lan's steadfast support and insightful guidance have been pivotal throughout my doctoral studies. He consistently made himself available for in-person consultations, helping me navigate complex academic obstacles with clarity and purpose. His willingness to share research materials, scholarly articles, writing techniques, and personal wisdom has profoundly shaped my development as a researcher. Recognizing my initial difficulties, Dr. Lan invested extra time in nurturing my academic capabilities, patiently guiding me through each challenge and helping me build a solid foundation of knowledge. This dissertation would not have been possible without his mentorship.

I owe immense gratitude to my parents, Mr. Cunhai Lin and Ms. Haimei Zhao. Their steadfast belief in me, and their boundless love and patience, sustained me through the demanding years of doctoral study. They have been my anchor and inspiration.

Special thanks to my dear friends Dr. Zhen Tian, Mr. Qi Zhang, and Lin Wu, whose companionship enriched my academic pursuits and daily life. Their insights from personal experience accelerated my learning and helped me adjust to doctoral research. Zhen, Qi, and Lin—your friendship has been one of the greatest gifts of this journey.

I am deeply appreciative of Dr. Christos Anagnostopoulos, my secondary supervisor, whose meticulous feedback and thoughtful discussions helped sharpen my research skills. His constructive critiques highlighted critical areas for development, while his encouragement during challenging periods reinforced my determination and self-belief.

My sincere appreciation extends to Dr. Chongfeng Wei, Prof. Dezong Zhao and Prof. David Flynn for their expert advice and thoughtful perspectives. I am equally grateful to my research group members, particularly Mr. Peizhuo Yu and my fellow colleagues, whose camaraderie and collaborative spirit lightened the burden of doctoral work. Your warm encouragement, ongoing support, and stimulating exchanges made this challenging journey far more manageable.

Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution. This thesis has been written and compiled by the author, Zhihao Lin, and certifies that the thesis presented here for examination for the PhD degree at the University of Glasgow.

Zhihao Lin

Contents

List of Publications

- **Zhihao Lin**, Lin Wu, Zhen Tian, Alessio Lomuscio, and Jianglin Lan. "Scalable and Safe Multi-Agent Coordination with Reconstructed Level-k Monte Carlo Tree Search" *The 25th International Conference on Autonomous Agents and Multiagent Systems* (AAMAS 2026).(Accepted)(doi: https://eprints.gla.ac.uk/376419/)

- **Zhihao Lin**, Jianglin Lan, Anh-Tu Nguyen, David Flynn. "Contingency-Aware Spatiotemporal Optimization for Safe Autonomous Vehicle Trajectory Planning" *IEEE Transactions on Intelligent Transportation Systems* (2025). (doi: https://doi.org/10.1109/TITS.2025.3597234)

- **Zhihao Lin**, Jianglin Lan, Christos Anagnostopoulos, Zhen Tian, David Flynn. "Safety-Critical Multi-Agent MCTS for Mixed Traffic Coordination at Unsignalized Intersections" *IEEE Transactions on Intelligent Transportation Systems* (2025). (doi: https://doi.org/10.1109/TITS.2025.3598727)

- **Zhihao Lin**, Jianglin Lan, Xianxian Zhao. "KAN-LSTM Enhanced Multi-Agent Advantage Actor-Critic Reinforcement Learning for Autonomous Ramp Merging" *IEEE Transactions on Vehicular Technology* (2025). (doi: https://doi.org/10.1109/TVT.2025.3593661)

- **Zhihao Lin**, Zhen Tian, Jianglin Lan, Qi Zhang, Ziyang Ye, and Hanyang Zhuang. "A Conflicts-Free, Speed-Lossless KAN-Based Reinforcement Learning Decision System for Interactive Driving in Roundabouts" *IEEE Transactions on Intelligent Transportation Systems* (2025). (doi: https://doi.org/10.1109/TITS.2025.3578279)

- **Zhihao Lin**, Zhen Tian, Jianglin Lan, Dezong Zhao, and Chongfeng Wei. "Uncertainty-Aware Roundabout Navigation: A Switched Decision Framework Integrating Stackelberg Games and Dynamic Potential Fields" *IEEE Transactions on Vehicular Technology* (2025). (doi: https://doi.org/10.1109/TVT.2025.3638264)

- **Zhihao Lin**, Qi Zhang, Zhen Tian, Peizhuo Yu, Ziyang Ye, Hanyang Zhuang. and Jianglin Lan. "SLAM2: Simultaneous Localization and Multimode Mapping for indoor dynamic environments." *Pattern Recognition* (2025). (doi: https://doi.org/10.1016/j.patcog.2024.111054)

- **Zhihao Lin**, Zhen Tian, Qi Zhang, Hanyang Zhuang, and Jianglin Lan. "Enhanced Visual SLAM for Collision-Free Driving with Lightweight Autonomous Cars." *Sensors*. (doi: https://doi.org/10.3390/s24196258)

- **Zhihao Lin**, Qi Zhang, Zhen Tian, Peizhuo Yu and Jianglin Lan, "DPL-SLAM: Enhancing Dynamic Point-Line SLAM Through Dense Semantic Methods,"*IEEE Sensors Journal*. (doi: https://doi.org/10.1109/JSEN.2024.3373892)

- Zhen Tian, **Zhihao Lin**, Dezong Zhao, Wenjing Zhao, David Flynn, and Chongfeng Wei. "Evaluating Scenario-based Decision-making for Interactive Autonomous Driving Using Criteria Matrix: A Survey." *IEEE Transactions on Intelligent Transportation System* (2025). (doi: https://doi.org/10.1109/TITS.2025.3636070)

- Zhen Tian, Dezong Zhao, **Zhihao Lin**, David Flynn, Wenjing Zhao, and Daxin Tian. "Balanced Reward-Inspired Reinforcement Learning for Autonomous Vehicle Racing." *In 6th Annual Learning for Dynamics and Control Conference*, pp. 628-640. PMLR, 2024. (doi: https://proceedings.mlr.press/v242/tian24a/tian24a.pdf)

- Zhen Tian, Dezong Zhao, **Zhihao Lin**, Wenjing Zhao, David Flynn, Yuande Jiang, Daxin Tian, Yuanjian Zhang, and Yao Sun. "Efficient and Balanced Exploration-driven Decision Making for Autonomous Racing Using Local Information." *IEEE Transactions on Intelligent Vehicles* (2025). (doi: https://doi.org/10.1109/TIV.2024.3432713)

- Zhen Tian, Dezong Zhao, **Zhihao Lin**, Wenjing Zhao, David Flynn, Daxin Tian, and Yao Sun. "Balanced Exploration and Attention-Inspired Decision Making for Autonomous Driving." *IEEE Transactions on Vehicular Technology* (2025). (doi: https://doi.org/10.1109/TVT.2025.3590637)

- **Zhihao Lin**, Xianxian Zhao, and Jianglin Lan. "Scalable multi-vehicle decision making for autonomous driving at unsignalized intersections" *IEEE Transactions on Intelligent Transportation Systems* (2025).(Under Review)

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

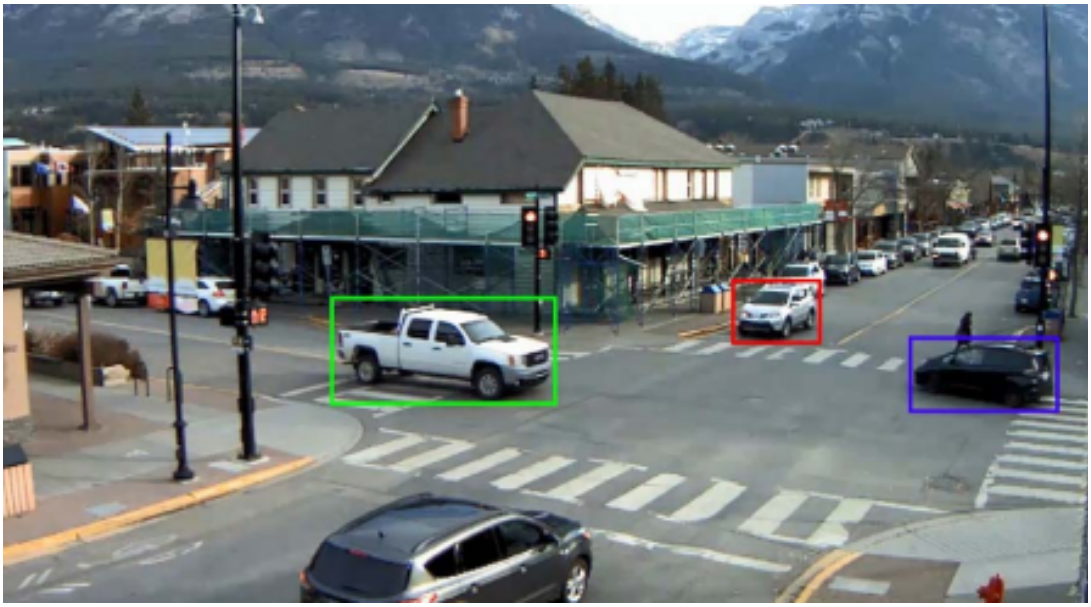


Figure 1.1: Autonomous vehicles navigating complex urban intersections.

The advancement of autonomous vehicles (AVs) represents one of the most transformative technological developments of the 21st century, promising to revolutionize transportation safety, efficiency, and accessibility [1]. As autonomous driving technology matures from controlled environments to complex urban settings, one of the most critical challenges emerges at unsignalized intersections, where multiple vehicles must coordinate their movements without centralized traffic control [2]. This challenge intensifies dramatically in mixed traffic scenarios, where AVs must safely and efficiently interact with human-driven

vehicles (HDVs) exhibiting uncertain and diverse driving behaviors. The fundamental question this thesis addresses is: how can AVs make scalable, safe, and strategic decisions in multi-agent traffic scenarios ranging from symmetric coordination problems to heterogeneous mixed traffic environments?

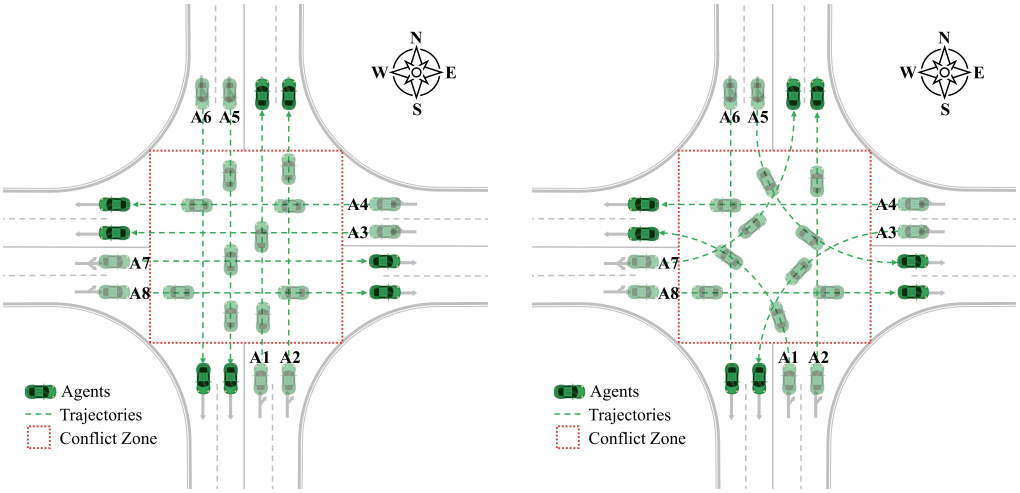


Figure 1.2: Symmetric intersection scenarios between agents.

Unsignalized intersections exemplify the broader multi-agent coordination challenge in autonomous driving as shown in Fig. 1.1. Unlike signalized intersections where traffic lights provide explicit coordination, unsignalized intersections require vehicles to negotiate passage through implicit communication, strategic reasoning, and real-time adaptation. This coordination becomes exponentially more complex as the number of interacting agents increases. Consider a symmetric scenario where eight vehicles simultaneously approach an intersection from symmetric positions, as shown in Fig. 1.2, all equidistant from the conflict zone. This configuration creates a coordination deadlock: no vehicle possesses inherent spatial or temporal advantage, leading to potential deadlock without principled coordination mechanisms. Such scenarios expose the fundamental limitations of conventional planning paradigms and demand novel approaches that can balance computational tractability with strategic sophistication.

Intersection coordination becomes substantially more demanding in mixed traffic environments, where AVs must coexist and cooperate with human drivers [3]. Human driving behavior introduces layers of uncertainty, heterogeneity, and bounded rationality that complicate decision-making. Human drivers exhibit diverse intentions and risk prefer-

ences that violate the perfect rationality assumptions underlying many game-theoretic approaches, rendering deterministic models inadequate [4]. An AV navigating through such environments must simultaneously reason about multiple objectives including collision avoidance, traffic efficiency, passenger comfort, and cooperative behavior, all while operating under real-time computational constraints.

Traditional approaches [5, 6] to multi-agent coordination and intersection management fall into several categories, each with inherent limitations [7]. Rule-based methods, while computationally efficient and interpretable, rely on predefined priority assignments or conflict resolution sequences that struggle to adapt to dynamic and symmetric scenarios. When multiple vehicles arrive simultaneously with equal priority, these methods either impose arbitrary tie-breaking rules or resort to overly conservative behaviors that sacrifice efficiency. Optimization-based approaches, though theoretically elegant, face the curse of dimensionality as the joint action space grows exponentially with the number of agents. Even for modest planning horizons, finding optimal joint trajectories for eight vehicles becomes computationally intractable, requiring simplifications that compromise solution quality or safety guarantees.

Learning-based methods [8–17], particularly deep reinforcement learning, have demonstrated impressive capabilities in modeling complex multi-agent interactions. These approaches can learn sophisticated coordination strategies from large-scale simulation or real-world data, capturing patterns that may be difficult to encode explicitly. However, they exhibit limitations that are particularly concerning in safety-critical applications. First, learned policies often lack interpretability, making it difficult to verify their behavior in novel situations or provide safety guarantees. Second, they require extensive training data covering diverse scenarios, yet may still fail in out-of-distribution situations. Third, their performance can degrade when the structure of the environment changes, such as varying numbers of agents or different intersection geometries, requiring retraining rather than adaptive reasoning. These limitations make purely learning-based approaches insufficient for safety-critical autonomous driving applications where collision avoidance must be guaranteed, not merely probable.

Game-theoretic methods [18–20] offer a principled framework for modeling strategic interactions among rational agents. Classical approaches such as Nash equilibrium and Stackelberg games can capture competitive and cooperative dynamics between vehicles. Nevertheless, they face two critical challenges [21, 22]. First, they assume perfect rationality and complete information, which rarely hold in real-world driving scenarios where human drivers exhibit bounded rationality and private information about their intentions. Second, when multiple Nash equilibria exist, agents may converge to different equilibria, leading to coordination failures. Moreover, computing equilibria in multi-agent settings becomes computationally prohibitive as the number of agents increases, limiting their applicability to small-scale problems.

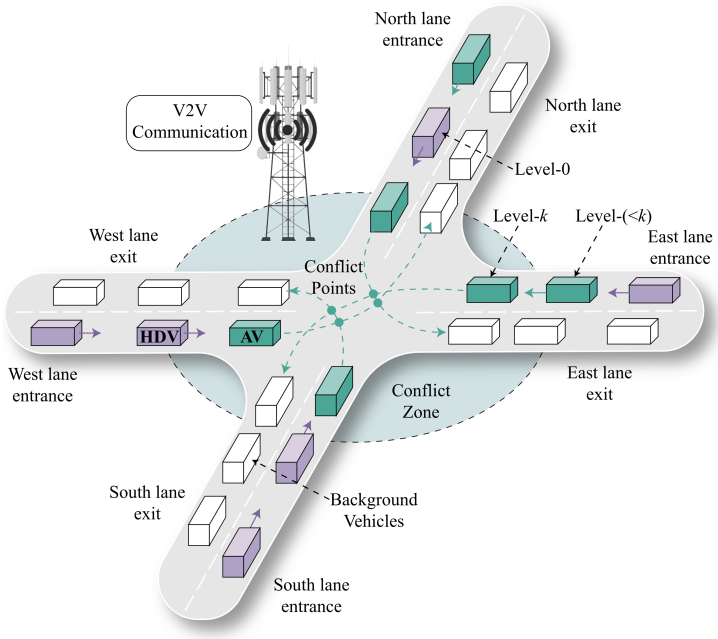


Figure 1.3: An unsignalized intersection with AVs and HDVs.

More recently, cognitive hierarchy models, particularly Level-k reasoning, have emerged as a behaviorally grounded approach to modeling strategic thinking in multi-agent systems as shown in Fig. 1.3. Level-k reasoning posits a hierarchy of reasoning depths, where Level-0 agents employ naive or non-strategic behaviors, and Level-k agents ($k \geq 1$) reason strategically by modeling others as Level-$(k-1)$ agents. This framework offers cognitive plausibility and has been successfully applied to various game-theoretic domains. However, traditional Level-k formulations suffer from a critical flaw in safety-critical applications:

they define Level-0 as random or myopic actors, leading to unrealistic and potentially unsafe baseline behaviors. When higher-level reasoning builds upon fundamentally unsafe foundations, the entire hierarchy lacks safety guarantees, limiting its direct applicability to autonomous driving.

Monte Carlo Tree Search (MCTS) [23] has emerged as a powerful online planning method that balances exploration and exploitation through simulation-based lookahead. Its success in complex domains such as Go and real-time strategy games demonstrates its potential for handling large state-action spaces [23, 24]. In autonomous driving contexts [25], MCTS offers advantages in handling uncertainty through stochastic simulation and can incorporate domain knowledge through reward shaping and pruning strategies. Despite these advantages, MCTS faces severe scalability challenges in multi-agent settings due to the combinatorial explosion of joint action spaces. As the number of agents increases, the branching factor of the search tree grows exponentially, making exhaustive exploration infeasible within real-time computational budgets.

The tension underlying all these approaches can be articulated as follows: strategic multi-agent coordination requires reasoning about joint behaviors and mutual influences, yet the computational complexity of such reasoning grows exponentially with the number of agents. Existing methods either sacrifice strategic sophistication for computational tractability through decomposition and simplification, or achieve strategic reasoning at the cost of computational intractability in realistic multi-agent scenarios [26, 27]. This thesis argues that resolving this tension requires rethinking how cognitive hierarchy, strategic reasoning, and search-based planning can be integrated in multi-agent systems. Rather than treating these paradigms as competing alternatives, we propose a unified framework that leverages their complementary strengths while mitigating their individual weaknesses.

Existing approaches to multi-agent coordination in autonomous driving exhibit significant limitations across multiple dimensions. Table 1.1 summarizes the key limitations of representative methods and how our framework addresses these challenges. Rule-based methods lack strategic reasoning capabilities and cannot adapt to uncertainty. Learning-

Table 1.1: Objective comparison of existing approaches and the proposed framework

Method	Strengths	Limitations
Rule-Based [28–31]	Interpretable; computationally lightweight; easy to deploy	Cannot handle strategic interactions; poor scalability with agent number; lacks adaptability to uncertainty
Learning-Based [15, 32]	Strong generalization from data; handles complex scenarios; end-to-end trainable	Requires massive training data; limited interpretability; weak safety guarantees; poor generalization to unseen scenarios
Game-Theoretic (Stackelberg [33], Nash [19])	Principled strategic reasoning; equilibrium-based optimality guarantees; no training data required	Standard formulations assume fully rational agents with deterministic best responses; incorporating probabilistic behavioural uncertainty (e.g., stochastic human driving) requires extensions such as Bayesian games, which significantly increase computational complexity
Vanilla MCTS [34]	Online planning without training; handles stochastic environments; flexible reward shaping	Joint action space grows as $\mathcal{O}(\lvert\mathcal{A}\rvert^N)$ per step; no strategic opponent modelling; treats other agents as environment noise; poor scalability in multi-agent settings
Ours	Level-k decomposition reduces multi-agent scaling to $\mathcal{O}(N \cdot K \cdot b_{\text{eff}} \cdot H)$[1]; strategic reasoning with integrated safety guarantees; probabilistic human modelling for mixed traffic; no training data required	Relies on reference-path assumptions; human behaviour modelled using simplified IDM-based dynamics; computational cost scales with reasoning depth k; current evaluation limited to intersection scenarios

based approaches require extensive training data and provide limited safety guarantees. Game-theoretic methods, while offering strategic reasoning, make idealized rationality assumptions that fail to capture human behavioral uncertainty. Standard MCTS approaches face exponential complexity in the joint action space and lack explicit opponent modeling.

1.2 Proposed Approach and Technical Evolution

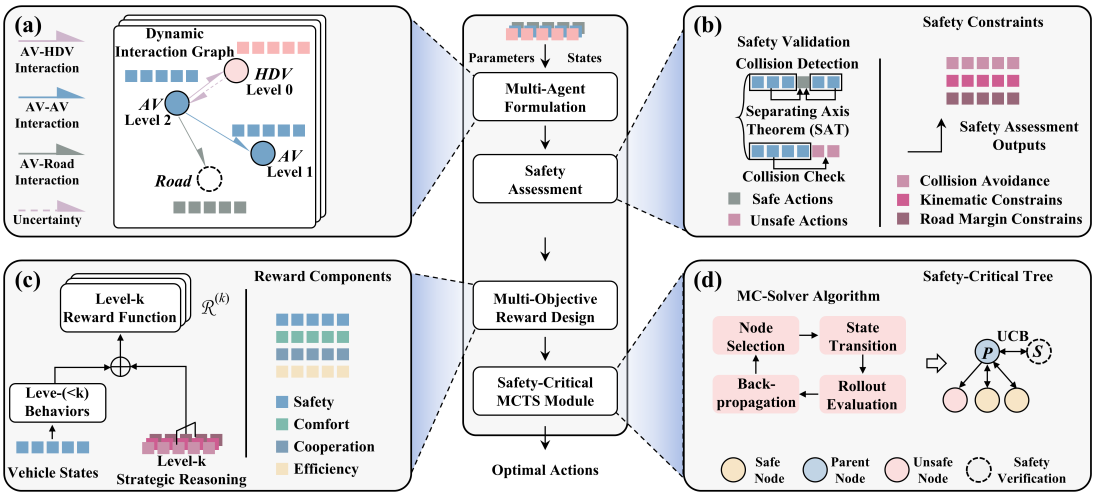


Figure 1.4: System architecture of the proposed MCTS-Level-k framework for scalable multi-vehicle decision making.

This thesis proposes a scalable multi-agent decision-making framework that integrates Level-k reasoning and MCTS for safety-critical autonomous driving applications as shown in Fig. 1.4. Our key insight is that Level-k reasoning and MCTS are not merely complementary but synergistic: Level-k reasoning provides hierarchical structure that decomposes the exponential joint action space into sequential reasoning steps, while MCTS provides adaptive sampling and uncertainty handling that enhances the realism and robustness of strategic reasoning. This integration is anchored by a critical redefinition of Level-0 not as a naive agent type, but as a universal safety initialization procedure that generates conservative baseline trajectories with guaranteed collision-free properties.

The theoretical foundation of our framework lies in reformulating the Level-k hierarchy so that safety is structurally embedded rather than imposed as an external constraint. Traditional Level-k formulations define Level-0 agents as non-strategic actors exhibiting random or myopic behaviors; while some conservative variants exist, they are not designed to provide explicit collision avoidance margins that propagate through the reasoning hierarchy. We redefine Level-0 as a safety-aware trajectory generation procedure that produces conservative baseline behaviors with explicit worst-case separation margins computed via the Separating Axis Theorem (Section 3.2.3). The key consequence for higher levels is that Level-k agents ($k \geq 1$) optimize against Level-$(k-1)$ trajectories that are collision-free

under the assumed dynamics model, rather than against arbitrary or random baselines. This propagates minimum safety margins upward through the hierarchy: a Level-1 agent responding to a safety-aware Level-0 baseline inherits the spatial separation guarantee embedded in that baseline, and so on for higher levels.

This baseline is not a description of how agents actually behave, but rather a procedural approximation of minimum safety margins under the assumed constant-velocity prediction model. We acknowledge that in practice, HDV behavioral uncertainty, sensor noise, and model mismatch mean that safety can only be ensured probabilistically rather than absolutely. The framework therefore provides structural safety properties under its modeling assumptions, with robustness to deviations handled through the uncertainty-adjusted thresholds in Section 5.4.

Based on this reformulated Level-k foundation, we integrate MCTS to handle the exploration-exploitation tradeoff in stochastic multi-agent environments. The integration operates through a carefully designed decomposition: rather than searching over the exponential joint action space of all agents simultaneously, our framework leverages the Level-k hierarchy to sequence the search process. Each agent conducts MCTS over its own action space while modeling other agents' responses through Level-$(k-1)$ reasoning. This decomposition reduces computational complexity from exponential in the number of agents to linear, making real-time planning feasible even in scenarios with eight or more vehicles. The MCTS backbone provides three critical capabilities: adaptive sampling that focuses computational resources on promising action sequences, stochastic rollouts that capture uncertainty in human driving behaviors, and progressive refinement through iterative tree expansion and backpropagation.

To ensure real-time constraints without sacrificing decision quality, we introduce three algorithmic techniques. First, we develop a safety-aware pruning mechanism that eliminates unsafe action sequences early in the search process, substantially reducing the effective branching factor. This pruning is enabled by the conservative Level-0 baseline, which provides tight lower bounds on minimum safe spacing between vehicles. Second, we design a Dual-Filtered Interaction Graph that dynamically identifies relevant interaction partners for each agent based on both geometric proximity and reasoning-level hierarchy.

This filtering addresses the scalability challenge in dense multi-agent scenarios by allowing each vehicle to focus computational resources on agents that significantly influence its decision, rather than exhaustively considering all pairwise interactions. Third, we implement trajectory caching that exploits the temporal coherence of multi-agent interactions, thereby avoiding redundant computation across planning cycles.

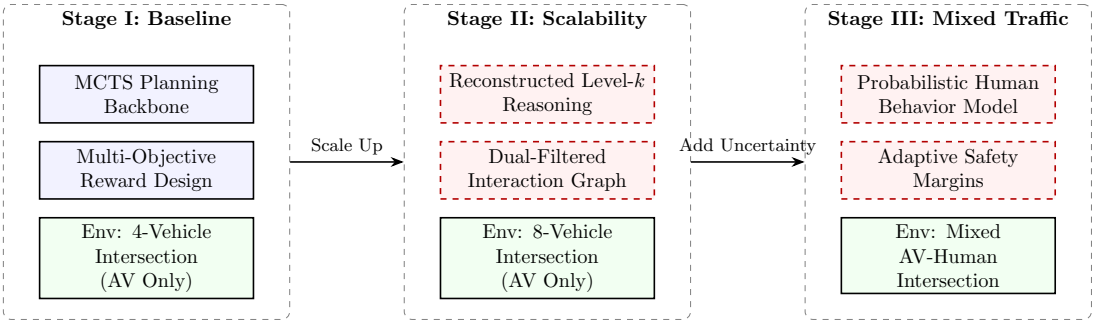


Figure 1.5: Technical evolution of the proposed multi-agent decision-making framework. The development progresses through three stages: (I) establishing the baseline MCTS planning for homogeneous coordination; (II) scaling to eight-agent scenarios via reconstructed Level-k reasoning and dual-filtered interaction graphs; and (III) extending to mixed traffic with probabilistic human modeling (Gaussian uncertainty over predicted HDV trajectories with online Bayesian style estimation)
and adaptive safety margins.

The development of our framework follows a systematic progression from homogeneous multi-agent scenarios to heterogeneous mixed traffic environments, with each stage introducing new complexities and corresponding algorithmic refinements as shown in Fig. 1.5. In the initial stage, we focus on symmetric multi-agent coordination problems involving only AVs. The canonical eight-vehicle intersection scenario serves as a stress test for our framework, as the maximal symmetry creates inherent ambiguity that challenges any coordination mechanism. Here, the reconstructed Level-k reasoning provides the differentiation needed to break symmetry through hierarchical strategic thinking, while MCTS explores alternative coordination strategies without arbitrary tie-breaking rules. Our framework achieves near-perfect success rates in these scenarios, demonstrating that principled strategic reasoning can resolve deadlock situations that confound rule-based or reactive approaches.

The second stage extends our framework to mixed traffic environments where AVs must interact with HDVs exhibiting uncertain and heterogeneous behaviors. This extension requires three major enhancements to the baseline framework. First, we develop a probabilistic human behavior model that captures diverse driving intentions, risk preferences, and decision-making styles beyond deterministic models such as the Intelligent Driver Model. This model combines physics-based motion primitives with learned distributions over driver preferences, suitable for realistic simulation of human responses during MCTS rollouts. Second, we introduce adaptive safety margins that adjust dynamically based on estimated human driver uncertainty, ensuring robust collision avoidance even when human behaviors deviate from expected patterns. Third, we design multi-objective reward functions that balance safety, efficiency, comfort, and cooperative behavior, reflecting the complex tradeoffs AVs must navigate in mixed traffic scenarios.

The third stage refines the Level-k and MCTS integration to handle the increased computational burden of mixed traffic reasoning. In homogeneous scenarios, all agents employ similar decision-making processes, simplifying the modeling of mutual responses. In mixed traffic, AVs must simultaneously reason about other AVs using game-theoretic models and human drivers using probabilistic behavioral models. We address this heterogeneity through a dual-filtered interaction graph that combines spatial relevance with reasoning-level hierarchy. Spatially, we identify interaction partners based on predicted trajectory overlap and conflict probability. Hierarchically, we assign reasoning levels based on agent types: AVs are assigned higher reasoning levels reflecting their strategic capabilities, while human drivers are modeled at lower levels reflecting bounded rationality and behavioral uncertainty. This dual filtering enables efficient yet accurate modeling of complex mixed traffic interactions.

Across all three stages, a consistent design principle applies: safety emerges as a structural property rather than being imposed as an external constraint. The reconstructed Level-0 establishes conservative baselines, MCTS pruning eliminates unsafe branches early, adaptive safety margins account for uncertainty, and the entire hierarchy preserves collision-free properties through bounded rationality. This design ensures that strategic sophistication and computational efficiency do not compromise safety guarantees, addressing a limitation of existing multi-agent planning methods.

The computational advantages of our framework are substantial. Compared to joint optimization approaches that scale exponentially with the number of agents, our sequential decomposition through Level-k reasoning achieves linear scaling. Empirical evaluations demonstrate computational speedups exceeding 60% compared to standard MCTS baselines in eight-agent scenarios, while maintaining or improving solution quality measured by collision avoidance rates and traffic efficiency metrics. Moreover, the framework exhibits graceful degradation: as the number of agents increases beyond eight, performance remains stable rather than collapsing catastrophically as observed in many competing approaches. This scalability is essential for real-world deployment where traffic density varies dynamically, requiring consistent performance across diverse conditions.

The computational advantages of our framework are substantial. Compared to joint optimization approaches scaling exponentially with the number of agents, our sequential decomposition through Level-k reasoning achieves linear scaling. Empirical evaluations demonstrate computational speedups exceeding 60% compared to standard MCTS baselines in eight-agent scenarios, while maintaining or improving solution quality measured by collision avoidance rates and traffic efficiency metrics. Moreover, the framework exhibits graceful degradation: as the number of agents increases beyond eight, performance remains stable rather than collapsing catastrophically like many competing approaches. This scalability is essential for real-world deployment where traffic density varies dynamically, requiring consistent performance across diverse conditions.

1.3 Thesis Contributions

This thesis develops a unified, scalable, and safety-aware multi-agent decision-making framework for autonomous driving. The framework integrates cognitive hierarchy theory, MCTS, and heterogeneous vehicle modeling to address coordination challenges in both homogeneous and mixed traffic environments. This framework advances the state-of-the-art in several fundamental dimensions, offering both theoretical insights and practical solutions to longstanding challenges in multi-agent coordination. The specific contributions are organized as follows:

1. **Reconstructed Level-k Reasoning with Safety Guarantees:** We fundamentally reconceptualize Level-k reasoning by redefining Level-0 as a safety initialization procedure rather than a naive agent type. This reconstruction embeds safety as an emergent structural property of the reasoning hierarchy, transforming Level-k from a descriptive cognitive model into a constructive planning framework with provable collision-free properties. Unlike traditional formulations where safety is imposed as external constraints, our approach ensures that conservative Level-0 baselines propagate through higher reasoning levels via bounded rationality, guaranteeing worst-case safety margins throughout the hierarchy.
2. **Integration of Level-k Reasoning and MCTS:** We propose a novel integration architecture where Level-k reasoning decomposes the exponential joint action space into sequential reasoning steps, while MCTS provides adaptive sampling and uncertainty handling. This synergy reduces computational complexity from exponential to linear in the number of agents, achieving 21 orders of magnitude reduction compared to joint optimization in eight-agent scenarios. The integration incorporates safety-aware pruning, trajectory caching, and iterative refinement to ensure both computational efficiency and decision quality.

3. **Dual-Filtered Interaction Graph for Scalable Multi-Agent Reasoning:** We introduce a dual-filtered interaction graph that combines geometric proximity with reasoning-level hierarchy to identify relevant interaction partners. This dual filtering enables each agent to focus computational resources on relevant interaction partners, addressing the scalability challenge in dense multi-agent scenarios. The approach achieves up to 60% computational speedup compared to exhaustive interaction modeling while maintaining high decision quality.
4. **Heterogeneous Vehicle Modeling for Mixed Traffic Environments:** We develop a comprehensive behavioral modeling framework that captures both deterministic AV dynamics and probabilistic human driver behaviors. The framework combines physics-based motion models with learned distributions over driver intentions and risk preferences, enabling realistic simulation during MCTS rollouts. Adaptive safety margins dynamically adjust to estimated human uncertainty, ensuring robust collision avoidance even under behavioral deviations.
5. **Safety-Aware Multi-Objective Reward Design:** We design a multi-layered reward composition strategy that systematically balances safety, efficiency, comfort, and cooperative behavior. The reward structure incorporates dynamic safety thresholds for vehicle-to-vehicle, vehicle-to-human, and vehicle-to-road interactions, providing comprehensive safety assessment across diverse interaction types. This formulation enables AVs to navigate complex tradeoffs inherent in mixed traffic coordination.
6. **Comprehensive Validation in Symmetric and Mixed Traffic Scenarios:** We conduct extensive simulation studies across multiple scenario classes, ranging from maximally symmetric eight-agent intersections to heterogeneous mixed traffic with varying traffic flows and driver behaviors. Our framework achieves 95-100% success rates in scenarios where baseline methods face deadlock or collision, demonstrating 30-40% improvements in both safety and efficiency metrics. The validation covers diverse traffic densities, intersection geometries, and human driving styles, establishing the framework's robustness and generalizability.

These contributions collectively establish a new paradigm for multi-agent decision-making in autonomous driving that reconciles strategic sophistication with computational tractability, safety guarantees with behavioral realism, and theoretical rigor with practical deployability. The framework addresses fundamental limitations of existing approaches while offering a modular and extensible architecture suitable for diverse autonomous driving applications.

1.4 Thesis Organization

The remainder of this thesis is organized as follows.

Chapter 2 provides a comprehensive review of related work, examining existing approaches to multi-agent coordination, game-theoretic reasoning, sampling-based planning methods, and mixed traffic modeling in autonomous driving contexts. We analyze the strengths and limitations of each approach, establishing the motivation for our integrated framework.

Chapter 3 establishes the baseline MCTS framework for multi-agent coordination at unsignalized intersections. We formalize the symmetric coordination problem, develop comprehensive agent models including vehicle dynamics and collision detection, and present the multi-objective optimization formulation. The MCTS planning algorithm is introduced with safety-validated expansion and rollout simulation. Experimental validation on four-agent left-turn scenarios demonstrates that MCTS with sufficient planning horizon ($H = 9$) achieves zero collision rate, establishing the baseline upon which subsequent chapters build. We also identify scalability limitations: the joint action space grows exponentially from 10^{42} for four agents to 10^{85} for eight agents, motivating the extensions in subsequent chapters.

Chapter 4 addresses these scalability limitations through two complementary mechanisms. The dynamic interaction graph with dual spatial-strategic filtering reduces effective opponent modeling from $N-1$ agents to approximately 3–4 relevant agents. The reconstructed Level-k cognitive hierarchy decomposes multi-agent coordination into sequential single-agent optimizations, where Level-0 serves as a universal safety initialization proced-

ure generating conservative baseline trajectories. The cascading safety property ensures that higher reasoning depth strengthens rather than compromises safety guarantees. The integrated MCTS-Level-k algorithm achieves 21 orders of magnitude complexity reduction through safety-aware pruning, trajectory caching, and interaction filtering. Experimental validation on symmetric eight-agent scenarios demonstrates zero collision rate where baseline methods exhibit 15–35% collisions, with computation times enabling real-time deployment.

Chapter 5 extends the framework to heterogeneous mixed traffic environments where AVs must coordinate with HDVs exhibiting diverse and uncertain behaviors. We develop style-aware behavior prediction based on the Intelligent Driver Model, using a parameter $\eta_h \in [0, 1]$ to represent driving style diversity. Time-varying uncertainty quantification captures the degradation of prediction confidence over the planning horizon. Adaptive safety thresholds respond to interaction-specific risks including relative velocity, heading conflicts, and spatial location. The unified V2H risk metric integrates seamlessly with Level-k reasoning for robust collision avoidance. Experimental validation across penetration rates from 20% to 100% demonstrates consistent performance: collision rates below 2% at 50% penetration despite diverse human driving behaviors, with graceful degradation from efficient coordination in AV-dominated traffic to defensive navigation in human-dominated scenarios.

Chapter 6 concludes the thesis by summarizing key findings and contributions. We discuss broader implications for multi-agent systems and autonomous driving, analyze limitations of the current framework, and outline promising directions for future research including adaptive level assignment, formal safety verification, and extensions to more complex traffic scenarios.

This thesis demonstrates that integrating cognitive hierarchy theory with search-based planning can address key challenges in multi-agent coordination. By reconstructing Level-k reasoning to prioritize safety, leveraging MCTS for adaptive exploration, and carefully modeling heterogeneous agent behaviors, we establish a framework that is simultaneously scalable, safe, and strategically sophisticated. This framework not only advances the theoretical understanding of multi-agent decision-making but also provides practical solutions

applicable to real-world autonomous driving systems. As AVs transition from controlled testing environments to complex urban traffic, frameworks that can guarantee safety while reasoning strategically about diverse interactions become increasingly critical. This thesis contributes a significant step toward that vision, offering a foundation upon which future multi-agent coordination systems can be built.

Chapter 2

Literature Review

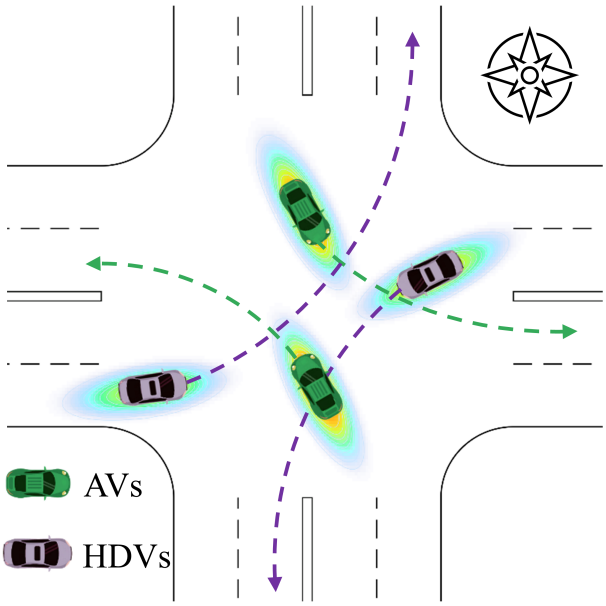
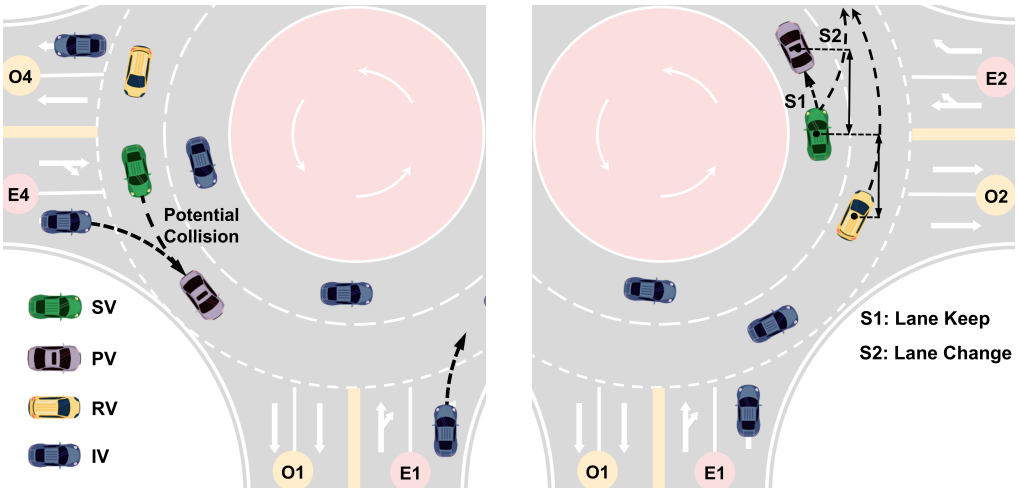


Figure 2.1: Mixed traffic interaction at an unsignalized intersection.

The challenges outlined in the introduction necessitate a comprehensive review of existing approaches to multi-agent coordination, strategic reasoning, and decision-making under uncertainty in autonomous driving contexts, as shown in Fig. 2.1. The scenario adopts a right-hand traffic convention, consistent with the predominant global standard. As the proposed framework is purely algorithmic, based on multi-agent MCTS and Level-k reasoning, it does not depend on any dataset tied to a specific driving convention, and generalises directly to left-hand traffic through a simple geometric mirroring.

This chapter systematically examines the state-of-the-art across multiple research domains, critically analyzing their strengths and limitations to establish the motivation for our integrated framework. We organize this review according to methodological paradigms rather than application domains, as the fundamental computational and safety challenges transcend specific traffic scenarios.

Our review reveals a persistent tension in existing work: methods that achieve strategic sophistication often sacrifice computational tractability or safety guarantees, while computationally efficient approaches typically employ overly simplified models of agent interactions. This tension motivates the central argument of this thesis: resolving it requires not incremental improvements, but a reformulation of how cognitive hierarchy, search-based planning, and safety constraints interact.



(a) (b)

Figure 2.2: Challenging scenarios in multi-vehicle interactions: (a) potential collision at an unsignalized intersection, where E1–E4 denote the four entrances and O1–O4 denote the four outlets of the intersection; (b) emergency lane-change with two strategies (S1, S2), where the subject vehicle (SV) interacts with the preceding vehicle (PV), rear vehicle (RV), and irrelevant vehicle (IV).

2.1 Multi-Agent Coordination Approaches

Multi-agent coordination in autonomous systems poses significant computational and safety challenges as illustrated in Fig. 2.2. The fundamental difficulty arises from the exponential growth of the joint state-action space as the number of agents increases, coupled with the need for real-time decision-making under uncertainty and partial observability.

2.1.1 Centralized versus Decentralized Coordination

Coordination approaches at unsignalized intersections can be broadly categorized by their decision-making architecture, as illustrated in Fig. 2.1. Early work in multi-agent intersection management largely focused on centralized approaches that assume full information sharing through vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication, along with a trusted coordinator [35–37]. These methods formulate coordination as a joint optimization problem where a central planner computes optimal trajectories for all agents simultaneously. The centralized paradigm offers theoretical optimality guarantees and can enforce global constraints such as collision avoidance through explicit coupling in the optimization formulation. Reservation-based systems exemplify this approach, where vehicles communicate arrival times and desired trajectories to a central intersection manager that grants or denies passage based on conflict resolution algorithms.

However, centralized coordination suffers from severe scalability limitations as joint action spaces grow exponentially with agent count [38]. Even with modern computational resources, joint trajectory optimization for more than four to five vehicles becomes intractable when considering realistic planning horizons and continuous action spaces. Beyond computational complexity, centralized approaches introduce single points of failure and require reliable V2I communication, limiting applicability in communication-constrained or infrastructure-degraded environments. Privacy concerns also arise, as vehicles must share trajectory intentions with external systems [39, 40].

In response to these limitations, recent research has shifted toward decentralized frameworks that distribute decision-making across agents while maintaining coordination quality [41]. As shown in Fig. 2.3, in mixed traffic scenarios with both AVs and HDVs, decentralized approaches enable each vehicle to reason about others' intentions at different cognitive levels (Level-0, Level-k) through V2V communication, without relying on centralized infrastructure. Decentralized approaches formulate coordination as a multi-agent optimization problem where each agent optimizes its own objective while accounting for other agents' decisions through prediction, communication, or implicit coordination mechanisms [42, 43]. This paradigm trades theoretical optimality for improved scalability, robustness to communication failures, and preservation of agent privacy.

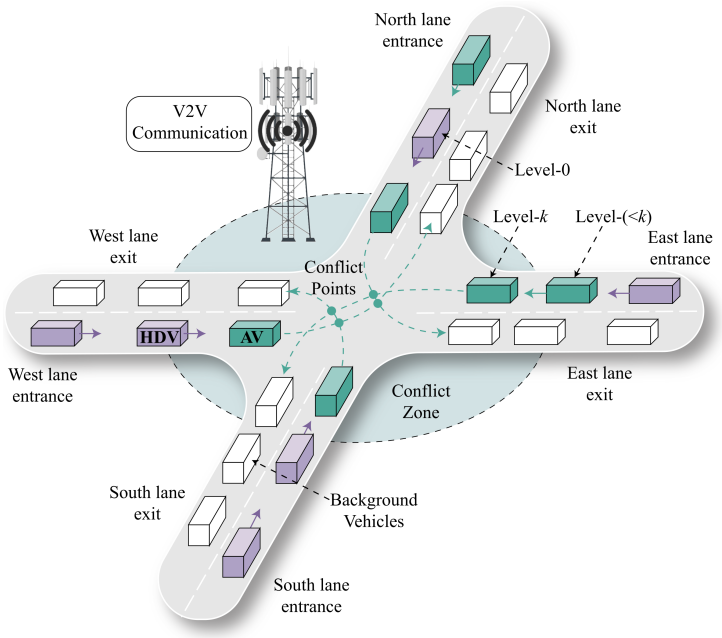


Figure 2.3: Illustration of unsignalized intersection scenario in mixed traffic. The scene includes AVs (shown in green), HDVs (shown in purple), and background vehicles (shown in white). Vehicles coordinate through V2V communication while reasoning at different cognitive levels (Level-0, Level-k, Level-$(k-1)$). Conflict zones and potential collision points are highlighted, demonstrating the complexity of multi-agent coordination in heterogeneous traffic environments.

The spectrum between fully centralized and fully decentralized coordination includes various hybrid architectures. Hierarchical approaches decompose the coordination problem into multiple levels, with high-level strategic decisions made centrally and low-level trajectory execution performed locally. Market-based coordination mechanisms treat intersection passage as a resource allocation problem, where agents bid for priority through auction protocols [14, 44]. Consensus-based methods iterate between local planning and information exchange until agents reach agreement on conflict-free trajectories. While consensus-based approaches can produce high-quality coordinated solutions, they typically require multiple rounds of communication and convergence is not guaranteed in dynamic environments with time-varying agent interactions, limiting their applicability in real-time intersection coordination.

Despite architectural diversity, all coordination approaches face the fundamental challenge of computational tractability versus coordination quality. Centralized methods achieve near-optimal coordination at prohibitive computational cost, while purely decentralized reactive approaches scale well but often produce conservative or suboptimal behaviors [45,

46]. Existing hybrid systems such as hierarchical planners and market-based methods partially address this tradeoff but introduce their own limitations: hierarchical decompositions rely on predefined priority structures that may fail under symmetric conflicts, while auction-based mechanisms require explicit communication infrastructure. Our work addresses this tradeoff through a communication-free decomposition strategy grounded in cognitive hierarchy theory, which sequences the optimization process without centralized coordination or predefined priority rules. Within the evaluated scenario of eight-agent unsignalized intersection coordination, this approach achieves lower collision rates and trajectory deviation compared to game-theoretic baselines (Stackelberg, Nash) and vanilla MCTS, while maintaining sub-100 ms per-step planning. We note that direct quantitative comparison with all existing hybrid systems is beyond the scope of this thesis, and claims of superiority are limited to the evaluated baselines and simulation conditions.

2.1.2 Communication Requirements and Coordination

Communication plays a critical role in multi-agent coordination, enabling agents to share intentions, negotiate priorities, and establish mutual understanding as shown in Fig. 2.3. V2V and V2I communication protocols have been extensively studied, with standards such as Cellular Vehicle-to-Everything providing technical foundations. Communication-based coordination can be explicit, where agents directly negotiate through structured protocols, or implicit, where agents infer others' intentions through observed behaviors [47–49].

However, reliance on communication introduces vulnerabilities. Communication channels may be unreliable due to interference, bandwidth limitations, or malicious attacks. Privacy concerns arise when vehicles must broadcast detailed trajectory information [50, 51]. Latency in communication can undermine time-critical coordination, particularly at high-speed intersections. Furthermore, the assumption of reliable communication excludes scenarios involving legacy vehicles without communication capabilities or unexpected communication infrastructure failures in complex environments.

These limitations have motivated research into implicit coordination mechanisms that enable multi-agent cooperation without explicit message passing. Implicit coordination relies on predictable behavioral patterns, social conventions, and mutual observation to establish shared expectations [52]. Game-theoretic frameworks provide formal models of implicit coordination through concepts such as focal points and common knowledge. Learning-based approaches can discover emergent coordination strategies through multi-agent training without explicit communication protocols.

Our framework operates in a middle ground, assuming agents can observe each other's states through onboard sensing but do not require structured communication protocols. This assumption aligns with current AV capabilities, where perception systems can track nearby vehicles but reliable inter-vehicle communication remains limited in mixed traffic scenarios involving HDVs. The Level-k reasoning component provides a structured model of implicit coordination, where agents coordinate through recursive belief modeling rather than explicit negotiation. This distinguishes our approach from purely reactive decentralised systems, which respond only to current observations without modeling opponent intent. Reactive systems scale well but tend to produce overly conservative or deadlock-prone behaviors in symmetric high-density scenarios, as demonstrated by the vanilla MCTS baseline in our simulations. By contrast, the Level-k hierarchy enables anticipatory planning over a finite horizon, producing more efficient and stable coordination without requiring communication infrastructure.

2.2 Game-Theoretic Strategic Reasoning

Game theory provides a mathematical framework for modeling strategic interactions among agents with conflicting objectives [18, 53, 54]. In autonomous driving contexts, game-theoretic formulations capture the inherent tension between individual efficiency and collective safety, enabling vehicles to reason about how their actions influence others' responses and vice versa in a dynamic traffic environment.

2.2.1 Nash Equilibrium and Stackelberg Games

Nash equilibrium concepts form the foundation of game-theoretic approaches to multi-agent coordination [18, 21, 22]. A Nash equilibrium represents a strategy profile where no agent can improve its utility through unilateral deviation, given that other agents maintain their equilibrium strategies. This solution concept captures stable outcomes where agents' beliefs about others' behaviors are mutually consistent and no agent intends to deviate.

Application of Nash equilibrium to intersection coordination typically formulates vehicle interactions as non-cooperative games where each agent's strategy space consists of possible trajectories or control sequences, and payoff functions encode preferences over safety, efficiency, and comfort [55]. Dynamic games extend this formulation to sequential decision-making, where agents observe the unfolding game state and update strategies over time. Differential games provide continuous-time formulations particularly suited to vehicle dynamics, where strategies are feedback control policies and equilibria are characterized through coupled Hamilton-Jacobi-Bellman equations.

However, Nash equilibrium approaches face fundamental computational and conceptual challenges. Computing equilibria requires exploring the full joint action space, making them computationally prohibitive beyond two to three agents [19, 33]. Even when equilibria can be computed efficiently through iterative best-response algorithms or gradient-based methods, the resulting strategies may not align with desired coordination outcomes. Multiple equilibria often exist in symmetric scenarios, requiring additional selection mechanisms that may conflict with safety objectives [56]. Moreover, the assumption of perfect rationality underlying Nash equilibrium rarely holds in mixed traffic where human drivers exhibit bounded rationality, inconsistent preferences, and limited cognition.

Stackelberg games [19, 57, 58] offer an alternative formulation where agents are organized in a leader-follower hierarchy. The leader commits to a strategy first, and followers best-respond to the leader's commitment. This sequential structure eliminates equilibrium multiplicity issues present in simultaneous-move games and provides a natural model for

scenarios where AVs can credibly signal intentions to human drivers. Recent work has demonstrated Stackelberg formulations for lane-changing and merging scenarios, where AVs act as leaders by clearly signaling intended maneuvers, enabling human drivers to respond predictably.

Despite these advantages, Stackelberg games introduce their own challenges. Determining appropriate leader-follower assignments may be unclear in symmetric scenarios or require communication protocols that may not be reliable. The assumption that leaders can commit to strategies before followers respond may not hold when decision-making occurs on similar timescales. Furthermore, Stackelberg equilibria remain computationally expensive to compute in multi-agent settings with more than two to three participants.

Hierarchical game-theoretic approaches have emerged to address scalability challenges through problem decomposition [38, 59]. These methods partition complex multi-agent scenarios into tractable subgames based on spatial or temporal separation [20]. For example, intersection coordination can be decomposed into pairwise or small-group interactions based on predicted collision risks. Hierarchical decomposition enables application of sophisticated game-theoretic solutions to local interactions while avoiding the combinatorial explosion of joint strategy spaces.

Recent developments include iterative linear-quadratic game solvers that exploit the structure of vehicle dynamics to compute approximate Nash equilibria efficiently. These methods linearize dynamics around nominal trajectories and employ iterative refinement to handle nonlinearities, achieving computational efficiency comparable to single-agent optimization [33, 60–62]. However, scalability remains limited to modest numbers of agents (typically four to six), and the approximations introduced through linearization and local convergence may compromise solution quality or safety guarantees [41].

The fundamental limitation of equilibrium-based approaches lies in their reliance on perfect rationality and complete information assumptions. In mixed traffic scenarios, human drivers do not optimize objective functions explicitly, exhibit inconsistent preferences across contexts and individuals, and operate under cognitive and perceptual limitations.

Applying game-theoretic models designed for perfectly rational agents to predict human behavior often produces systematic mismatches, leading to coordination failures or unsafe interactions. This observation has motivated the development of bounded rationality models, including the Level-k reasoning framework examined in the next section.

2.2.2 Computational Tractability and Equilibrium Selection

Beyond conceptual limitations, game-theoretic approaches face severe computational challenges in real-time autonomous driving applications. Computing Nash equilibria in general-form games is PSPACE-complete, placing it among the most computationally demanding problems in complexity theory [63, 64]. While special structure in vehicle coordination problems—such as continuous strategy spaces, differentiable payoffs, and convex constraints—enables more efficient solution algorithms, the fundamental complexity grows exponentially with the number of agents and planning horizon.

Approximate solution methods trade optimality for computational feasibility through various relaxation strategies. Iterative best-response algorithms compute sequences of local optima that may converge to Nash equilibria under certain conditions, but convergence is not guaranteed and can be slow when the game exhibits weak contraction properties [65]. Gradient-based methods formulate equilibrium computation as constrained optimization and employ standard solvers, but suffer from local optima in non-convex games typical of multi-agent driving scenarios. Sampling-based approaches approximate equilibria through Monte Carlo simulation or evolutionary algorithms, but require careful tuning to balance exploration and convergence.

Even when equilibria can be computed efficiently, selecting among multiple equilibria presents additional challenges. Generic games typically admit multiple Nash equilibria with different payoff profiles and stability properties. In symmetric intersection scenarios, the symmetry structure guarantees existence of symmetric equilibria but also creates multiple asymmetric equilibria corresponding to different priority assignments. Without additional coordination mechanisms, agents may coordinate on suboptimal or unsafe equilibria, or fail to coordinate altogether [66, 67].

Equilibrium refinement concepts from game theory provide principled selection criteria. Subgame perfection eliminates equilibria sustained by non-credible threats in sequential games. Pareto optimality selects equilibria that cannot be improved for all agents simultaneously. Risk dominance favors equilibria robust to uncertainty in others' strategy choices. However, these refinements often fail to uniquely identify equilibria in complex multi-agent scenarios, and computing refined equilibria increases computational cost.

The computational intractability of equilibrium concepts motivates our framework's departure from pure game-theoretic formulations. Rather than seeking exact equilibria through exhaustive search or iterative optimization, we employ Level-k reasoning to construct approximate equilibria through bounded recursive reasoning, combined with MCTS to explore high-value regions of the strategy space through selective sampling. This hybrid approach achieves computational efficiency comparable to heuristic methods while retaining strategic sophistication derived from game-theoretic principles.

2.2.3 Cognitive Hierarchy and Level-k Reasoning

Level-k reasoning emerged from behavioral economics as an alternative to perfect rationality assumptions underlying classical game theory. The framework models bounded rationality through a hierarchy of reasoning depths, where agents at different levels employ increasingly sophisticated models of opponent behavior [26, 27, 68]. As illustrated in Fig. 2.3, this cognitive hierarchy approach enables vehicles to reason at different levels (Level-0, Level-k, Level-$(k-1)$) in mixed traffic scenarios, providing more realistic models of human decision-making while maintaining computational tractability compared to traditional equilibrium computation methods.

The classical Level-k framework defines a recursive reasoning structure [69]. Level-0 agents employ naive or non-strategic behaviors, often modeled as random action selection or simple heuristics such as maintaining constant velocity. Level-1 agents best-respond to the assumption that all other agents are Level-0, optimizing their own objectives given

these behavioral predictions [70]. More generally, Level-k agents (for $k \geq 1$) model all other agents as Level-$(k-1)$ and compute best responses accordingly. This recursive structure naturally breaks symmetry in coordination problems, as agents at different reasoning levels arrive at different strategies even when facing identical physical situations.

The cognitive plausibility of Level-k reasoning has been validated through extensive experimental studies in behavioral economics and psychology. Human subjects exhibit reasoning patterns consistent with Level-1 to Level-3 thinking across various strategic games, with higher reasoning levels becoming progressively rarer. The distribution of reasoning levels varies across populations and contexts, with experienced decision-makers and strategic thinkers more likely to employ higher-level reasoning. This empirical grounding makes Level-k models particularly attractive for mixed traffic scenarios where AVs must interact with human drivers exhibiting heterogeneous reasoning capabilities in real-world environments [71].

Application of Level-k reasoning to autonomous driving has demonstrated promising results in small-scale scenarios. [26] integrated Level-k reasoning with trajectory prediction, showing improved accuracy in modeling human driving behavior compared to models assuming perfect rationality [27]. Their work successfully predicted interactions between two to three agents using model predictive control at each reasoning level, but did not scale to dense traffic scenarios involving larger numbers of vehicles. Similarly, recent work applied Level-k models to pedestrian-agent interactions, demonstrating that bounded rationality assumptions better capture human decision-making than perfect rationality [68]. However, these implementations focused primarily on single pedestrian scenarios rather than multi-agent coordination involving multiple vehicles [69].

Theoretical analysis of Level-k reasoning in multi-agent systems has identified both strengths and fundamental limitations. [70] provided a comprehensive survey of autonomous agents modeling other agents, highlighting Level-k as a promising approach for tractable multi-agent planning. They identified the key computational challenge of solving each level's optimization problem efficiently, particularly as agent numbers increase [71]. While their analysis is thorough, implementations for scenarios beyond four to five agents remained elusive, leaving a gap between theory and practice [72].

The critical limitation of classical Level-k formulations lies in the definition of Level-0 behavior. Modeling Level-0 as random or myopic actors creates fundamentally unsafe baselines that propagate through the entire reasoning hierarchy. When Level-1 agents optimize assuming others behave randomly, and Level-2 agents optimize assuming others optimize against random agents, the resulting strategies lack safety guarantees. A single Level-0 vehicle behaving randomly in the intersection scenario shown in Fig. 2.3 could lead to collisions that higher-level reasoning fails to prevent, as the recursive structure builds upon unsafe foundations.

This fundamental flaw has limited the application of Level-k reasoning to safety-critical autonomous driving despite its cognitive plausibility and computational advantages. Previous implementations either restrict Level-0 to deterministic behaviors that lack realism, introduce external safety constraints that undermine the elegance of the recursive structure, or accept residual collision risks that are unacceptable in real-world deployment. Our framework addresses this limitation through a reconceptualization that redefines Level-0 not as a behavioral model but as a safety initialization procedure, transforming Level-k from a descriptive theory into a planning framework with provable safety.

Beyond safety concerns, classical Level-k implementations face scalability challenges [73]. Computing best responses at each level requires solving optimization problems over continuous trajectory spaces, which becomes computationally expensive as planning horizons extend and state-action spaces grow. Naive implementations evaluate Level-$(k-1)$ predictions at every Level-k optimization step, creating recursive computational dependencies that multiply cost exponentially with reasoning depth [74]. Furthermore, maintaining beliefs over multiple agents' reasoning levels and computing best responses to heterogeneous populations introduces combinatorial complexity even when individual optimization problems remain tractable.

Recent work has begun addressing these computational challenges through various approximations and algorithmic innovations [75]. Hierarchical decomposition strategies partition multi-agent scenarios into smaller subproblems based on spatial or temporal locality, enabling parallel computation of Level-k responses for non-interacting agent groups. Sampling-based approximations replace exhaustive best-response computation

with Monte Carlo evaluation over discrete action sets, trading optimality for computational efficiency. Trajectory caching and reuse mechanisms exploit temporal coherence in multi-agent interactions, avoiding redundant computations across planning cycles [76]. However, these techniques have been demonstrated primarily in small-scale scenarios and have not been systematically integrated into comprehensive frameworks scaling to eight or more agents in real-time.

The theoretical properties of Level-k reasoning provide both opportunities and challenges for intersection coordination. The asymmetric reasoning structure naturally resolves symmetry in coordination problems, as agents at different levels compute different strategies even from identical initial conditions. This symmetry-breaking property is particularly valuable in scenarios such as eight-vehicle intersections where all agents arrive simultaneously with equal geometric priority. However, the recursive nature of Level-k reasoning requires careful management of computational resources, as reasoning depth directly impacts both solution quality and computational cost [77].

2.3 Sampling-Based Planning Methods

Sampling-based planning algorithms provide an alternative paradigm that explores state-action spaces through selective sampling rather than exhaustive enumeration. These methods have proven particularly effective in high-dimensional continuous spaces where systematic discretization becomes infeasible, making them natural candidates for multi-agent trajectory planning in autonomous driving contexts.

2.3.1 Monte Carlo Tree Search Fundamentals

MCTS has emerged as one of the most successful sampling-based planning algorithms, achieving superhuman performance in complex domains such as Go, chess, and real-time strategy games [23]. MCTS builds asymmetric search trees through iterative simulation, progressively refining value estimates and expanding the tree toward promising regions of the search space. The algorithm's fundamental appeal lies in its ability to balance exploration of uncertain regions with exploitation of known high-value actions without requiring domain-specific heuristics.

The standard MCTS algorithm operates through four phases repeated iteratively: selection, expansion, simulation, and backpropagation. During selection, the algorithm traverses the existing search tree from the root node representing the current state, choosing child nodes according to a selection policy that balances exploration and exploitation. The most common selection policy, Upper Confidence Bounds (UCB) applied to Trees, treats node selection as a multi-armed bandit problem, choosing actions that maximize the sum of empirical value estimates and an exploration bonus proportional to visit count uncertainty in the search process.

Upon reaching a leaf node of the current tree, the expansion phase adds one or more child nodes representing unexplored actions from that state. The newly added node is then evaluated through simulation, where a rollout policy generates a complete trajectory from the expanded state to a terminal condition or fixed horizon. This rollout policy may be random, domain-specific heuristics, or learned policies depending on the application. The simulation produces an outcome value that is backpropagated up the tree, updating value estimates and visit counts for all nodes along the path from the expanded node to the search tree root.

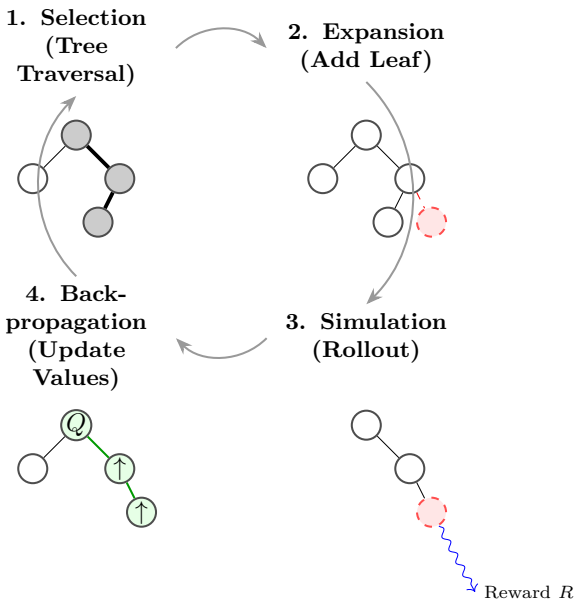


Figure 2.4: Standard MCTS algorithm workflow. The algorithm iteratively performs four phases: (1) Selection: traverse the tree using the UCB policy; (2) Expansion: add new child nodes; (3) Simulation: rollout to a terminal state; (4) Backpropagation: update values along the path. This process builds an asymmetric tree focusing computational resources on promising regions.

This iterative process as shown in Fig. 2.4 gradually builds an asymmetric tree that devotes more computational resources to promising regions while maintaining some exploration of alternatives. After a fixed computational budget (measured in simulation count or wall-clock time), the algorithm selects an action from the root node according to some final selection criterion, typically choosing the most-visited child or the child with highest average value. The chosen action is executed, the root node advances accordingly, and the process repeats for the next decision cycle.

The theoretical foundations of MCTS provide convergence guarantees under certain conditions. For finite-horizon problems with bounded rewards, the algorithm converges almost surely to optimal policies as simulation count approaches infinity, provided the selection policy ensures all state-action pairs are visited infinitely often. The rate of convergence depends on the branching factor, planning horizon, and effectiveness of the rollout policy in evaluating positions. In practice, finite computational budgets mean MCTS produces approximate solutions, with quality improving monotonically as more simulations are performed during the search process.

Application of MCTS to autonomous driving has shown promise in handling uncertainty and complex state spaces [25,78]. Unlike offline reinforcement learning that requires extensive pre-training, MCTS performs online planning that adapts to current environmental states without requiring prior experience in similar situations. The algorithm naturally incorporates stochasticity through probabilistic rollout policies and can model uncertain opponent behaviors by sampling from belief distributions during simulation. These properties make MCTS particularly suitable for mixed traffic scenarios where human driver behaviors exhibit variability and cannot be predicted deterministically.

The work in [25] combined MCTS with social attention mechanisms for highway driving, demonstrating that sampling-based planning can handle uncertainty in other drivers' intentions. Their approach used learned rollout policies to simulate realistic traffic interactions and employed attention mechanisms to focus computational resources on relevant nearby vehicles. However, scalability remained limited to two to three interacting vehicles due to the exponential growth of joint action spaces and the prohibitive computational cost of realistic, high-fidelity simulation.

Extensions incorporating upper confidence bounds have shown improved exploration-exploitation tradeoffs [79]. By treating action selection as a bandit problem, UCB-based selection policies provide theoretical guarantees on regret accumulation and enable MCTS to identify near-optimal actions with logarithmic sample complexity. These theoretical properties have motivated widespread adoption of UCB-based selection in MCTS implementations across diverse domains.

2.3.2 MCTS in Multi-Agent Settings

Extending MCTS to multi-agent settings introduces fundamental challenges distinct from single-agent planning. In single-agent MCTS, the environment provides stochastic transitions but does not adapt strategically to the agent's policy. In multi-agent settings, other agents observe the ego agent's actions and respond strategically, creating complex interdependencies that violate the Markov assumption underlying standard MCTS convergence proofs in traditional settings.

Naive application of single-agent MCTS to multi-agent scenarios assumes other agents follow fixed policies or behave randomly during rollouts [25]. This assumption enables tractable planning but produces overly optimistic value estimates when other agents actually respond strategically to the ego agent's actions. When all agents simultaneously execute their "optimal" actions computed under fixed-policy assumptions, collisions become inevitable as the optimistic predictions fail to account for mutual interference [80].

Several extensions address opponent modeling in multi-agent MCTS. Centralized MCTS formulations treat multi-agent coordination as a joint planning problem, where each tree node represents a joint state and each edge represents a joint action by all agents [81, 82]. This approach can compute coordinated strategies but suffers from exponential growth in branching factor with agent count. For n agents each with b available actions, the joint action space contains b^n branches at each node, making deep search prohibitive even for modest agent counts and action sets.

Decentralized MCTS implementations allow each agent to build its own search tree over local actions while modeling other agents through belief distributions or deterministic predictions [83]. These approaches achieve better scalability but require each agent to maintain beliefs over other agents' policies, which itself becomes computationally expensive as agent count increases. Approximate belief representations through scenario sampling or policy compression reduce computational burden but introduce approximation errors that may compromise safety or coordination quality.

Communication-based multi-agent MCTS enables agents to share information about tree structures, value estimates, or intended actions during planning. This information exchange can improve coordination by aligning agents' beliefs and enabling them to simulate each other's responses more accurately. However, communication requirements may be impractical in bandwidth-constrained or adversarial environments, and communication latency can undermine real-time planning when agents' trees evolve asynchronously.

Recent work has explored integrating MCTS with partially observable Markov decision processes for cooperation-aware planning [65, 82]. These methods model uncertainty over other agents' observations and intentions through belief states, enabling robust planning under incomplete information. Simulation-based belief updates during MCTS rollouts

provide approximate solutions to partially observable Markov decision process (POMDP) planning that avoid the computational intractability of exact belief-space planning. However, the computational demands of POMDP-based MCTS have limited scalability to three to four agents in practice [84].

Qi et al. [80] directly applied MCTS to intersection management through centralized planning, demonstrating effective coordination of up to four agents. Their approach relied on reliable V2I communication and centralized computation, introducing single points of failure and scalability limitations. While their results demonstrated MCTS viability for intersection scenarios, the centralized architecture limits applicability to scenarios where decentralized decision-making is required due to inherent communication constraints or significant privacy considerations in public road networks.

The fundamental challenge in multi-agent MCTS lies in the tension between accurate opponent modeling and computational tractability. Sophisticated opponent models that capture strategic reasoning and adaptation improve planning quality but incur exponential computational cost due to recursive simulation and belief maintenance [85]. Conversely, simplified opponent models reduce computational burden but may introduce systematic bias in value estimation, potentially leading to coordination failures or safety violations. In this work, we adopt a structured simplification strategy that balances these competing requirements. Specifically, human-driven vehicles are modeled using physics-based dynamics with uncertainty, while autonomous agents are modeled through Level-k reasoning. This hybrid formulation preserves essential behavioral characteristics while maintaining tractable computation, enabling robust performance in mixed traffic scenarios. We acknowledge that these modeling choices introduce limitations: constant-velocity prediction at Level-0 becomes less accurate under aggressive acceleration or deceleration, and IDM captures only nominal longitudinal car-following behavior. These limitations and their implications for real-world deployment are discussed in Chapter 6.

This tradeoff has limited prior multi-agent MCTS implementations to scenarios with few agents (typically two to four). Our framework addresses this through integration with Level-k reasoning, detailed in Chapter 4.

2.3.3 Computational Complexity and Pruning Strategies

The computational complexity of MCTS scales with the product of branching factor, planning horizon, and simulation cost. In autonomous driving applications, these factors can be substantial: continuous action spaces discretized into dozens of primitives create large branching factors, safety requirements necessitate planning horizons of several seconds (corresponding to dozens of decision steps), and realistic simulation of vehicle dynamics and multi-agent interactions introduces significant per-simulation cost.

Pruning strategies reduce effective branching factors by eliminating provably suboptimal or unsafe actions early in the search process. Progressive widening techniques gradually increase the number of child nodes explored from each parent as visit count increases, focusing initial exploration on a small action set and expanding only when sufficient samples suggest benefit. Domain-specific pruning can eliminate actions that violate kinematic constraints, geometric feasibility, or safety conditions before adding them to the search tree.

Safety-aware pruning is particularly effective in autonomous driving [86]. Many potential actions can be identified as unsafe through geometric reasoning or conservative reachability analysis without requiring detailed simulation. Eliminating these unsafe actions from the search tree dramatically reduces effective branching factors while guaranteeing that explored policies satisfy minimum safety criteria. Our framework employs safety-aware pruning enabled by the conservative Level-0 baseline, which provides tight lower bounds on minimum safe spacing that can be checked efficiently during the tree expansion and node generation process.

Trajectory caching and reuse exploit temporal coherence in multi-agent planning problems. When the environment evolves slowly relative to planning frequency, portions of the search tree computed at time t remain relevant at time $t+1$. Reusing subtrees from previous planning cycles avoids redundant computation and enables deeper search within fixed computational budgets. However, tree reuse requires careful management of stale information, as opponent policies and environmental conditions may change significantly between successive planning cycles.

Action abstraction reduces branching factors by grouping similar actions into equivalence classes. Rather than considering all possible discretizations of continuous control inputs, the planner reasons over higher-level action abstractions such as "merge left aggressively" or "yield to oncoming traffic." These abstract actions are expanded into detailed trajectories only after tree search identifies promising high-level strategies. Multi-resolution action abstractions enable coarse initial exploration followed by refined optimization of selected strategies within the search process.

Despite these optimization techniques, vanilla MCTS implementations struggle with multi-agent scenarios involving more than four to five vehicles under real-time constraints. The exponential growth in joint action spaces overwhelms pruning and caching strategies when agent count exceeds modest thresholds. This scalability barrier has motivated our integration of MCTS with Level-k reasoning, which fundamentally restructures the search problem to avoid joint action space explosion through sequential decomposition guided by cognitive hierarchy.

2.4 Mixed Traffic Modeling and Human Behavior

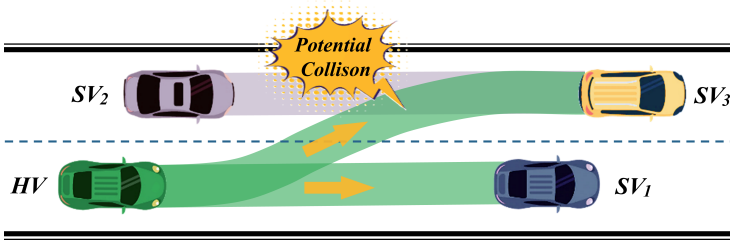


Figure 2.5: During lane-changing, the autonomous host vehicle (HV) needs to interact with surrounding vehicles (SVs) to make accurate decisions and plan collision-free trajectories. The green transparent region represents potential trajectories of the HV, with the arrows indicating HV's movement directions.

The transition from controlled testing environments with only AVs to real-world deployment necessitates effective modeling of mixed traffic scenarios where autonomous and HDVs coexist as shown in Fig. 2.5. This heterogeneity introduces fundamental challenges distinct from homogeneous multi-agent coordination, as AVs must reason about agents operating under fundamentally different decision-making paradigms.

2.4.1 Human Driver Behavior Models

Human driving behavior exhibits complex patterns shaped by perception, cognition, social norms, and individual preferences [87, 88]. Early approaches to modeling human drivers employed physics-based models that capture fundamental aspects of car-following, lane-changing, and gap acceptance behaviors. The Intelligent Driver Model (IDM) is widely adopted for longitudinal control, capturing smooth acceleration, deceleration, and traffic jam formation through a differential equation relating acceleration to desired velocity and safe following distance [89].

IDM's success stems from its ability to reproduce emergent traffic patterns observed empirically, including stop-and-go waves and capacity drops at bottlenecks, using only a small number of interpretable parameters [90]. Extensions to IDM incorporate lane-changing decisions through gap acceptance models that evaluate available spaces in adjacent lanes based on safety and efficiency criteria. The Minimizing Overall Braking Induced by Lane changes (MOBIL) lane-changing model computes incentive criteria balancing the lane-changer's acceleration improvement against impacts on surrounding vehicles, capturing cooperative aspects of human lane-changing decisions.

However, deterministic models like IDM capture only average or nominal behaviors and fail to represent the significant variability observed across human drivers and contexts. Individual drivers exhibit different preferences for following distances, acceleration aggressiveness, and risk tolerance. The same driver may behave differently depending on urgency, fatigue, distraction, or traffic density. This behavioral heterogeneity cannot be captured by single-parameter-set deterministic models and requires probabilistic frameworks that represent distributions over behaviors rather than single trajectories.

Probabilistic extensions to physics-based models address behavioral variability through parameter distributions or stochastic perturbations [77]. Rather than assuming all drivers follow IDM with fixed parameters, mixed traffic simulators sample parameters from learned distributions representing population heterogeneity. Time-varying parameter mod-

els capture context-dependent behavior changes, such as more aggressive driving in sparse traffic or more conservative behavior near intersections. Stochastic differential equation formulations add noise terms to deterministic dynamics, representing moment-to-moment variability in human control inputs.

Data-driven approaches [76, 91–93] leverage growing availability of naturalistic driving datasets to learn behavioral models directly from observations. Inverse reinforcement learning infers reward functions that rationalize observed driving behaviors, enabling prediction of human responses in novel situations by optimizing the learned rewards. Recurrent neural networks and transformers capture temporal dependencies in driving sequences, learning to predict future trajectories from historical observations. Graph neural networks model interactions among multiple vehicles, capturing how human drivers adjust behaviors based on surrounding traffic configurations.

Despite impressive predictive performance on benchmark datasets, purely data-driven models face challenges in safety-critical deployment. Learned models may fail catastrophically on out-of-distribution inputs not represented in training data, producing predictions that violate physical constraints or lead to unsafe interactions. The black-box nature of neural network models limits interpretability and makes formal verification difficult. Training data biases may encode and perpetuate unsafe or inequitable behaviors observed in human driving populations.

Hybrid approaches combine physics-based structure with data-driven learning to balance interpretability and expressiveness. Physics-informed neural networks incorporate differential equation constraints into loss functions, ensuring learned models respect fundamental physical laws. Modular architectures separate perception, prediction, and decision-making, applying learning selectively to components where data-driven methods excel while retaining interpretable structure for safety-critical elements. Our framework adopts this hybrid philosophy, combining IDM-based nominal behaviors with learned uncertainty distributions that capture variability while maintaining interpretable structure.

2.4.2 Intent Recognition and Uncertainty Quantification

Beyond predicting trajectories, effective interaction with human drivers requires reasoning about their intentions and quantifying uncertainty in predictions [94]. Intent recognition infers high-level goals such as lane-change intentions, turn directions at intersections, or yielding decisions from observable vehicle states and motions. Bayesian inference frameworks update intent beliefs as observations accumulate, providing probabilistic assessments that explicitly represent uncertainty.

Hidden Markov models (HMMs) represent a classical approach to intent recognition, modeling observable vehicle motions as emissions from latent intent states [94]. Transition probabilities capture temporal dynamics of intent changes, while emission probabilities relate intents to observable motion patterns. Inference algorithms such as forward-backward passes or Viterbi decoding compute posterior intent distributions from observation sequences. However, HMM expressiveness is limited by discrete state spaces and Markovian assumptions that may not capture complex intent dynamics.

Goal-recognition planning formulates intent recognition as inverse planning, inferring which goals would rationalize observed actions as approximately optimal [85]. This approach leverages the principle that human behaviors, while not perfectly optimal, tend to be reasonable with respect to underlying objectives. By computing which goals would make observed actions sensible, goal-recognition methods can infer intentions from partial observation sequences. However, computational cost grows with the complexity of the planning problem, limiting real-time applicability in complex traffic scenarios.

Learning-based intent recognition employs supervised or self-supervised learning on labeled trajectory datasets [95]. Convolutional or recurrent architectures extract spatial-temporal features from observation sequences and classify into intent categories through softmax outputs interpreted as probability distributions. Attention mechanisms enable models to focus on relevant aspects of complex traffic scenes when inferring intentions. However, learned classifiers require extensive labeled training data and may not generalize reliably to novel scenarios.

Uncertainty quantification extends beyond intent recognition to characterize confidence in trajectory predictions and behavioral model parameters [96]. Ensemble methods train multiple predictors on bootstrap samples or random initializations, interpreting prediction variance across ensemble members as epistemic uncertainty. Bayesian neural networks place distributions over network weights, propagating parameter uncertainty through the network to produce predictive distributions. Monte Carlo dropout approximates Bayesian inference by sampling different dropout masks at test time, providing computationally efficient uncertainty estimates.

Conformal prediction provides distribution-free uncertainty quantification with finite-sample coverage guarantees [97]. By calibrating prediction intervals on held-out data, conformal methods construct sets of plausible predictions that contain true outcomes with specified probability regardless of the underlying data distribution. This approach offers formal guarantees suitable for safety-critical applications without requiring restrictive distributional assumptions.

Our framework incorporates intent recognition and uncertainty quantification through adaptive safety margins that expand or contract based on prediction confidence [98]. When human driver behaviors are highly uncertain—such as at intersection approaches where turn intentions remain ambiguous—safety margins increase to account for multiple possible responses. As additional observations resolve uncertainty, safety margins adapt accordingly, enabling efficient coordination when predictions are confident while maintaining robustness when uncertainty is high.

2.4.3 Cooperative and Competitive Interaction Modeling

Human driving involves both cooperative and competitive behaviors depending on context, urgency, and social norms. Cooperative behaviors include yielding to merging vehicles, maintaining consistent speeds to facilitate lane changes, and following traffic conventions such as zipper merging. Competitive behaviors include aggressive gap closures to prevent lane changes, acceleration to claim priority at uncontrolled intersections, and violation of formal priority rules when perceived safe.

Social force models capture cooperation through attractive forces toward destinations and repulsive forces from obstacles and other vehicles, with force magnitudes and directions encoding social norms such as maintaining personal space and avoiding sudden relative motions [99]. However, simple social force formulations struggle to represent strategic aspects of competitive interactions where drivers actively influence others' behaviors.

Game-theoretic frameworks naturally model the spectrum from cooperation to competition through payoff structures [100]. Cooperative games encode shared objectives such as collision avoidance and traffic efficiency, admitting Pareto-optimal equilibria where no agent can improve without harming others. Non-cooperative games represent conflicting objectives such as priority claims at intersections, with equilibria reflecting strategic compromises between competing interests [101]. Mixed-motive games combine cooperative and competitive elements, capturing realistic scenarios where drivers share safety objectives but compete for priority and efficiency.

However, human drivers do not explicitly solve game-theoretic problems in real-time. Descriptive models of strategic interaction require cognitively plausible representations of bounded rationality, limited lookahead, and heuristic reasoning [102]. Level-k models provide such representations through recursive reasoning hierarchies that match observed human strategic thinking better than perfect rationality assumptions. This cognitive plausibility motivates our framework's use of Level-k reasoning to model human strategic behaviors while maintaining computational tractability.

Cultural and contextual variations in cooperation-competition balance complicate modeling. Driving norms vary across regions, with some cultures favoring more cooperative yielding behaviors while others exhibit more competitive priority claims [103]. Time pressure and traffic density shift individual drivers toward more competitive behaviors as urgency increases or perceived opportunities decrease. AVs deployed globally must adapt to local norms or risk coordination failures and negative social perceptions.

Our framework addresses cooperation-competition modeling through multi-objective reward functions that explicitly balance safety, efficiency, comfort, and cooperative behaviors. By adjusting relative weights on these objectives, the framework can represent diverse driving styles ranging from highly cooperative (prioritizing others' efficiency) to more assertive (prioritizing own efficiency subject to safety constraints). This flexibility enables adaptation to different cultural contexts and individual human driver styles through online parameter estimation or offline calibration [104–106].

2.5 Safety-Critical Decision Making

Safety represents the paramount constraint in autonomous driving, requiring formal guarantees that go beyond probabilistic assurances or empirical testing. Safety-critical decision-making frameworks must ensure collision avoidance under worst-case conditions while maintaining real-time computational feasibility and preserving sufficient maneuverability for achieving complex mission objectives.

2.5.1 Collision Avoidance and Safety Constraints

Collision avoidance constitutes the most fundamental safety requirement in autonomous driving. Various mathematical frameworks formalize collision-free constraints and provide mechanisms for enforcing them during planning and control. The most basic approach represents vehicles as geometric shapes—typically rectangles or circles—and enforces non-overlap constraints between shapes. However, simple geometric constraints become overly conservative when applied rigidly, as they do not account for relative velocities, future trajectories, or maneuverability margins.

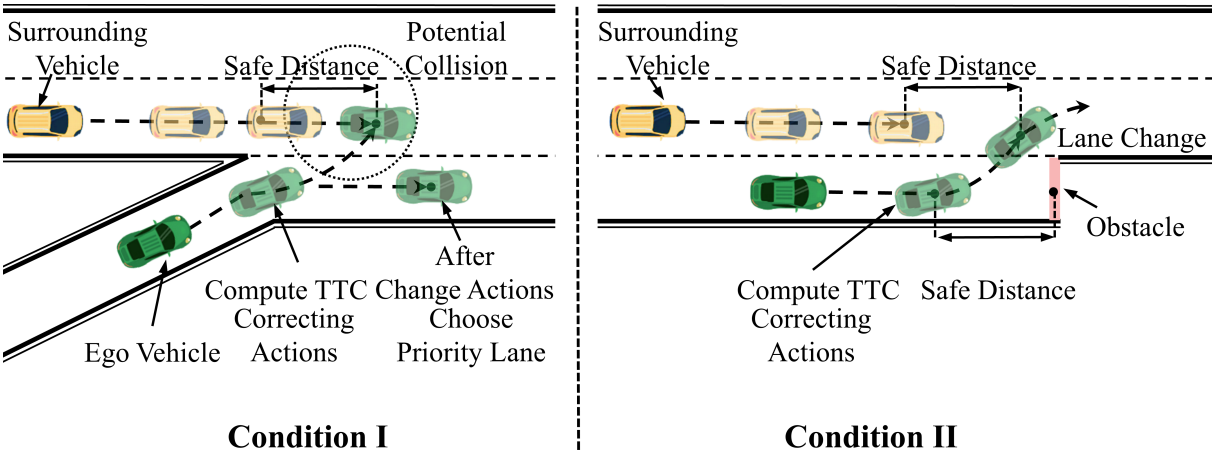


Figure 2.6: Examples of Initial lane-selection (Condition I) and in-ramp lane-selection (Condition II).

Time-to-collision (TTC) metrics refine geometric constraints by incorporating velocities and predicted trajectories as shown in Fig. 2.6. TTC measures the time until collision would occur if current velocities are maintained, providing an intuitive safety metric that accounts for both spacing and relative motion. Safety constraints then require TTC to exceed minimum thresholds that allow sufficient reaction time [107]. However, TTC becomes undefined when relative closing velocity is zero or negative, limiting applicability to approaching scenarios.

Time-to-closest-point-of-approach (TTCPA) generalizes TTC by computing when vehicles will be nearest to each other along current trajectories, even if actual collision does not occur. Combined with distance-at-closest-approach, TTCPA-based metrics provide continuous safety assessments valid across diverse interaction geometries. Adaptive thresholds that vary with velocity, uncertainty, and road conditions enable context-sensitive safety constraints that balance conservatism with efficiency [108].

Responsibility-sensitive safety proposes a formal framework for assigning responsibility in potential collision scenarios based on right-of-way rules and physical constraints [77]. Responsibility-Sensitive Safety (RSS) defines dangerous situations where collision cannot be avoided through any physically possible maneuver, and requires that AVs never cause dangerous situations for others while responding appropriately when others create dan-

gerous situations. By formalizing blame assignment and required responses, RSS aims to provide verifiable safety properties and legal clarity. However, RSS relies on conservative assumptions about others' behaviors that may produce excessively cautious driving in practice.

Reachability analysis computes sets of states reachable under all possible control inputs and disturbances, enabling formal verification that unsafe states remain unreachable. Forward reachable sets represent states the ego vehicle can reach, while backward reachable sets represent states from which collisions are inevitable [86]. Comparing these sets against obstacle predictions determines whether collision avoidance is guaranteed, possible, or impossible. Hamilton-Jacobi reachability provides computational tools for continuous-state systems, though scalability remains challenging for high-dimensional systems.

Our framework incorporates collision avoidance through multiple layers of defense. Conservative Level-0 baselines establish worst-case safety margins that guarantee collision-free trajectories under strong assumptions about others' behaviors. Higher reasoning levels preserve these margins through bounded optimization that explores efficiency improvements while maintaining safety buffers. Adaptive safety margins adjust dynamically based on prediction uncertainty, expanding when human behaviors are ambiguous and contracting when high-confidence predictions enable tighter coordination [103].

2.5.2 Control Barrier Functions and Formal Methods

Control Barrier Functions (CBFs) provide a unifying mathematical framework for enforcing safety constraints in continuous-time dynamical systems. A CBF is a scalar function that decreases along unsafe trajectories and increases along safe trajectories, with safety guaranteed when CBF values remain positive. By formulating safety constraints as CBF conditions, controllers can enforce collision avoidance through pointwise inequality constraints rather than global trajectory optimization [109, 110].

The integration of Control Lyapunov Functions (CLFs) with CBFs enables simultaneous optimization of stability and safety objectives through quadratic programming. At each time step, quadratic programming (QP) solvers find control inputs minimizing deviation from desired control while satisfying CBF safety constraints and CLF stability conditions. This real-time optimization framework provides formal safety guarantees without requiring full trajectory planning, making it suitable for fast control loops.

CBF-based methods [109,110] have been applied to various autonomous driving scenarios including adaptive cruise control, lane keeping, and intersection coordination. Extensions incorporate multiple CBFs for handling multiple obstacles, high-relative-degree CBFs for systems where safety constraints depend on higher derivatives, and time-varying CBFs for moving obstacles. However, CBF approaches face challenges in multi-agent settings where other agents' future behaviors must be predicted to evaluate barrier function evolution.

Recent work has explored learning-based methods for constructing CBFs from data, addressing the difficulty of manually designing barrier functions for complex systems. Neural network CBFs learn safe regions through supervised learning on safe-unsafe trajectory data, while reinforcement learning approaches discover CBFs that balance safety and performance through reward shaping. However, providing formal guarantees for learned CBFs remains challenging, as training data may not cover all possible scenarios and neural network verification remains computationally expensive [111, 112].

While CBF methods provide elegant formal guarantees for continuous control, they require accurate models of other agents' future behaviors to evaluate barrier function evolution. In multi-agent settings with strategic interactions, this requirement becomes problematic. Our framework instead achieves safety through structural properties of Level-k reasoning, where conservative Level-0 baselines provide safety guarantees without requiring explicit barrier function computation.

2.5.3 Risk Assessment and Emergency Maneuvers

Beyond nominal collision avoidance, AVs must recognize and respond to emergency situations where standard planning and control may be insufficient. Risk assessment frameworks quantify the severity and imminence of potential collisions, enabling appropriate escalation of responses from nominal planning to emergency maneuvers [113].

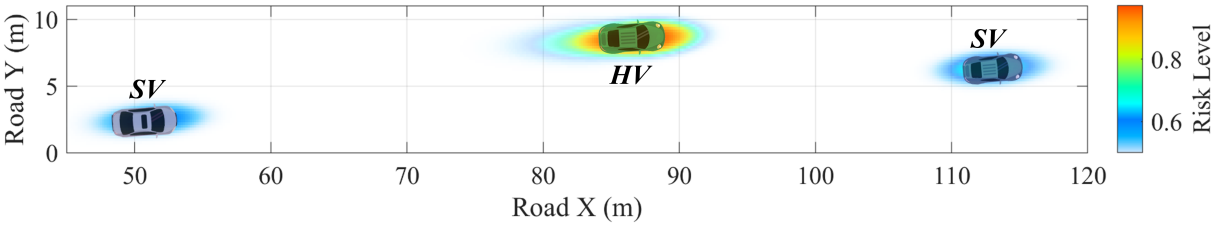


Figure 2.7: Illustration of the dynamic risk field. The host vehicle (HV) is positioned at the centre of the road segment, with two surrounding vehicles (SV) in adjacent lanes. Risk levels are computed using exponential barrier functions of inter-vehicle distance (Eq. (3.12)), visualised as colour-coded fields ranging from low risk (blue, ≤ 0.6) to high risk (red, ≥ 0.9). The asymmetric risk distribution around HV reflects its higher velocity relative to the surrounding vehicles.

Hierarchical risk assessment as shown in Fig. 2.7 evaluates multiple interaction types with different safety thresholds and response strategies. V2V interactions require coordination through strategic planning. Vehicle-to-human (V2H) interactions must account for behavioral uncertainty. Vehicle-to-road (V2R) interactions involve static obstacles and road boundaries that constrain feasible maneuvers. Our framework systematically addresses these three interaction types through specialized assessment modules with context-appropriate safety thresholds.

Dynamic risk fields represent collision risk as continuous spatial-temporal distributions rather than binary safe-unsafe classifications [114]. Risk magnitudes vary smoothly with proximity to obstacles, relative velocities, and prediction uncertainties. Planning through risk fields enables smooth adaptation of trajectories in response to changing risk landscapes, avoiding abrupt strategy switches that compromise passenger comfort. Gradient-based optimization in risk fields provides computationally efficient trajectory refinement.

Emergency maneuver planning activates when risk assessments exceed critical thresholds or when standard planning fails to find safe solutions. Emergency maneuvers prioritize collision avoidance over efficiency and comfort, employing maximum braking, aggressive steering, or combinations thereof to escape imminent collisions [115]. However, emergency maneuvers must account for vehicle dynamics limits, as excessive braking or steering can induce loss of control that transforms one hazard into another.

Fallback strategies provide final layers of defense when all planning and emergency responses fail. Minimum-risk conditions specify target states such as stopping in current lane or pulling to roadside that minimize expected harm. Transition planning computes safe trajectories to minimum-risk conditions when mission objectives become infeasible [116]. Our framework incorporates these concepts through safety-aware pruning that eliminates actions leading to states from which minimum-risk conditions cannot be reached.

2.6 Application Domains in Autonomous Driving

While the preceding sections reviewed methodological approaches, their effectiveness ultimately depends on successful application to concrete autonomous driving scenarios. We briefly survey key application domains to contextualize our work and establish the practical relevance of the challenges addressed.

2.6.1 Intersection Management Systems

Unsignalized intersection coordination exemplifies the core challenges of multi-agent planning in autonomous driving [117]. The fundamental difficulty arises from the need to resolve conflicts among vehicles with competing passage objectives within limited spatial and temporal margins. Traditional management approaches rely on reservation-based protocols, auction mechanisms, or priority rules, but these struggle in symmetric scenarios where no natural priority exists [118–120].

Early centralized intersection management systems demonstrated feasibility of V2I coordination for modest traffic volumes. Vehicles communicate arrival times and desired trajectories to intersection managers that grant or deny passage based on conflict detection. However, centralized approaches inherit limitations of single points of failure, communication requirements, and computational scalability [62]. As traffic density increases beyond four to six simultaneously interacting vehicles, centralized computation becomes prohibitive and communication latency undermines real-time coordination.

Learning-based intersection coordination has shown promise in adapting to diverse traffic patterns through reinforcement learning [121, 122]. Multi-agent Reinforcement Learning (RL) discovers emergent cooperation strategies without explicit communication or priority rules. However, learned policies often fail to generalize to novel configurations such as varying agent counts or geometric layouts not represented in training distributions. Retraining requirements for each deployment scenario limit practical applicability [63, 64].

Recent work has begun addressing larger-scale coordination involving six to eight agents, though often with simplified dynamics, perfect communication assumptions, or relaxed real-time constraints [52]. Prior methods have not jointly addressed eight-agent symmetric coordination under realistic dynamics, limited communication, and strict real-time requirements as tackled in our work. The symmetric eight-agent intersection scenario serves as our primary validation domain, representing one of the most challenging forms of decentralized coordination due to maximal symmetry and the total absence of any natural priority ordering.

2.6.2 Ramp Merging and Highway Scenarios

On-ramp merging presents distinct challenges from intersection coordination due to continuous traffic flow, high-speed dynamics, and the need to identify appropriate merge gaps [54, 120]. Successful merging requires coordination with mainline traffic to find or create acceptable gaps while avoiding both dynamic vehicles and static road boundaries at ramp terminations, as shown in Fig. 2.8.

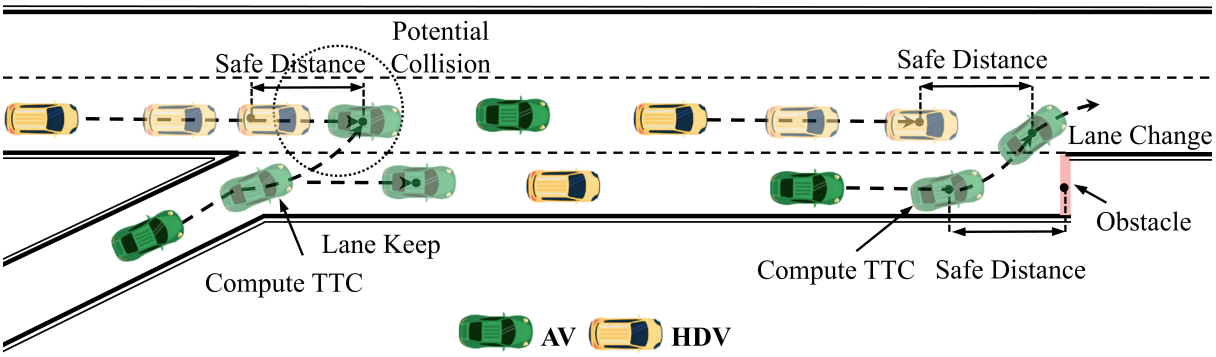


Figure 2.8: Illustration of safety-critical scenarios during ramp merging. The diagram shows key interactions between AVs and HDVs, including safe distance maintenance, collision detection, and lane-changing decisions.

Classical approaches separate longitudinal and lateral control, with gap acceptance models determining when to initiate lane changes and trajectory planners generating smooth merge paths. However, this decomposition can produce inconsistent behaviors when discrete mode transitions disrupt continuous control. More recent approaches employ integrated planning that jointly optimizes merge timing and trajectory shape, achieving smoother and more human-like merging behaviors.

MCTS has been applied to ramp merging scenarios to handle uncertainty in human driver responses during gap negotiations [23, 25]. By simulating possible human reactions during rollouts, MCTS enables AVs to evaluate risky versus conservative merge strategies under behavioral uncertainty. However, scalability remains limited to two to three interacting vehicles due to joint action space explosion.

CBFs provide formal safety guarantees for ramp merging by encoding collision avoidance constraints as pointwise inequalities [123]. CBF-based controllers ensure that lane changes maintain safe distances from mainline vehicles and road boundaries throughout maneuvers. Integration with CLFs enables optimization of efficiency and comfort objectives subject to safety constraints. While our primary focus is intersection coordination, the framework's safety mechanisms share conceptual similarities with CBF approaches in their fundamental design principles.

2.6.3 Roundabout Navigation

Roundabout navigation combines elements of both intersection coordination and continuous flow management. The circular geometry creates continuous conflict zones without discrete decision points, while entry and exit maneuvers require yield decisions and gap acceptance similar to intersection negotiation. Curved trajectories complicate perception and prediction due to sensor occlusions and reduced visibility, increasing behavioral uncertainty compared to straight intersections.

Existing approaches to roundabout navigation include game-theoretic methods for modeling strategic yield decisions, learning-based policies for discovering efficient circulation strategies, and potential field methods for reactive obstacle avoidance [120, 124]. However, each faces limitations: game-theoretic approaches struggle with computational efficiency in time-critical curved scenarios, learning methods lack generalization across varying roundabout geometries, and potential field methods suffer from local minima in circular topologies during complex maneuvers.

Recent work has explored uncertainty-aware prediction specifically for roundabout scenarios, recognizing that curved geometries and lane-specific dynamics increase prediction difficulty [57]. Probabilistic trajectory prediction methods quantify confidence in human driver behaviors, enabling risk-aware planning that adapts safety margins to prediction uncertainty. Integration with game-theoretic or MCTS planning remains an active research direction for achieving robust multi-agent coordination.

While roundabout scenarios as shown in Fig. 2.9 present interesting extensions of our framework's core ideas, we maintain primary focus on intersection coordination as the canonical multi-agent planning challenge. The methodological innovations developed for symmetric intersection scenarios—reconstructed Level-k reasoning, MCTS integration, dual-filtered interaction graphs—generalize to roundabouts and other traffic domains, though specific implementations may require domain-specific adaptations to geometric and dynamic constraints.

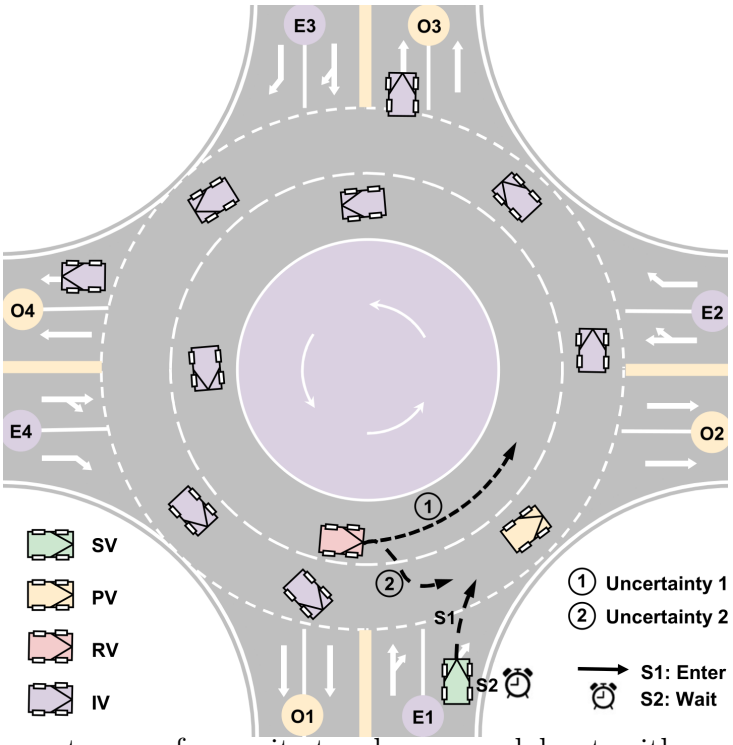


Figure 2.9: A four-entrance, four-exit, two-lane roundabout with an example collision scenario involving an AV. SV, PV, RV, and IV denote subject vehicle, preceding vehicle, rear vehicle, and irrelevant vehicle, respectively.

2.7 Chapter Summary and Research Gaps

This chapter has systematically reviewed the state-of-the-art in multi-agent coordination, strategic reasoning, and safety-critical decision-making for autonomous driving. Our analysis reveals several persistent challenges that motivate the comprehensive coordination framework developed and evaluated in this thesis.

2.7.1 Fundamental Limitations of Existing Paradigms

he preceding review has analysed each paradigm with respect to both its strengths and limitations. Existing approaches face a fundamental trilemma among computational tractability, strategic sophistication, and safety guarantees.

Existing approaches face a fundamental trilemma among computational tractability, strategic sophistication, and safety guarantees. Centralized optimization methods achieve near-optimal coordination but suffer from exponential complexity that renders them impractical beyond four to five agents. Game-theoretic approaches provide principled strategic reasoning but face intractable equilibrium computation and equilibrium selec-

tion problems in symmetric scenarios, while making idealized rationality assumptions that fail in mixed traffic with uncertain human behaviors and incomplete information. Learning-based methods demonstrate impressive performance in trained scenarios but require massive training data, lack generalization to novel configurations, and cannot provide formal safety guarantees required for real-world deployment.

Level-k reasoning offers a cognitively plausible alternative through bounded rationality modeling, but classical formulations suffer from a critical flaw: defining Level-0 as random or naive actors creates fundamentally unsafe baselines that propagate through the reasoning hierarchy. This limitation has prevented application of Level-k reasoning to safety-critical autonomous driving despite its computational advantages and behavioral realism. Similarly, MCTS provides efficient online planning through selective exploration without training requirements but faces exponential complexity ($O(|A|^N)$) in multi-agent settings due to joint action space growth and lacks explicit opponent modeling, treating other agents as environmental noise.

The safety assessment literature has made significant progress in modeling collision avoidance constraints through various mathematical frameworks including TTC metrics, responsibility-sensitive safety formulations, and CBFs. However, existing methods often treat safety as an external constraint imposed on top of strategic planning, rather than as an emergent property of the decision-making structure itself. Furthermore, mixed traffic scenarios introduce additional complexity through human behavioral uncertainty that most frameworks handle through overly conservative assumptions or probabilistic extensions that lack principled integration with strategic reasoning during multi-agent interactions.

Beyond these fundamental gaps, practical deployment considerations remain underexplored. Most existing work evaluates methods in scenarios involving two to four agents with simplified dynamics or communication assumptions. Scaling to eight agents under realistic constraints including limited communication, continuous dynamics, and strict

real-time requirements represents a significant challenge that prior work has not comprehensively addressed. Furthermore, the interpretability and verifiability of decision-making processes become increasingly critical as autonomous systems transition from testing to deployment, yet many high-performing learning-based approaches remain black boxes.

Table 1.1 summarizes the key limitations of representative approaches and how our framework addresses these challenges across four major paradigms: rule-based methods that lack strategic reasoning, learning-based approaches with weak safety guarantees, game-theoretic methods with idealized assumptions, and standard MCTS with exponential complexity in multi-agent settings.

2.7.2 Positioning of This Thesis

Three critical research gaps emerge from this analysis that this thesis addresses through a unified framework combining reconstructed Level-k reasoning with MCTS for multi-agent coordination.

Reformulating Level-k Reasoning for Safety-Critical Planning. We fundamentally redefine Level-0 from a behavioral model of naive actors to a conservative safety initialization procedure that generates collision-free baseline trajectories. This reconceptualization transforms safety from an external constraint into a structural property that propagates through the reasoning hierarchy: Level-1 agents optimizing against safety-aware Level-0 baselines inherit the spatial separation margins embedded in those baselines, and Level-2 agents similarly benefit from the conservative Level-1 responses they model. While this propagation does not constitute a formal guarantee at higher levels due to modeling errors under simplified opponent predictions, it provides systematic safety margins under the adopted modeling assumptions, ensuring improved collision-free properties compared to standard Level-k formulations with arbitrary Level-0 behaviors. This addresses the fundamental safety limitations that have prevented Level-k adoption in autonomous driving while preserving its computational advantages and behavioral realism.

Synergistic Integration of Level-k Reasoning and MCTS. We propose a novel integration architecture where Level-k reasoning decomposes the exponential joint action space into sequential reasoning steps, while MCTS provides adaptive sampling and uncertainty handling. No existing framework successfully integrates these two paradigms in a manner that leverages their complementary strengths while mitigating individual weaknesses. Our integration incorporates carefully designed computational mechanisms including dual-filtered interaction graphs for scalability, safety-aware pruning for collision avoidance, and trajectory caching for efficiency. This synergy reduces computational complexity from exponential $\mathcal{O}(|\mathcal{A}|^N)$ to linear $\mathcal{O}(N \cdot |\mathcal{A}|)$ in the number of agents (see Chapter 4.5.4 and Table 4.5 for detailed analysis) while maintaining strategic sophistication comparable to game-theoretic approaches, particularly effective in symmetric multi-agent coordination problems where traditional approaches fail.

Unified Heterogeneous Modeling for Mixed Traffic Environments. We develop a comprehensive framework that seamlessly integrates homogeneous multi-agent coordination with heterogeneous mixed traffic scenarios within a common theoretical foundation. Most existing approaches specialize in either pure AV coordination or human-vehicle interaction, failing to provide unified architectures that handle both cases. Our framework combines deterministic AV models with probabilistic human behavior predictions through adaptive safety margins that dynamically adjust to estimated human uncertainty. This integration addresses the limitations of game-theoretic methods that assume perfect rationality, enabling robust collision avoidance even under behavioral deviations while balancing safety, efficiency, comfort, and cooperative behavior through multi-objective optimization.

The subsequent chapters address these identified gaps systematically. Chapter 3 establishes the foundational MCTS framework and identifies its scalability limitations in multi-agent settings. Chapter 4 resolves these limitations through reconstructed Level-k reasoning and dual-filtered interaction graphs. Chapter 5 extends the framework to mixed traffic environments with probabilistic human modeling. Chapter 6 concludes with a summary of findings and future research directions.

Chapter 3

Multi-Agent Coordination via Monte Carlo Tree Search

This chapter establishes the foundational framework for multi-agent coordination at unsignalized intersections using Monte Carlo Tree Search (MCTS). We begin by formalizing the symmetric intersection coordination problem that motivates our approach, then develop comprehensive agent models encompassing vehicle dynamics, collision detection, and trajectory representation. The formulation progresses systematically from individual agent modeling to multi-objective optimization, culminating in the MCTS-based planning algorithm that enables real-time decision-making. Through simulation-based evaluation on a four-agent left-turn scenario, we demonstrate that MCTS provides an effective solution for small-scale coordination problems while identifying scalability limitations that motivate the extensions developed in subsequent chapters.

3.1 Problem Statement

We consider a decentralized multi-agent coordination problem where N autonomous vehicles simultaneously approach an unsignalized intersection from multiple directions. Each direction provides two lanes, and vehicles are initially positioned equidistant from the intersection center, creating a symmetric configuration as illustrated in Fig. 3.1. This setup represents a fundamental challenge in multi-agent coordination, as the absence of traffic signals or predetermined priority rules forces all vehicles to resolve conflicts through strategic planning rather than external regulation.

The symmetric configuration introduces three fundamental challenges that distinguish this problem from conventional traffic coordination:

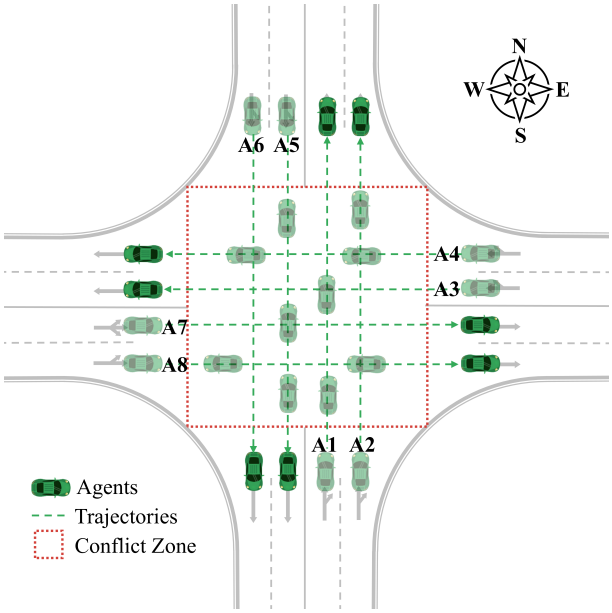


Figure 3.1: Symmetric eight-agent intersection scenario. Agents approach from orthogonal directions at equal distances, executing left-turn maneuvers that create crossing conflicts.

Challenge 1: Simultaneous Arrival. Unlike scenarios where vehicles arrive with temporal separation allowing simple priority assignment based on first-come-first-served rules, our symmetric configuration forces all agents to confront direct strategic conflicts at the intersection center simultaneously. This eliminates the sequential queueing structure present in typical traffic flow and requires genuine multi-agent negotiation.

Challenge 2: Exponential Action Space Growth. The joint action space grows exponentially with both agent count N and planning horizon H. For a single decision step, the joint action space scales as $\mathcal{O}(|\mathcal{A}|^N)$, where $\mathcal{A}$ denotes the discrete action space available to each agent; over a full planning horizon of H steps, this compounds to $\mathcal{O}(|\mathcal{A}|^{N \cdot H})$. For the four-agent scenario with $|\mathcal{A}| = 15$ actions and $H = 9$ steps, Eq. (3.27) yields approximately 10^{42} joint trajectories, rendering exhaustive enumeration computationally intractable.

Challenge 3: Deadlock Risk. Perfect symmetry dramatically increases deadlock risk, as no agent possesses inherent spatial or temporal advantage to break ties. Without principled coordination mechanisms, agents may either yield indefinitely in mutual deference or proceed simultaneously into collision.

To address these challenges, we adopt a decentralized framework where each agent $i \in \{1,\ldots,N\}$ plans locally based on observations of nearby agents. The planning objective for each agent can be abstractly formulated as finding an optimal policy π_i^* that minimizes expected cumulative cost over a finite receding horizon H:

$$\begin{aligned} \pi_i^* &= \arg\min_{\pi_i \in \Pi_i} \mathbb{E}\left[\sum_{t=0}^{H-1} c_i(S^t, a_i^t)\right] \\ \text{s.t.} \quad & s_i^t \in \mathscr{S}_{\text{safe}}^i, \quad \forall t \in [0, H-1], \\ & a_i^t \in \mathscr{A}_i, \quad \forall t \in [0, H-1], \end{aligned} \tag{3.1}$$

where $\pi_i : \mathscr{S}_i \to \mathscr{A}_i$ maps local states to actions, Π_i denotes the space of feasible policies, $c_i(\cdot)$ represents the instantaneous cost function encoding safety, efficiency, and comfort objectives, S^t denotes the global state comprising all agents' states at time t, and the constraints ensure trajectory safety through state feasibility $\mathscr{S}_{\text{safe}}^i$ over the planning horizon. In practice, since each agent only has access to local observations, we replace the global state S^t with the local observation S_i^t, which comprises the agent's own state s_i^t together with the states of its observable neighbors.

The following sections develop the detailed formulation, from agent models to the MCTS planning algorithm designed for multi-agent coordination.

3.2 Agent Modeling

This section establishes the mathematical representation of individual autonomous vehicles, including state and action spaces, kinematic dynamics, and collision detection mechanisms utilized in the planning process.

3.2.1 State and Action Spaces

The state of agent i at discrete time step t is represented by a four-dimensional vector:

$$s_i^t = [x_i^t, y_i^t, v_i^t, \theta_i^t]^\top \in \mathscr{S}_i \subset \mathbb{R}^4, \tag{3.2}$$

where x_i^t and y_i^t denote the Cartesian position coordinates in a global reference frame centered at the intersection, $v_i^t \in [0, v_{\max}]$ represents the longitudinal velocity constrained by maximum velocity $v_{\max}$, and $\theta_i^t \in [-\pi, \pi]$ specifies the heading angle measured counterclockwise from the positive x-axis. The global state $S^t = \{s_1^t, \ldots, s_N^t\} \in \mathscr{S}^N$ aggregates individual agent states into a joint representation of the multi-agent system at time t.

Each agent selects actions from a discrete action space $\mathscr{A}_i$ with 15 control primitives. Each primitive combines longitudinal acceleration $a_{i,\text{lon}}^t$ and angular velocity ω_i^t:

$$a_i^t = [a_{i,\text{lon}}^t, \omega_i^t]^\top \in \mathscr{A}_i, \tag{3.3}$$

The complete action space is specified in Table 3.1. The primitives are organised into four categories: pure longitudinal control (maintain, braking, acceleration), pure lateral control (steering at two intensities), and combined longitudinal-lateral actions (acceleration or braking with steering). This discretisation balances trajectory expressiveness with computational tractability, providing sufficient coverage of feasible manoeuvres at unsignalised intersections.

Table 3.1: Complete discrete action space $\mathscr{A}_i$ with 15 control primitives

Index	Primitive	a_{lon} (m/s^2)	ω (rad/s)
1	Maintain	0.0	0.0
2	Low brake	-1.5	0.0
3	Mid brake	-3.5	0.0
4	High brake	-5.0	0.0
5	Low acceleration	1.5	0.0
6	Mid acceleration	2.5	0.0
7	High acceleration	4.5	0.0
8	Low left steer	0.0	$\pi/4$
9	Low right steer	0.0	$-\pi/4$
10	Mid left steer	0.0	$\pi/2$
11	Mid right steer	0.0	$-\pi/2$
12	Accelerate + left steer	1.5	$\pi/4$
13	Accelerate + right steer	1.5	$-\pi/4$
14	Brake + left steer	-1.5	$\pi/4$
15	Brake + right steer	-1.5	$-\pi/4$

3.2.2 Vehicle Dynamics

Agent motion evolves according to a kinematic bicycle model that captures the essential characteristics of vehicle dynamics while maintaining efficiency for real-time planning:

$$x_i^{t+1} = x_i^t + v_i^t \cos(\theta_i^t)\Delta t, \tag{3.4a}$$

$$y_i^{t+1} = y_i^t + v_i^t \sin(\theta_i^t)\Delta t, \tag{3.4b}$$

$$v_i^{t+1} = \text{clip}(v_i^t + a_{i,\text{lon}}^t \Delta t, 0, v_{\max}), \tag{3.4c}$$

$$\theta_i^{t+1} = \theta_i^t + \omega_i^t \Delta t, \tag{3.4d}$$

where $\Delta t = 0.2s$ denotes the discrete time step and the $\text{clip}(\cdot)$ function enforces velocity bounds. This formulation assumes instantaneous response to control inputs, which is reasonable for planning at the time scales typical of intersection coordination where vehicle actuator dynamics can be neglected compared to the strategic planning horizon of the coordination process.

3.2.3 Collision Detection

Collision detection forms a critical component of safety assessment in multi-agent planning. We employ the Separating Axis Theorem (SAT) to perform precise geometric collision checking between oriented rectangular vehicle representations. Each agent i is modeled as an oriented rectangle with length l_i and width w_i. Given an agent state s_i^t, the four vertices of the occupied rectangle are computed as:

$$\mathbf{v}_i^m = \begin{bmatrix} x_i^t \\ y_i^t \end{bmatrix} + \mathbf{R}(\theta_i^t) \cdot \mathbf{p}^m, \quad m \in \{1,2,3,4\}, \tag{3.5}$$

where $\mathbf{R}(\theta_i^t)$ is the two-dimensional rotation matrix:

$$\mathbf{R}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \tag{3.6}$$

and $\mathbf{p}^m$ represents the corner offset vector from the agent center to vertex m expressed in the vehicle's body-fixed frame. Specifically, $\mathbf{p}^m$ is a constant vector determined by the vehicle dimensions l_i and w_i:

$$\mathbf{p}^1 = \begin{bmatrix} l_i/2 \\ w_i/2 \end{bmatrix}, \quad \mathbf{p}^2 = \begin{bmatrix} -l_i/2 \\ w_i/2 \end{bmatrix}, \quad \mathbf{p}^3 = \begin{bmatrix} -l_i/2 \\ -w_i/2 \end{bmatrix}, \quad \mathbf{p}^4 = \begin{bmatrix} l_i/2 \\ -w_i/2 \end{bmatrix}. \tag{3.7}$$

Potential collisions between agents at states s_i and s_j are determined by applying the Separating Axis Theorem:

$$\text{Collision}(s_i, s_j) = \neg\exists \mathbf{n} : \text{proj}_{\mathbf{n}}(\mathbf{V}_i(s_i)) \cap \text{proj}_{\mathbf{n}}(\mathbf{V}_j(s_j)) = \emptyset, \tag{3.8}$$

where $\mathbf{V}_i = \{\mathbf{v}_i^1, \mathbf{v}_i^2, \mathbf{v}_i^3, \mathbf{v}_i^4\}$ denotes the vertex set, $\mathbf{n}$ represents potential separating axes (edge normals), and $\text{proj}_{\mathbf{n}}(\cdot)$ denotes projection onto axis $\mathbf{n}$. If no separating axis exists, the polygons overlap, indicating a collision.

Beyond inter-agent collision avoidance, each agent must satisfy road boundary constraints. The feasible state space is defined as:

$$\mathscr{S}_{\text{safe}}^i = \{s_i \in \mathscr{S}_i : \mathscr{V}_i(s_i) \subset \mathscr{R}_{\text{valid}} \wedge v_i \in [0, v_{\max}]\}, \tag{3.9}$$

where $\mathscr{V}_i(s_i)$ denotes the geometric polygon occupied by agent i at state s_i, and $\mathscr{R}_{\text{valid}}$ represents the drivable region bounded by road geometry.

3.2.4 Trajectory Representation in Frenet Coordinates

To facilitate efficient trajectory planning, we represent agent trajectories using a Frenet coordinate system aligned with reference paths. For each agent i, a reference path Γ_i is defined through the intersection according to the agent's intended maneuver (straight, left turn, or right turn). The vehicle's position relative to this reference path is parameterized using arc length s along the path and lateral offset d perpendicular to the path.

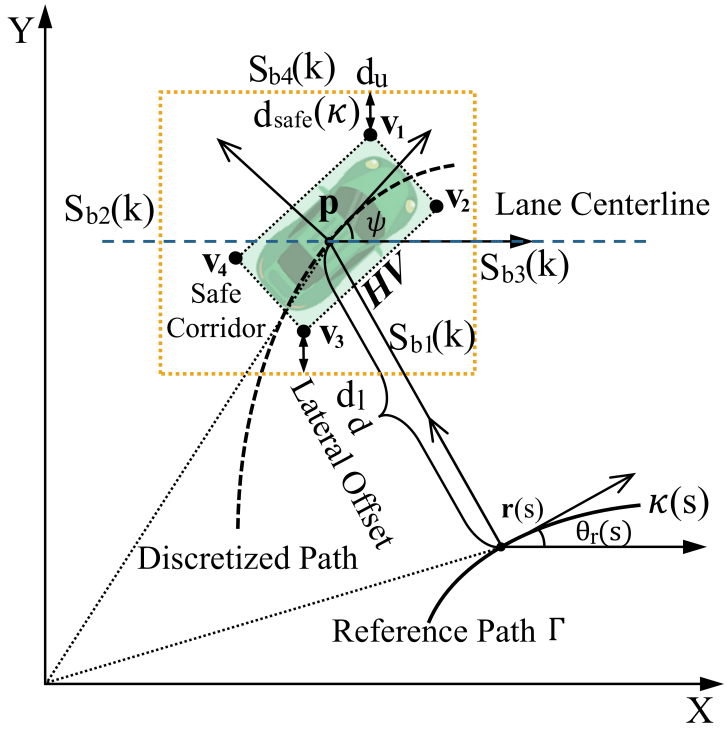


Figure 3.2: Frenet coordinate representation. The reference path Γ defines the nominal trajectory with arc-length parameter s, curvature $\kappa(s)$, and tangent angle $\theta_r(s)$. The vehicle position $\mathbf{p}$ is described by arc length s and lateral offset d, with heading angle ψ relative to the lane centerline. The safe corridor at discrete time step k is bounded by four boundary segments $S_{b1}(k)$–$S_{b4}(k)$, where d_u and d_l denote the upper and lower lateral safety margins, and $d_{\text{safe}}(\kappa)$ is the curvature-dependent safety distance. Vertices $\mathbf{v}_1$–$\mathbf{v}_4$ define the corridor polygon enclosing the host vehicle (HV).

The transformation between Cartesian coordinates (x,y) and Frenet coordinates (s,d) is given by:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{r}(s) + d \cdot \begin{bmatrix} -\sin(\theta_r(s)) \\ \cos(\theta_r(s)) \end{bmatrix}, \tag{3.10}$$

where $\mathbf{r}(s)$ represents the reference path centerline position and $\theta_r(s)$ is the path tangent angle at arc length s, as illustrated in Fig. 3.2. This representation simplifies trajectory optimization by decoupling longitudinal progress from lateral positioning.

3.3 Multi-Objective Optimization Formulation

Having established individual agent models, we now formulate the decision-making problem as a multi-objective optimization that balances competing considerations of safety, efficiency, comfort, and path adherence.

3.3.1 Cost Function Design

Each agent seeks to minimize a cumulative cost functional over the planning horizon while satisfying hard constraints on collision avoidance and kinematic feasibility. The instantaneous cost function aggregates multiple objectives through weighted summation:

$$c_i(S_i^t, a_i^t) = w_{\mathrm{s}} \cdot c_{\mathrm{s}}^i(S_i^t) + w_{\mathrm{e}} \cdot c_{\mathrm{e}}^i(s_i^t) + w_{\mathrm{d}} \cdot c_{\mathrm{d}}^i(s_i^t) + w_{\mathrm{c}} \cdot c_{\mathrm{c}}^i(a_i^t), \tag{3.11}$$

where the weights $w_{\mathrm{s}}, w_{\mathrm{e}}, w_{\mathrm{d}}, w_{\mathrm{c}}$ control the relative importance of safety, efficiency, trajectory deviation, and comfort objectives, respectively. Here, S_i^t denotes the local observation of agent i, comprising its own state s_i^t and the states of its observable neighbors, and the cost components c_{s}^i, c_{e}^i, c_{d}^i, and c_{c}^i correspond to the safety, efficiency, path adherence, and comfort terms, which are specified in the following subsections. This multi-objective design reflects the principle that autonomous driving must balance multiple stakeholder concerns: regulators mandate safety, passengers expect efficient travel, and vehicle dynamics impose comfort constraints. The weights follow a deliberate priority ordering $w_{\mathrm{s}} \gg w_{\mathrm{d}} > w_{\mathrm{c}} > w_{\mathrm{e}}$, reflecting the hierarchical importance of these objectives in safety-critical driving scenarios. Specifically, the safety weight w_{s} is set approximately one order of magnitude larger than the remaining weights to ensure that collision avoidance dominates all other objectives. The path adherence weight w_{d} is set larger than comfort w_{c} because lane-keeping is a hard operational requirement, whereas comfort represents a soft preference. Efficiency w_{e} receives the smallest weight as speed regulation is naturally bounded by the safety and path-following constraints. These relative magnitudes were determined empirically through systematic parameter sweeps on the four-agent left-turn scenario, and held fixed across all subsequent simulations to ensure consistency. Specific values for each reasoning level are provided in Section 3.3.

3.3.1.1 Safety Cost

The safety cost penalizes proximity to both other agents and static road boundaries through exponential barrier functions:

$$c_{\mathrm{s}}^{i}(S_i^t) = \sum_{j \neq i} \exp\left(-\frac{d_{ij}^2}{2\sigma_{\mathrm{safe}}^2}\right) + \sum_{b \in \mathscr{B}_{\mathrm{road}}} \exp\left(-\frac{d_{ib}^2}{2\sigma_{\mathrm{safe}}^2}\right), \tag{3.12}$$

where d_{ij} is the minimum distance between agents i and j, $\mathscr{B}_{\mathrm{road}}$ denotes the set of road boundary elements, d_{ib} is the distance from agent i to boundary element b, and σ_{safe} is the safety threshold parameter (set to $\sigma_{\mathrm{safe}} = 2\,\mathrm{m}$ in all simulations). The exponential form ensures costs rise sharply as spacing approaches the safety threshold, providing strong gradient signals that guide trajectories away from hazardous configurations.

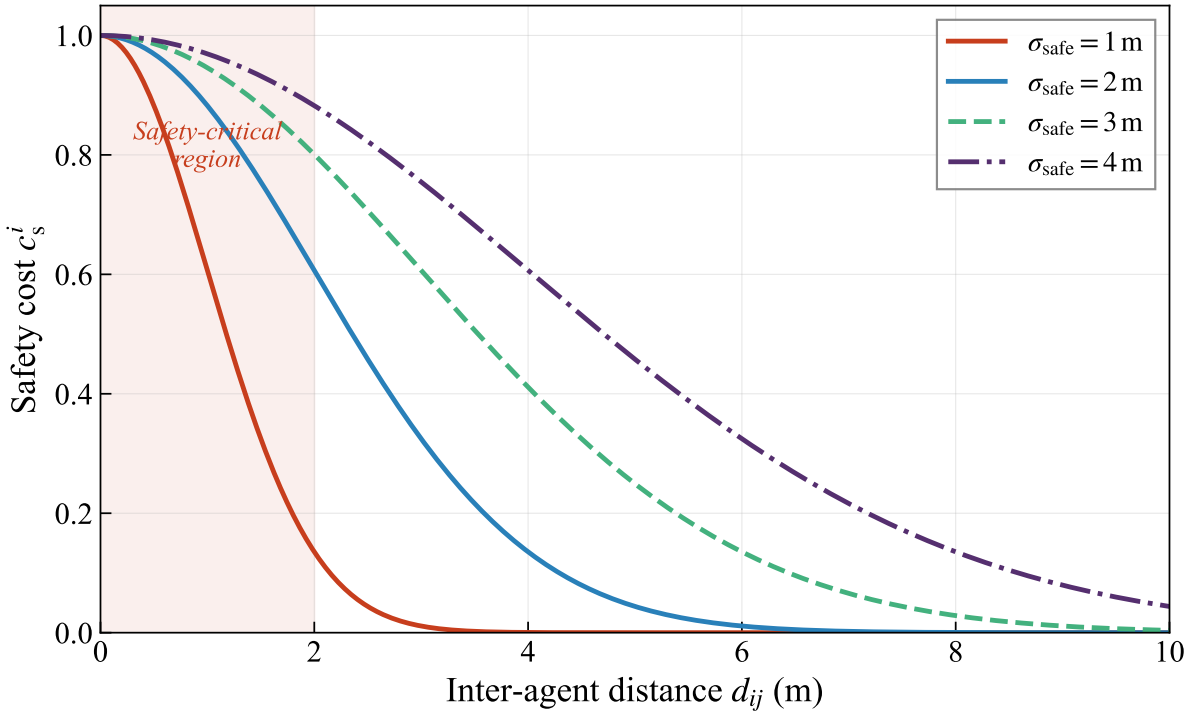


Figure 3.3: Safety cost c_{s}^{i} vs. inter-agent distance d_{ij} for selected σ_{safe}.

The exponential barrier function is selected for three reasons. First, it is continuously differentiable, providing smooth gradient signals throughout the state space that facilitate effective tree search in MCTS. Second, it exhibits asymmetric sensitivity: costs remain negligible at large inter-agent distances but increase sharply as $d_{ij} \to 0$, naturally encoding the non-linear urgency of proximity in safety-critical scenarios. Third, the parameter σ_{safe}

provides an intuitive and geometrically meaningful tuning knob that directly controls the effective safety margin radius. Figure 3.3 illustrates the profile of c_{s}^{i} as a function of d_{ij} for representative values of σ_{safe}, confirming that the cost rises steeply within the safety-critical range and decays gracefully at larger distances.

3.3.1.2 Efficiency Cost

The efficiency cost encourages progress toward the agent's goal by penalizing deviations from a reference velocity:

$$c_{\mathrm{e}}^{i}(s_i^t) = |v_{\mathrm{ref}} - v_i^t|, \tag{3.13}$$

where v_{ref} represents the desired cruising speed. This linear penalty encourages maintaining the reference velocity while allowing the overall cost function to trade off speed against safety when necessary.

3.3.1.3 Trajectory Deviation Cost

The deviation cost penalizes departure from the reference path, ensuring agents remain within designated lanes:

$$c_{\mathrm{d}}^{i}(s_i^t) = \min_{p \in \mathscr{P}_i} \|[x_i^t, y_i^t]^\top - p\|_2, \tag{3.14}$$

where $\mathscr{P}_i$ denotes the set of reference path waypoints for agent i. The minimum distance formulation provides smooth cost gradients with respect to the agent's position (x_i^t, y_i^t) even when the agent's position does not align with discrete waypoints.

3.3.1.4 Comfort Cost

The comfort cost discourages abrupt control changes that cause uncomfortable motion:

$$c_{\mathrm{c}}^{i}(a_i^k) = \begin{cases} \|a_i^k - a_i^{k-1}\|_2^2, & k > 0 \\ 0, & k = 0 \end{cases}, \tag{3.15}$$

where k denotes the discrete planning step index to distinguish from the continuous time variable t, and the quadratic form penalizes large changes in control inputs. Smooth control trajectories enhance passenger comfort and reduce mechanical wear. The comfort cost considers longitudinal acceleration changes only, rather than lateral acceleration. This is a deliberate design choice motivated by the symmetric eight-agent intersection scenario: agents follow designated reference paths whose lateral deviations are already penalized by the trajectory deviation cost c_{d}^{i}. Imposing additional penalties on lateral acceleration would over-constrain agent maneuverability, potentially increasing collision risk by preventing necessary evasive adjustments. Safety is instead enforced as the primary objective through the dominant weight w_{s} in Eq. (3.11).

3.3.2 Markov Decision Process Formulation

The multi-agent coordination problem can be formulated as a Markov Decision Process (MDP) from each agent's perspective. For agent i, the MDP is defined as:

$$\mathscr{M}_i = (\mathscr{S}_i, \mathscr{A}_i, P_i, r_i, H), \tag{3.16}$$

where $\mathscr{S}_i$ is the local state space, $\mathscr{A}_i$ is the discrete action space, $P_i : \mathscr{S}_i \times \mathscr{A}_i \to \mathscr{S}_i$ is the deterministic transition function derived from the dynamics model (3.4), r_i is the reward function defined as the negative cost, and H is the planning horizon.

The reward function aligns with standard MDP conventions where agents maximize expected cumulative reward:

$$r_i(S_i^t, a_i^t) = -c_i(S_i^t, a_i^t), \tag{3.17}$$

enabling application of standard planning algorithms to find optimal policies.

The objective is to find an optimal policy π_i^* maximizing expected cumulative reward:

$$\pi_i^* = \arg\max_{\pi_i} \mathbb{E}\left[\sum_{t=0}^{H-1} r_i(S_i^t, a_i^t) \mid a_i^t = \pi_i(s_i^t)\right]. \tag{3.18}$$

However, directly solving this optimization faces computational challenges due to the coupling between agents through the safety cost component $c_{\mathrm{s}}^{i}(S_i^t)$, which depends on all agents' positions. In principle, optimal coordination requires joint optimization over all agents' policies, but the exponential growth of the joint action space renders such approaches intractable. The following section introduces Monte Carlo Tree Search as a practical solution enabling real-time planning through selective exploration.

3.4 MCTS Planning Algorithm

MCTS provides a principled approach to sequential decision-making that achieves near-optimal performance without exhaustive enumeration of the action space. This section describes the MCTS algorithm adapted for complex multi-agent intersection coordination scenarios.

3.4.1 Algorithm Overview

MCTS builds an asymmetric search tree through iterative simulation, progressively refining action value estimates and expanding the tree toward promising regions of the action space. Each iteration consists of four phases: selection, expansion, simulation (rollout), and backpropagation. The algorithm operates in a receding horizon fashion, replanning at each time step based on updated observations.

3.4.2 Tree Structure and Node Representation

The search tree $\mathscr{T}_i$ for agent i consists of nodes representing states encountered during exploration. Each node $n \in \mathscr{T}_i$ maintains the following components:

$$n = \{s(n), a(n), d(n), Q(n), N(n), \mathscr{C}(n)\}, \tag{3.19}$$

where $s(n) \in \mathscr{S}_i$ is the state reached by executing action $a(n) \in \mathscr{A}_i$ from the parent node, $d(n) \in \{0, \ldots, H\}$ is the depth from the root, $Q(n) \in \mathbb{R}$ is the cumulative value estimate, $N(n) \in \mathbb{N}$ is the visit count, and $\mathscr{C}(n)$ is the set of child nodes, as illustrated in Fig. 3.4.

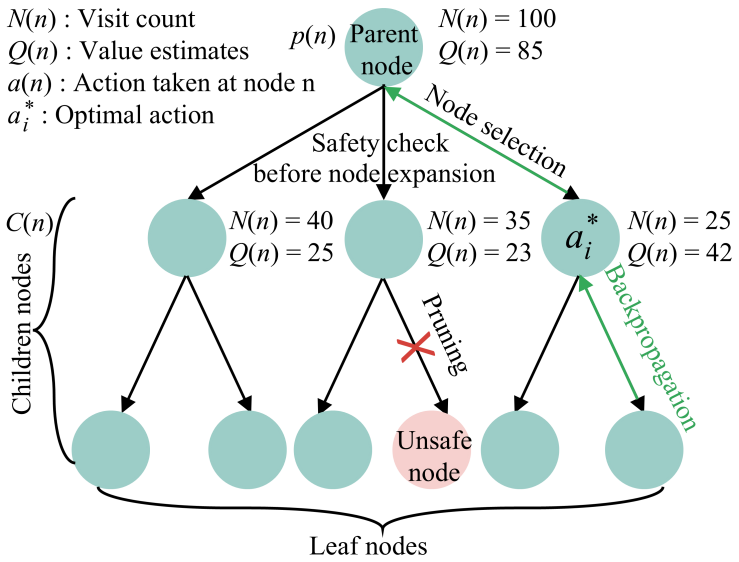


Figure 3.4: MCTS tree structure illustration. Each node maintains visit count $N(n)$, cumulative value $Q(n)$, and associated action $a(n)$. The tree grows asymmetrically through selective expansion guided by the UCT criterion.

3.4.3 Selection via Upper Confidence Bounds

Each MCTS iteration begins with a selection phase that traverses the tree from root to a leaf node. At each internal node, the algorithm selects actions according to the Upper Confidence Bound applied to Trees (UCT) criterion:

$$a^* = \arg\max_{a \in \mathscr{A}_{\text{valid}}(n)} \left\{ \frac{Q(n,a)}{N(n,a)} + c\sqrt{\frac{\ln N(n)}{N(n,a)}} \right\}, \tag{3.20}$$

where $\mathscr{A}_{\text{valid}}(n) \subseteq \mathscr{A}_i$ contains actions satisfying kinematic and road boundary constraints, $Q(n,a)$ is the cumulative value for action a from node n, $N(n,a)$ is the visit count for that action, $N(n) = \sum_{a'} N(n,a')$ is the total visits to node n, and $c > 0$ is a constant.

The UCT formula balances exploitation (first term: empirical mean reward) against exploration (second term: uncertainty bonus for less-visited actions). The logarithmic dependence ensures exploration bonus decreases over time, allowing convergence toward exploitation of superior actions.

3.4.4 Expansion with Safety Validation

Upon reaching a leaf node n_{leaf}, the expansion phase adds child nodes representing previously unexplored actions. Before expansion, actions are validated for safety:

$$\mathscr{A}_{\text{safe}}(n) = \{a \in \mathscr{A}_i : \Phi_{\text{safe}}(s(n), a) = \text{true}\}, \tag{3.21}$$

where the safety predicate Φ_{safe} verifies:

1. *Kinematic feasibility*: The resulting state satisfies dynamics constraints (3.4).
2. *Road boundary compliance*: The vehicle remains within $\mathscr{R}_{\text{valid}}$.
3. *Collision avoidance*: No collision with predicted positions of other agents.

For collision checking, other agents' future positions are predicted using a simplified constant velocity model throughout the search:

$$\hat{s}_j^{t+\tau} = \begin{bmatrix} x_j^t + v_j^t \cos(\theta_j^t)\tau \\ y_j^t + v_j^t \sin(\theta_j^t)\tau \\ v_j^t \\ \theta_j^t \end{bmatrix}, \quad j \neq i, \tag{3.22}$$

Here, τ denotes the prediction time offset from the current time t, which is conceptually distinct from the single-step time increment Δt in Eq. (3.4). While Δt is a fixed simulation clock step used to iteratively propagate the ego agent's state forward one step at a time, τ parameterises a direct extrapolation from the current moment to an arbitrary future instant, i.e., $\tau \in \{\Delta t, 2\Delta t, \ldots, H\Delta t\}$. Two distinct variables are therefore introduced to reflect these different computational roles: iterative state propagation for the ego agent versus direct constant-velocity extrapolation for neighbouring agents.

This simplifying assumption enables efficient prediction but may produce optimistic collision assessments when other agents accelerate or change direction. We note that this chapter presents a decentralized baseline where each agent runs an independent MCTS planner with constant-velocity opponent predictions, rather than a joint multi-agent optimizer. This intentional simplification isolates the MCTS planning backbone and is addressed in Chapter 4 through Level-k strategic opponent modeling.

A new child node n_{new} is created with:

$$s(n_{\text{new}}) = f(s(n_{\text{leaf}}), a_{\text{new}}), \tag{3.23}$$

where $f(\cdot)$ denotes the state transition according to dynamics (3.4).

3.4.5 Rollout Simulation

After expansion, the algorithm evaluates the new node through rollout simulation. Starting from $s(n_{\text{new}})$, actions are sampled according to a default policy π_{default} until reaching the horizon depth H:

$$V_{\text{rollout}}(n_{\text{new}}) = \sum_{k=d(n_{\text{new}})}^{H-1} r_i(S_i^k, a_i^k), \tag{3.24}$$

where $a_i^k \sim \pi_{\text{default}}(s_i^k)$, and $S_i^k = \{s_i^k, \hat{s}_1^k, \ldots, \hat{s}_{i-1}^k, \hat{s}_{i+1}^k, \ldots, \hat{s}_N^k\}$ combines the ego agent's simulated state with predicted states $\hat{s}_j^k$ of other agents via (3.22).

The default policy typically employs simple heuristics such as constant velocity continuation or reference path following, providing baseline estimates of trajectory values without expensive optimization.

3.4.6 Backpropagation

The rollout value is backpropagated from the expanded node to the root, updating statistics along the path:

$$Q(n) \leftarrow Q(n) + V_{\text{rollout}}, \tag{3.25a}$$

$$N(n) \leftarrow N(n) + 1, \tag{3.25b}$$

for each node n on the path from n_{new} to the root n_0. The empirical mean $Q(n)/N(n)$ in the UCT criterion thus represents average observed return across all rollouts passing through node n.

Algorithm 1: MCTS Planning for Multi-Agent Coordination

Input: Current state s_i^t, other agents' states $\{s_j^t\}_{j\neq i}$, horizon H, iterations K
Output: Optimal action a_i^*
Initialize root node n_0 with $s(n_0) = s_i^t$
for each of K iterations **do**
  // **Selection**
  $n \leftarrow n_0$
  while n is not a leaf node **do**
    $a \leftarrow \arg\max_{a'} \mathrm{UCT}(n, a')$ {Eq. (3.20)}
    $n \leftarrow \mathrm{child}(n, a)$
  end while
  // **Expansion**
  if $d(n) < H$ **then**
    $\mathscr{A}_{\mathrm{safe}} \leftarrow \mathrm{SafetyValidation}(s(n))$ {Eq. (3.21)}
    Select $a_{\mathrm{new}} \in \mathscr{A}_{\mathrm{safe}}$ not yet expanded
    Create child n_{new} with $s(n_{\mathrm{new}}) = f(s(n), a_{\mathrm{new}})$
    $n \leftarrow n_{\mathrm{new}}$
  end if
  // **Rollout**
  $V \leftarrow \mathrm{Simulate}(s(n), H - d(n))$ {Eq. (3.24)}
  // **Backpropagation**
  while $n \neq \mathrm{null}$ **do**
    $Q(n) \leftarrow Q(n) + V; \quad N(n) \leftarrow N(n) + 1$
    $n \leftarrow \mathrm{parent}(n)$
  end while
end for
return $\arg\max_{a \in \mathscr{C}(n_0)} Q(n_0, a)/N(n_0, a)$

3.4.7 Action Selection

After completing K MCTS iterations, the optimal action is extracted:

$$a_i^* = \arg\max_{a \in \mathscr{C}(n_0)} \frac{Q(n_0, a)}{N(n_0, a)}, \tag{3.26}$$

selecting the action with highest empirical mean value from the root node.

The complete MCTS planning procedure is summarized in Algorithm 1.

3.5 Experimental Validation

This section validates the effectiveness of the MCTS planning framework through comprehensive experiments on a four-agent left-turn scenario. We evaluate performance across multiple metrics including safety, efficiency, and computational cost, comparing different planning horizon configurations to demonstrate the importance of sufficient lookahead in multi-agent coordination.

3.5.1 Experimental Setup

3.5.1.1 Scenario Description

The experimental scenario involves four autonomous vehicles approaching a symmetric unsignalized intersection from orthogonal directions, as illustrated in Fig. 3.5. Each vehicle starts at a distance of $18\,\mathrm{m}$ from the intersection center and executes a left-turn maneuver. This configuration creates dense crossing conflicts at the intersection center, requiring sophisticated coordination to avoid collisions while maintaining traffic efficiency.

The scenario parameters are summarized in Table 3.2.

Table 3.2: Scenario and Algorithm Parameters

Parameter	Symbol	Value
Number of agents	N	4
Initial distance to center	d_0	$18\,\mathrm{m}$
Maximum velocity	$v_{\max}$	$10\,\mathrm{m/s}$
Reference velocity	v_{ref}	$7\,\mathrm{m/s}$
Time step	Δt	$0.2\,\mathrm{s}$
Action space size	$\lvert\mathscr{A}\rvert$	15
Vehicle length	l	$4.5\,\mathrm{m}$
Vehicle width	w	$2.4\,\mathrm{m}$
Safety threshold	σ_{safe}	$2.0\,\mathrm{m}$

3.5.1.2 Baseline Methods

To provide comprehensive comparison, we evaluate our MCTS framework against several established multi-agent coordination approaches:

- **Stackelberg Game** [33]: A hierarchical game-theoretic approach where agents are assigned leader-follower roles, with followers optimizing their responses given the leader's committed strategy.
- **Nash Equilibrium** [19]: A simultaneous game approach seeking strategy profiles where no agent can unilaterally improve its outcome, representing classical game-theoretic coordination.
- **Vanilla MCTS** [125]: A standard single-agent Monte Carlo Tree Search formulation that treats other agents as part of the environment under a constant-velocity assumption, without explicit opponent modelling or Level-k strategic reasoning. This baseline isolates the contribution of our Level-k integration by sharing the same tree search backbone while omitting the cognitive hierarchy.

All baseline methods use 1000 maximum iterations. Our MCTS framework uses $K = 300$ iterations per planning step with two horizon configurations: $H = 4$ (short) and $H = 9$ (long). The lower iteration count for MCTS reflects its per-step replanning nature, whereas baseline methods solve for complete trajectories. The horizon values $H = 4$ and $H = 9$ were selected to represent short-term reactive planning and long-term anticipatory planning respectively, with $H = 9$ providing sufficient lookahead to cover the full intersection traversal in our simulations. The iteration count $K = 300$ was determined empirically as the minimum value achieving stable Q-value estimates within the sub-100 ms real-time budget per planning step.

3.5.1.3 Evaluation Metrics

We employ the following metrics to quantify coordination performance:

- **Collision Rate (%)**: Percentage of trials resulting in inter-agent collisions, directly measuring safety performance.
- **Arrival Time (s)**: Average time for all agents to clear the intersection, measuring coordination efficiency.

- **Computation Time (ms)**: Per-step planning time, measuring real-time feasibility.
- **Trajectory Deviation (m)**: Mean deviation from the reference path, measuring path-following quality. This metric is related to, but distinct from, the lateral offset d in Fig. 3.2. The lateral offset d is an instantaneous geometric state variable defined in the Frenet frame, representing the perpendicular distance to the reference path at a single time step. In contrast, trajectory deviation is an aggregated evaluation metric computed as the mean of $|d_i^t|$ over all agents and time steps during execution. Thus, d serves as the per-step geometric quantity from which trajectory deviation is derived.
- **Minimum Distance Distribution**: Distribution of closest inter-agent distances throughout execution, providing detailed safety analysis.

All experiments are conducted over 40 independent trials with randomized initial position perturbations sampled from a uniform distribution $\mathscr{U}(-0.05, 0.05)$ m applied independently to each agent's longitudinal starting position along the reference path, to assess robustness.

3.5.2 Qualitative Analysis

Figure 3.5 provides a qualitative comparison of agent coordination between our proposed MCTS-based framework (top row) and the vanilla MCTS baseline (bottom row) in a symmetric four-agent unsignalized intersection scenario.

At $t = 2\,\mathrm{s}$, our method already demonstrates anticipatory and smooth coordination. Agents A3 and A4 maintain moderate speeds ($5.6\,\mathrm{m/s}$ and $7.0\,\mathrm{m/s}$) while preparing for their left turns, whereas agents A1 and A2 adjust to $4.8\,\mathrm{m/s}$ and $4.6\,\mathrm{m/s}$ to implicitly establish a safe passing order. This early speed modulation reflects effective conflict-aware reasoning, allowing agents to negotiate priority without abrupt braking or hesitation.

In contrast, the vanilla MCTS baseline exhibits noticeably less coherent behavior. Agents show larger velocity variance (4.9, 3.5, 6.3, and $5.6\,\mathrm{m/s}$ for A1–A4), indicating reactive adjustments rather than proactive coordination. Such inconsistent motion suggests that without our structured reasoning mechanism, agents struggle to form a stable interaction pattern under symmetry.

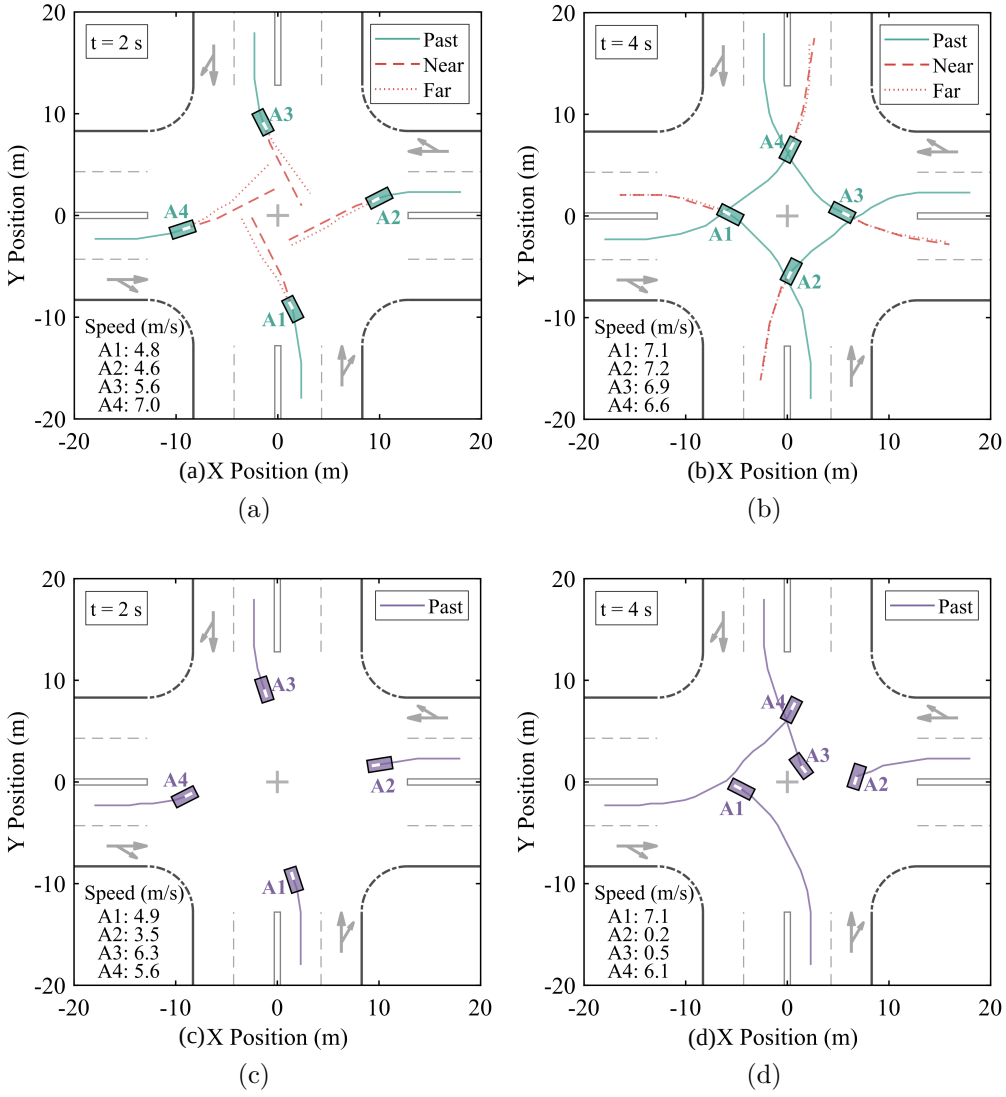


Figure 3.5: Comparison of agent coordination. Top row (a-b): Our method shows smooth coordination with minimal trajectory deviation. Bottom row (c-d): Vanilla MCTS exhibits larger deviations and longer delays.

By $t = 4\,\mathrm{s}$, the advantages of our approach become more pronounced. The proposed planner enables agents A1 and A2 to clear the intersection smoothly at $7.1\,\mathrm{m/s}$ and $7.2\,\mathrm{m/s}$, while agents A3 and A4 enter in a synchronized manner at $6.9\,\mathrm{m/s}$ and $6.6\,\mathrm{m/s}$. This coordinated flow demonstrates efficient implicit turn-taking with minimal trajectory deviation.

Meanwhile, the baseline suffers from significant delays and coordination breakdown. Agents A2 and A3 nearly come to a stop ($0.2\,\mathrm{m/s}$ and $0.5\,\mathrm{m/s}$), while A1 proceeds at $7.1\,\mathrm{m/s}$, creating unsafe velocity differentials and increasing the risk of deadlock. These abrupt slowdowns and deviations highlight the inability of vanilla MCTS to consistently resolve symmetric conflicts.

Overall, this qualitative comparison confirms that our framework achieves smoother, more stable multi-agent coordination, whereas the vanilla baseline often leads to reactive behavior, excessive deviation, and inefficient intersection traversal.

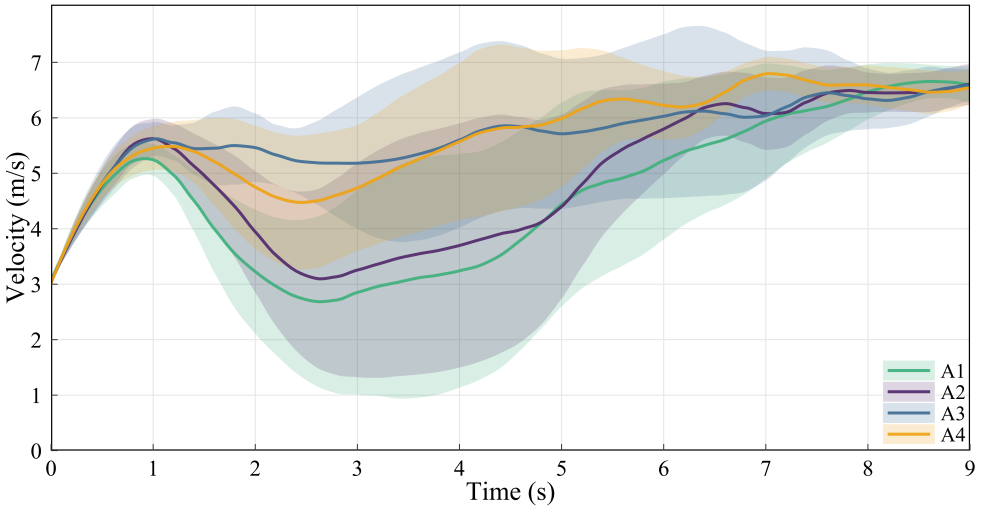
3.5.3 Temporal Dynamics Analysis

Figure 3.6 presents detailed temporal analysis of the long-horizon MCTS planner ($H = 9$), revealing the coordination patterns that emerge through the planning process.

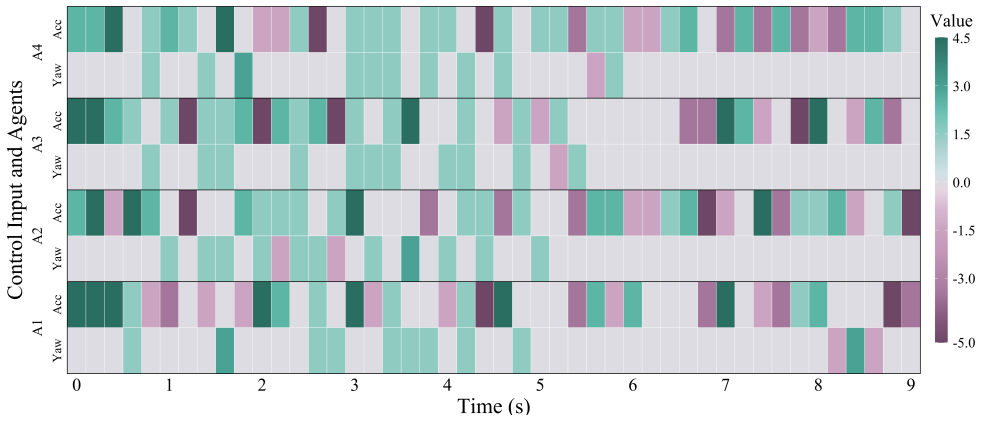
The velocity profiles in Fig. 3.6a show distinct coordination phases. During $t = 1$–$3\,\mathrm{s}$, agents A1 and A2 exhibit synchronized deceleration, reaching velocity minima around 3–$4\,\mathrm{m/s}$ to yield priority to agents A3 and A4. This implicit yielding behavior breaks the inherent symmetry of the scenario without explicit communication. Meanwhile, agents A3 and A4 maintain relatively stable speeds around 5–$6\,\mathrm{m/s}$, implicitly claiming earlier passage through the intersection.

After $t = 4\,\mathrm{s}$, all agents converge toward uniform velocities around 6–$7\,\mathrm{m/s}$ as they complete their left turns and exit the conflict zone. The shaded regions represent 95% confidence intervals across 40 trials, with narrow bands particularly for A3 and A4 indicating consistent strategic behavior enabled by the MCTS framework.

The control heatmap in Fig. 3.6b visualizes decision-making intensity throughout the scenario, where color intensity encodes control magnitude (dark purple indicating strong deceleration around $-5\,\mathrm{m/s^2}$, cyan indicating acceleration). Three notable patterns emerge:



(a) Velocity profiles of all agents over time



(b) Control input heatmap for all agents over time

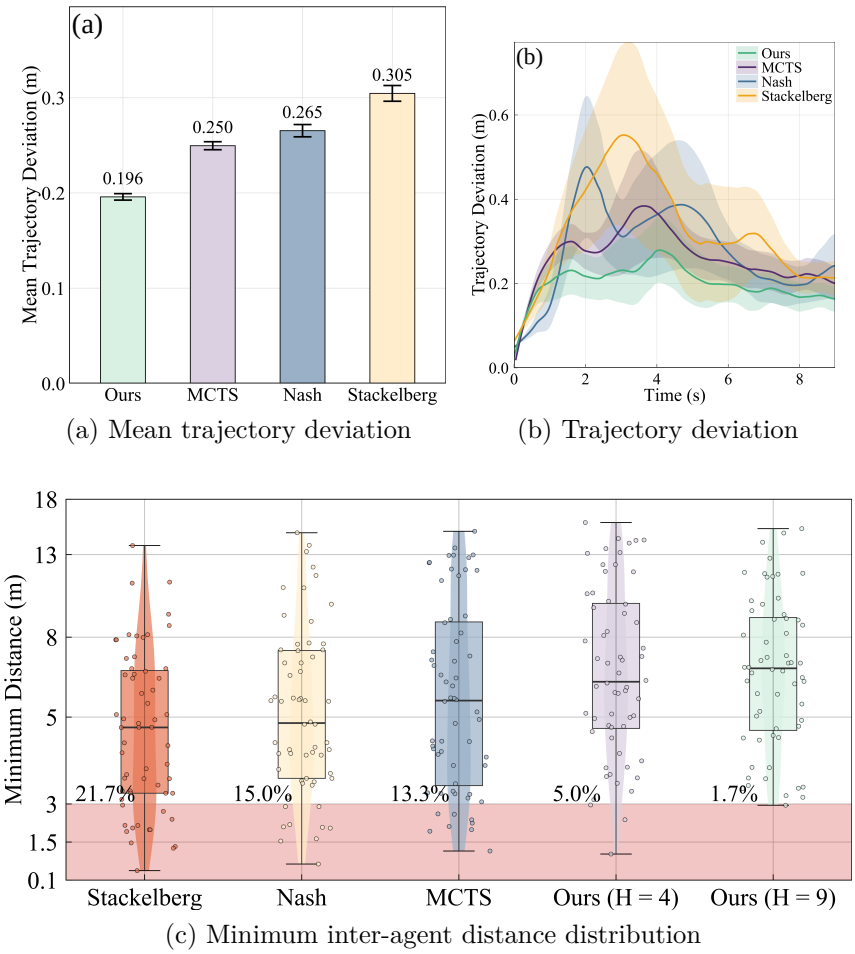
Figure 3.6: Temporal dynamics of MCTS ($H = 9$). (a) Velocity profiles with 95% confidence intervals; all agents share identical initial speeds, with position-only perturbations $\mathscr{U}(-0.05, 0.05)$ m. (b) Control input heatmap: Acc (m/s^2) and Yaw (rad/s) for all agents; darker/lighter regions denote deceleration/acceleration.

1. *Coordinated deceleration*: Intensive braking by A1 and A2 during $t = 1$–3 s (dark bands) establishes the yielding behavior that breaks symmetry. Notably, agent A3 also exhibits significant deceleration during this phase, followed by subsequent acceleration to reclaim its target speed. This suggests that A3 engages in anticipatory speed regulation rather than purely maintaining priority: even as a non-yielding agent, A3 temporarily adjusts its longitudinal profile to ensure safe spatial separation at the conflict point before accelerating to complete its left turn efficiently.

2. *Implicit turn-taking*: The heatmap reveals that agent A3 exhibits the most frequent and intense deceleration events, with approximately four distinct braking episodes concentrated around $t = 1\,\mathrm{s}$, $2\,\mathrm{s}$, and $3\,\mathrm{s}$. Rather than indicating coordination failure, this pattern reflects A3's active role in dynamically adjusting its passage timing to maintain safe inter-agent separation while navigating the most geometrically constrained conflict zone. The alternating control intensities across agents — with A1 and A2 exhibiting sustained braking while A3 applies repeated short bursts of deceleration and A4 maintains comparatively smoother control — collectively demonstrate an emergent negotiated passage order achieved without explicit inter-agent communication.
3. *Smooth transitions*: Gradual color changes in the acceleration channel indicate smooth longitudinal control evolution, validating the comfort cost component from Equation (3.15). Approximately 90% of longitudinal control inputs satisfy $|a| \leq 3\,\mathrm{m/s^2}$, confirming that coordination is achieved through strategic anticipation rather than aggressive reactive maneuvering. The yaw rate channel further corroborates this smoothness: gradual transitions in yaw rate across all agents indicate that lateral dynamics remain well-regulated throughout the left-turn maneuver, with no abrupt steering reversals observed. This simultaneous smoothness in both longitudinal and lateral channels confirms that the comfort cost effectively suppresses jerky control in both dimensions, producing naturalistic and passenger-friendly motion profiles.

It is worth noting that even if all agents were initialised with identical speeds, the proposed framework would still achieve safe and efficient coordination. The safety validation predicate Φ_{safe} in Eq. (3.21) eliminates unsafe actions during tree expansion at every planning step, ensuring that no collision trajectory can be selected regardless of the degree of initial symmetry. Furthermore, the receding horizon replanning strategy re-evaluates the action space at each time step, so any residual symmetry is progressively broken through the stochastic nature of MCTS rollouts combined with minor numerical differences in agent positions accumulated over time. Consequently, the randomized position perturbations in our experiments serve to assess robustness rather than to artificially break symmetry as a prerequisite for safety.

3.5.4 Statistical Performance Comparison



(a) Mean trajectory deviation (b) Trajectory deviation

(c) Minimum inter-agent distance distribution

Figure 3.7: Statistical analysis across methods in the four-agent scenario. (a) Mean trajectory deviation comparison. (b) Temporal evolution of trajectory deviation during the conflict phase. (c) Distribution of minimum inter-agent distances, where the red shaded region indicates

the safety threshold $d_{\text{safe}} = 3\,\text{m}$.

Figure 3.7 provides statistical comparison across all evaluated methods.

3.5.4.1 Trajectory Deviation Analysis

The mean trajectory deviation results in Fig. 3.7a demonstrate clear performance stratification. Our MCTS framework achieves the lowest mean deviation at $0.196\,\text{m}$, representing improvements of 22% over vanilla MCTS ($0.250\,\text{m}$), 26% over Nash equilibrium ($0.265\,\text{m}$), and 36% over Stackelberg game ($0.305\,\text{m}$).

The Stackelberg approach shows the worst performance, attributable to its rigid leader-follower assumption that breaks down in symmetric scenarios where no natural hierarchy exists. Nash equilibrium performs moderately but struggles with the coordination problem due to multiple equilibria in symmetric games, leading to inconsistent strategy selection.

The temporal evolution in Fig. 3.7b reveals performance differences during the critical conflict phase ($t = 2$–$5\,\mathrm{s}$). Our long-horizon MCTS (green curve) maintains deviations consistently below $0.25\,\mathrm{m}$, demonstrating stable coordination throughout. In contrast, Nash equilibrium (blue) and Stackelberg (orange) exhibit peaks up to $0.55\,\mathrm{m}$ around $t = 3$–$4\,\mathrm{s}$, indicating coordination breakdown during the most critical phase when multiple agents simultaneously occupy the intersection. The vanilla MCTS (red) shows intermediate performance with peaks around $0.38\,\mathrm{m}$. After $t = 8\,\mathrm{s}$, trajectory deviations decrease across all methods as agents successfully clear the intersection and return to their reference paths, reflecting the natural reduction in conflict once agents exit the high-density interaction zone.

3.5.4.2 Safety Analysis via Minimum Distance Distribution

The minimum distance distribution in Fig. 3.7c provides detailed safety assessment. We define *near-collision events* as instances where minimum inter-agent distance falls below $d_{\text{safe}} = 3\,\mathrm{m}$, distinct from actual collisions reported in Table 3.3.

The minimum distance distribution in Fig. 3.7c provides detailed safety assessment. We define the safety threshold $d_{\text{safe}} = 3\,\mathrm{m}$ as the minimum acceptable inter-agent distance; instances where separation falls below this threshold are classified as *near-collision events*, distinct from actual collisions (physical overlap). This threshold exceeds the cost function parameter $\sigma_{\text{safe}} = 2\,\mathrm{m}$ to provide a conservative safety margin for analysis.

Our MCTS framework with $H = 9$ achieves the best safety profile with only 1.7% near-collision events (instances where minimum distance falls below d_{safe}) and a median separation of $7\,\mathrm{m}$. This conservative behavior emerges naturally from sufficient planning horizon enabling anticipatory collision avoidance.

The vanilla MCTS exhibits degraded safety, with 5.0% near-collision events, approximately three times higher than our method. This degradation reflects its reactive interaction handling: without explicit coordination mechanisms, conflicts are often resolved only at the last moment, leading to abrupt avoidance maneuvers and increased risk.

Baseline methods exhibit significantly worse safety performance: Stackelberg game shows 21.7% near-collision events due to coordination failures when the leader-follower hierarchy conflicts with scenario geometry; Nash equilibrium reaches 15.0% due to equilibrium selection issues in symmetric scenarios; the vanilla MCTS achieves 13.3%, outperforming game-theoretic baselines but significantly trailing ours.

3.5.5 Quantitative Performance Summary

Table 3.3: Performance Comparison in Four-Agent Left-Turn Scenario

Method	Collision Rate (%)	Arrival Time (s)	Computation Time (ms)	Max. Iters
Stackelberg	17.5	9.7 ± 2.2	54.4 ± 14.1	1000
Nash	12.5	9.1 ± 1.7	74.6 ± 12.2	1000
MCTS	7.5	8.5 ± 2.6	60.4 ± 17.5	300
Ours ($H = 9$)	**0.0**	$\mathbf{5.3 \pm 1.3}$	$\mathbf{21.2 \pm 6.3}$	**300**

Table 3.3 summarizes the quantitative performance of all methods in the four-agent left-turn scenario. Overall, our proposed framework achieves consistent superiority across safety, efficiency, and computational cost.

Safety: Our method achieves perfect safety with a 0.0% collision rate, significantly outperforming Stackelberg (17.5%), Nash (12.5%), and vanilla MCTS (7.5%). This result demonstrates that our structured planning mechanism can reliably resolve symmetric interaction conflicts without collision.

Efficiency: In terms of traversal efficiency, agents complete the maneuver in $\mathbf{5.3 \pm 1.3}$ seconds, compared to 8.5–9.7 seconds for baseline methods. This corresponds to an improvement of approximately 38%–45%, indicating that our approach avoids the hesitation and delays commonly observed in game-theoretic and unstructured search-based coordination.

Computational Cost: Remarkably, our method also achieves the lowest computation time per planning step ($21.2 \pm 6.3\,\text{ms}$), representing a 65%–72% reduction relative to Stackelberg, Nash, and vanilla MCTS baselines. Notably, this efficiency is achieved with only 300 MCTS iterations, compared to 1000 iterations required by the game-theoretic solvers, highlighting the advantage of selective and structured exploration.

While these results confirm the effectiveness of our framework in four-agent coordination, the computational burden of joint reasoning still grows rapidly with the number of agents. As discussed in Section 3.1, scaling from four to eight agents expands the joint action space from approximately 10^{42} to 10^{85} (computed as $|\mathscr{A}|^{N\times H} = 15^{4\times 9} \approx 10^{42}$ and $15^{8\times 9} \approx 10^{85}$ respectively), making direct planning increasingly intractable.Chapter 4 addresses this scalability challenge through Level-k reasoning, which decomposes the joint optimization into a sequence of tractable single-agent subproblems.

3.6 Chapter Summary and Discussion

This chapter has established the foundational MCTS framework for multi-agent coordination at unsignalized intersections, demonstrating its preliminary effectiveness in a four-agent symmetric left-turn scenario through simulation-based evaluation under simplified and controlled conditions.

3.6.1 Summary of Contributions

We developed a complete planning framework comprising:

1. *Agent Modeling*: A comprehensive representation of autonomous vehicles including four-dimensional state space (3.2), discrete action space (3.3), kinematic bicycle dynamics (3.4), and precise collision detection via the Separating Axis Theorem (3.8).
2. *Multi-Objective Optimization*: A cost function (3.11) balancing safety (3.12), efficiency (3.13), path adherence (3.14), and comfort (3.15), cast as a Markov Decision Process amenable to tree search methods.

3. *MCTS Planning Algorithm*: A complete implementation (Algorithm 1) featuring UCT-based selection (3.20), safety-validated expansion (3.21), rollout simulation (3.24), and value backpropagation (3.25).

No collisions were observed in the four-agent left-turn simulation scenario, with additional improvements including:

- 40% reduction in arrival time compared to game-theoretic baselines within the evaluated scenario
- 65–71% reduction in computation time
- Consistent coordination behavior with narrow confidence intervals across trials

These results establish MCTS as an effective approach for small-scale multi-agent coordination, providing the baseline upon which subsequent chapters build.

3.6.2 Limitations and Scalability Challenges

While the MCTS framework demonstrates strong performance in the four-agent scenario, fundamental scalability limitations emerge when considering larger agent populations. These limitations motivate the extensions developed in subsequent chapters.

3.6.2.1 Exponential Action Space Growth

The joint action space grows exponentially with both agent count N and planning horizon H. For the four-agent scenario with $H = 9$ steps and $|\mathscr{A}| = 15$ actions, the theoretical joint action space contains:

$$|\mathscr{A}|^{N \cdot H} = 15^{4 \times 9} = 15^{36} \approx 10^{42} \text{ joint trajectories.} \tag{3.27}$$

When scaling to eight agents—a configuration representing realistic intersection density with vehicles approaching from multiple lanes in each direction—Eq. (3.27) gives

$$|\mathscr{A}|^{N \cdot H} = 15^{8 \times 9} = 15^{72} \approx 10^{85} \text{ joint trajectories.} \tag{3.28}$$

This 10^{43}-fold increase in search space size represents a fundamental barrier: even with MCTS's selective exploration, the probability of discovering coordinated strategies through random sampling diminishes exponentially as the problem scale increases.

3.6.2.2 Computational Time Scaling

We evaluate the empirical computation time as the number of agents increases from $N = 2$ to $N = 8$. The results exhibit a clear exponential scaling trend, consistent with the theoretical complexity of vanilla MCTS-based joint planning.

In particular, scenarios with up to four agents can be solved within a real-time budget of 100 ms per planning step. However, when the agent count increases to eight, the computation time exceeds 500 ms, violating real-time constraints. This rapid growth renders direct joint MCTS impractical for dense traffic coordination and highlights the necessity of structured decomposition or approximation mechanisms to achieve scalability.

This scaling behavior stems from two compounding factors:

1. *Collision checking overhead*: Each node expansion requires collision validation against all $N-1$ other agents' predicted trajectories, with cost scaling as $\mathcal{O}(N)$ per expansion.
2. *Increased tree depth requirement*: Larger agent populations create more complex interaction patterns requiring deeper search to discover coordinated strategies, but deeper search exponentially increases the number of nodes explored.

3.6.2.3 Coordination Quality Degradation

Beyond computational constraints, the quality of discovered strategies degrades as agent count increases. With $K = 300$ MCTS iterations distributed across an exponentially larger search space, the probability of each promising trajectory receiving sufficient exploration diminishes. This manifests as:

- Increased collision rates due to insufficient exploration of safe trajectories
- Higher trajectory deviation as agents fail to discover smooth coordination patterns
- Greater variance across trials reflecting inconsistent strategy discovery

Preliminary experiments with eight agents show collision rates exceeding 25% even with increased iteration budgets, demonstrating that brute-force scaling of vanilla MCTS cannot address the fundamental complexity barrier.

3.6.3 Motivation for Subsequent Chapters

The scalability limitations identified above motivate two key extensions developed and evaluated in the following chapters of this thesis:

Chapter 4: Interaction Graph and Level-k Reasoning. To address the exponential action space growth, we introduce a dynamic interaction graph that filters spatially and strategically irrelevant agents from each vehicle's planning problem. Combined with Level-k cognitive hierarchy reasoning, this decomposition reduces the effective problem size from joint optimization over all agents to sequential single-agent planning with bounded opponent modeling. The resulting MCTS-Level-k framework achieves linear computational scaling in agent count while maintaining coordination quality.

Chapter 5: Mixed Traffic with Human Drivers. Real-world deployment requires coordination not only among autonomous vehicles but also with human-driven vehicles exhibiting uncertain and diverse behaviors. We extend the framework to incorporate probabilistic human behavior models, uncertainty quantification, and adaptive safety mechanisms that ensure robust coordination despite behavioral unpredictability.

Together, these extensions transform the baseline MCTS framework established in this chapter into a comprehensive coordination system capable of handling realistic multi-agent scenarios with guaranteed safety and real-time performance.

Chapter 4

Scalable Coordination via Interaction Graph and Level-k Reasoning

Chapter 3 established that Monte Carlo Tree Search provides effective coordination for four-agent intersection scenarios, achieving zero collision rate with real-time computational performance. However, as identified in Section 3.6.2, vanilla MCTS faces scalability barriers when agent count increases beyond four. This chapter addresses these limitations by introducing two complementary mechanisms: a dynamic interaction graph that filters irrelevant agents from each vehicle's planning problem, and a Level-k cognitive hierarchy that decomposes multi-agent reasoning into tractable sequential optimizations.

The key observation is that not all agents are equally relevant to each vehicle's decision-making. An agent approaching from the north need not reason about detailed interactions with agents approaching from the east if their trajectories will never intersect. Similarly, in a cognitive hierarchy where agents reason about opponents' responses, higher-level agents need not explicitly model peers at equal reasoning depths. By exploiting both spatial and strategic structure, we achieve linear computational scaling in agent count (formally analysed in Section 4.5.4 and Table 4.5) while maintaining coordination quality and safety guarantees.

We validate the proposed MCTS-Level-k framework through comprehensive experiments on eight-agent symmetric intersection scenarios, demonstrating that the approach successfully resolves coordination challenges that cause vanilla MCTS to fail with collision rates exceeding 15%. The framework achieves zero collision rate, over 20% improvement in arrival rate compared to vanilla MCTS (and over 100% compared to Stackelberg), and 44% reduction in computation time.

4.1 Scalability Challenge: From Four to Eight Agents

Before introducing our solutions, we first quantify the scalability barrier that motivates this chapter's contributions. The transition from four-agent to eight-agent scenarios represents not merely a doubling of problem size, but an exponential explosion in computational complexity that fundamentally changes the nature of the coordination challenge.

4.1.1 Exponential Complexity Growth

Recall from Chapter 3 that the joint action space for multi-agent coordination grows as $\mathcal{O}(|\mathcal{A}|^{N \cdot H})$, where $|\mathcal{A}|$ denotes the action space size per agent, N is the number of agents, and H is the planning horizon. For the four-agent scenario with $|\mathcal{A}| = 15$ actions and $H = 9$ steps, the joint space contains approximately 10^{42} possible trajectory combinations. While enormous, this space remains tractable for MCTS's selective exploration.

Scaling to eight agents transforms this challenge fundamentally:

$$\underbrace{15^{4\times 9}}_{\text{4 agents}} = 10^{42} \quad \longrightarrow \quad \underbrace{15^{8\times 9}}_{\text{8 agents}} = 10^{85}. \tag{4.1}$$

This 10^{43}-fold increase exceeds the number of atoms in the observable universe ($\approx 10^{80}$), representing a complexity regime where brute-force exploration becomes fundamentally impossible regardless of computational resources.

4.1.2 Empirical Performance Degradation

The theoretical complexity explosion manifests as severe practical performance degradation. Table 4.1 presents vanilla MCTS performance across agent counts. Note that these results employ MCTS *without* the safety-validated expansion procedure from Section 3.5, isolating the scalability effects from safety mechanisms. The resulting collision rates at $N = 4$ (10.7%) contrast with the zero-collision performance reported in Chapter 3, where safety-validated expansion was active, confirming that both scalability solutions *and* safety mechanisms are essential for reliable coordination.

Table 4.1: Vanilla MCTS Performance Degradation with Increasing Agent Count

Agent Count N	Collision Rate (%)	Arrival Rate (%)	Computation Time (ms)	Success Rate (%)
4	10.7 ± 5.2	91.5 ± 8.6	60.4 ± 17.5	89.3
6	18.3 ± 7.1	78.2 ± 12.3	156.7 ± 34.2	81.7
8	34.3 ± 12.8	47.1 ± 18.7	387.4 ± 89.6	65.7

Note: Results obtained using vanilla MCTS without safety-validated expansion to isolate scalability effects.

Three critical observations emerge from this analysis:

Safety degradation: Collision rate increases from 10.7% at $N = 4$ to 34.3% at $N = 8$, a threefold deterioration that renders the approach unsuitable for safety-critical deployment. With one-third of trials resulting in collisions, vanilla MCTS cannot provide the safety guarantees required for autonomous vehicle coordination.

Efficiency collapse: Arrival rate drops from 91.5% to 47.1%, indicating that fewer than half of vehicles successfully navigate the intersection within acceptable time bounds. This degradation reflects increased deadlock frequency as agents struggle to discover coordinated passing sequences.

Computational intractability: Computation time increases from 60.4 ms to 387.4 ms, exceeding the 100 ms real-time threshold by nearly fourfold. This growth stems from two compounding factors: collision checking overhead scaling as $\mathcal{O}(N^2)$ per tree node, and the need for deeper search to discover coordination strategies in larger action spaces.

4.1.3 Root Cause Analysis

The performance degradation stems from a fundamental mismatch between vanilla MCTS's exploration strategy and the structure of multi-agent coordination problems. MCTS achieves efficiency through selective exploration guided by value estimates, but this selectivity assumes that random sampling will eventually discover promising regions of the action space. As agent count increases, the probability of randomly sampling coordinated joint trajectories diminishes exponentially, causing MCTS to waste computational resources exploring unproductive regions.

Consider the coordination requirements for eight-agent intersection crossing. Successful coordination requires establishing an implicit passing order among all eight agents, with each agent's timing precisely calibrated to avoid conflicts with seven others. The probability that random rollout policies discover such coordinated sequences by chance approaches zero in the 10^{85}-dimensional joint action space.

A second limitation is that vanilla MCTS treats all agents as equally relevant, requiring collision checking against all $N-1$ opponents at every tree node. This exhaustive approach ignores the spatial structure of intersection coordination: only agents with intersecting trajectories pose collision risks. An agent traveling north-to-south need not reason about east-to-west traffic if their paths never cross.

These observations motivate the two-pronged approach developed in this chapter: *spatial filtering* through dynamic interaction graphs reduces the effective opponent count by identifying geometrically relevant interactions, while *strategic filtering* through Level-k reasoning decomposes multi-agent optimization into sequential single-agent problems with bounded opponent modeling.

4.2 Dynamic Interaction Graph

The first component of our scalability solution is a dynamic interaction graph that captures only strategically and spatially relevant agent relationships. Rather than requiring each agent to model all $N-1$ opponents, the interaction graph identifies the subset of agents whose trajectories may conflict, dramatically reducing the problem dimensionality.

4.2.1 Graph Representation

The interaction graph at time t is represented as a directed graph $G^t = (A, E^t)$, where $A = \{1, 2, \ldots, N\}$ is the set of agent indices and $E^t \subseteq A \times A$ is the set of directed interaction edges. The graph is directed because influence relationships need not be symmetric: agent i may need to account for agent j's behavior while agent j can safely ignore agent i due to asymmetries in their trajectories or relative positions. For instance, a vehicle already clearing the intersection at high speed influences the planning of an approaching vehicle, but need not itself account for the slower approaching agent whose trajectory poses no conflict to its own path.

A directed edge $(j, i) \in E^t$ indicates that agent j influences the decision-making of agent i at time t. The time-varying nature of the graph reflects that interaction topology evolves as agents move and their potential conflicts change over time.

4.2.2 Spatial Filtering via Trajectory Conflict Prediction

Spatial filtering identifies which agents may pose collision risks to the ego agent over a finite prediction horizon, thereby warranting explicit consideration during planning. The filtering mechanism operates by forward-simulating predicted trajectories and detecting geometric conflicts.

For each agent $j \neq i$, we predict future positions using the constant velocity model introduced in Equation (3.22):

$$\hat{s}_j^k = \begin{bmatrix} x_j^t + k v_j^t \cos(\theta_j^t)\Delta t \\ y_j^t + k v_j^t \sin(\theta_j^t)\Delta t \\ v_j^t \\ \theta_j^t \end{bmatrix}, \tag{4.2}$$

where $\hat{s}_j^k \triangleq \hat{s}_j(t + k\Delta t)$ represents the predicted state of agent j at future time step k assuming constant velocity v_j and heading θ_j, over the planning horizon $k \in \{1, \ldots, H\}$.

A directed edge (j, i) is included in the spatial interaction set if and only if agents i and j exhibit potential trajectory conflicts:

$$(j,i) \in E^t \iff \exists k \in \{1,\ldots,H\} : \text{Conflict}(\hat{s}_i^k, \hat{s}_j^k) = \text{true}, \tag{4.3}$$

where the conflict predicate evaluates whether predicted states $\hat{s}_i^k$ and $\hat{s}_j^k$ at future time step k result in proximity below the conflict threshold:

$$\text{Conflict}(\hat{s}_i^k, \hat{s}_j^k) = \begin{cases} \text{true}, & \text{if } d(\hat{s}_i^k, \hat{s}_j^k) < d_{\text{conflict}}, \\ \text{false}, & \text{otherwise}, \end{cases} \tag{4.4}$$

where $d(\hat{s}_i^k, \hat{s}_j^k)$ denotes the minimum Euclidean distance between the oriented rectangular representations of agents i and j at predicted states, computed via the Separating Axis Theorem from Equation (3.8), and d_{conflict} is the conflict detection threshold (set larger than the safety threshold σ_{safe} for early warning). In this work, d_{conflict} is set as a static constant $d_{\text{conflict}} = 5\,\text{m}$, independent of agent speed or heading. This design choice is appropriate for the unsignalized intersection scenario considered here, where vehicle speeds are moderate and a fixed spatial margin provides sufficient early warning for conflict detection. A speed-dependent or time-to-collision-based dynamic threshold would be more suitable for high-speed highway scenarios but introduces unnecessary complexity in the low-speed intersection setting.

4.2.3 Interaction Set Construction

For each agent i, the spatial interaction set $\mathcal{N}_i^{\text{spatial}}$ contains all agents with predicted trajectory conflicts:

$$\mathcal{N}_i^{\text{spatial}} = \{j \in A \setminus \{i\} : (j,i) \in E^t\}. \tag{4.5}$$

This set typically contains far fewer agents than the complete opponent set $A \setminus \{i\}$. In eight-agent intersection scenarios, spatial filtering reduces the average interaction set size from 7 (all opponents) to approximately 3–4 geometrically relevant agents, depending on the specific configuration and maneuver types.

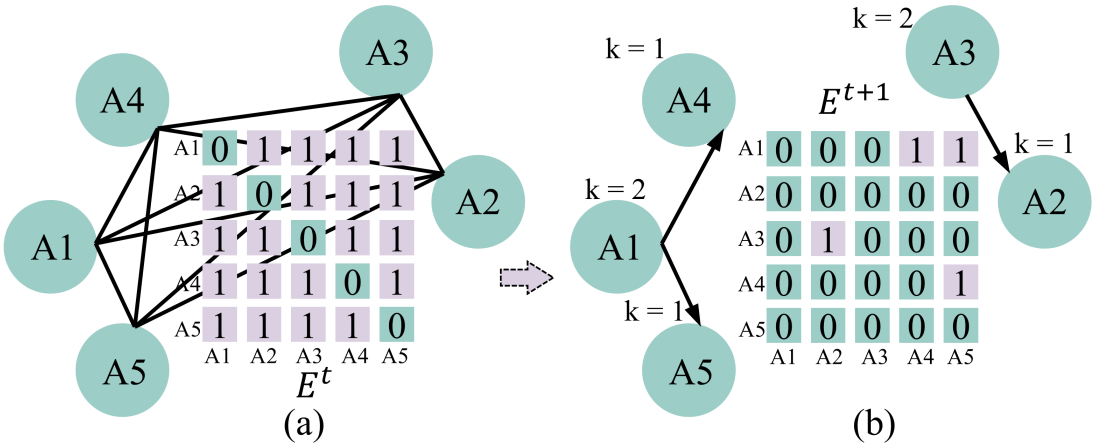


Figure 4.1: Illustration of spatial filtering in the interaction graph. (a) Complete interaction graph with all pairwise connections. (b) Filtered graph retaining only edges where trajectory conflicts are predicted. Agent A1 need only consider A4 and A5, ignoring spatially distant agents A2, A3.

Figure 4.1 illustrates the filtering effect. The complete graph (left) contains $N(N-1) = 56$ directed edges for eight agents. After spatial filtering (right), the graph retains only edges for predicted conflicts, dramatically reducing the modeling burden. In this example, the relevant opponent set for agent A1 is reduced from seven to only two.

4.2.4 Complexity Reduction Analysis

The computational benefit of spatial filtering manifests in multiple algorithmic phases. During MCTS expansion, the safety validation procedure from Equation (3.21) is modified to perform collision checking only against agents in the filtered set $\mathcal{N}_i^{\text{spatial}}$ rather than all opponents:

$$\text{Safe}(s_i, a_i) = \bigwedge_{j \in \mathcal{N}_i^{\text{spatial}}} \neg\text{Collision}(f(s_i, a_i), \hat{s}_j^{t+1}), \tag{4.6}$$

where $f(s_i, a_i)$ denotes the successor state from applying action a_i to state s_i via the dynamics model (3.4), and the conjunction ranges only over spatially relevant opponents.

This filtering reduces per-node collision checking from $\mathcal{O}(N)$ to $\mathcal{O}(|\mathcal{N}_i^{\text{spatial}}|)$ operations. With $|\mathcal{N}_i^{\text{spatial}}| \approx 3$–$4$ versus $N-1=7$ opponents in eight-agent scenarios, spatial filtering achieves approximately 50% reduction in collision checking overhead.

However, spatial filtering alone does not address the fundamental challenge of multi-agent strategic reasoning. Even with reduced opponent sets, each agent must still reason about how filtered opponents will behave—a recursive modeling problem that remains computationally demanding. The following section introduces Level-k reasoning to decompose this strategic complexity.

4.3 Level-k Cognitive Hierarchy

While spatial filtering reduces the number of agents requiring consideration, it does not address the strategic complexity of predicting opponent behaviors. Each agent must anticipate how others will respond to its actions, creating a recursive reasoning problem that classical game-theoretic approaches resolve through equilibrium computation—an exponentially expensive procedure. This section introduces Level-k cognitive hierarchy as an alternative that achieves strategic sophistication through bounded recursive reasoning.

4.3.1 Cognitive Hierarchy Structure

Level-k reasoning posits a hierarchy of reasoning depths where agents at each level best-respond to beliefs about agents at lower levels. The hierarchy is defined recursively:

- **Level-0**: A baseline behavior model that does not involve strategic reasoning. In our framework, Level-0 serves as a *safety initialization procedure* rather than a behavioral type.
- **Level-1**: Agents at Level-1 optimize their policies assuming all opponents follow Level-0 baseline trajectories.
- **Level-2**: Agents at Level-2 optimize assuming opponents employ Level-1 strategies, enabling anticipation of strategic responses.

This bounded hierarchy provides computational tractability: rather than computing game-theoretic equilibria requiring exponential joint action enumeration, each agent solves a sequence of single-agent optimization problems with fixed opponent models.

4.3.2 Level-0: Conservative Safety Initialization

A critical innovation in our framework is the reformulation of Level-0 from a naive behavioral model to a universal safety initialization procedure. Classical Level-k defines Level-0 as acting randomly, creating unsafe foundations that propagate through higher reasoning levels. We instead define Level-0 as a conservative trajectory generation procedure that establishes safety margins for all agents.

During Level-0 initialization, each agent i computes a baseline trajectory by treating others as obstacles with constant velocity, then solving a safety-prioritized optimization:

$$\tau_i^{(0)} = \arg\min_{\tau \in \mathscr{T}_i} \sum_{k=0}^{H-1} \left[w_{\mathrm{s}}^{(0)} c_{\mathrm{s}}^i(S_i^k) + w_{\mathrm{e}}^{(0)} c_{\mathrm{e}}^i(s_i^k) + w_{\mathrm{d}}^{(0)} c_{\mathrm{d}}^i(s_i^k) + w_{\mathrm{c}}^{(0)} c_{\mathrm{c}}^i(a_i^k) \right], \tag{4.7}$$

where $\tau_i^{(0)} = \{s_i^1, s_i^2, \ldots, s_i^H\}$ denotes the Level-0 baseline trajectory, $\mathscr{T}_i$ represents the space of feasible trajectories satisfying dynamics and constraint requirements, other agents' future states within S_i^k are predicted using the constant velocity model (4.2), and the weight configuration emphasizes safety with $w_{\mathrm{s}}^{(0)} \gg w_{\mathrm{e}}^{(0)}$ (specifically, $w_{\mathrm{s}}^{(0)} = 100$, $w_{\mathrm{d}}^{(0)} = 10$,

$w_c^{(0)} = 5$, $w_e^{(0)} = 2$), reflecting the priority ordering $w_s^{(0)} \gg w_d^{(0)} > w_c^{(0)} > w_e^{(0)}$: safety is the dominant objective, followed by path adherence, comfort, and efficiency. This ordering ensures that Level-0 agents prioritize collision avoidance above all else, while maintaining reasonable path tracking and smooth control before optimizing for speed.

Critically, Level-0 optimization employs extended safety margins. Each vehicle's geometric footprint is expanded by a buffer ε_0 during collision checking:

$$\mathscr{V}_i^{\text{ext}} = \mathscr{V}_i \oplus \mathscr{B}_{\varepsilon_0}, \tag{4.8}$$

where $\mathscr{V}_i$ represents the actual vehicle geometry, $\oplus$ denotes the Minkowski sum operation, and $\mathscr{B}_{\varepsilon_0}$ is a disk of radius $\varepsilon_0 = 0.6\,\text{m}$. This extended footprint ensures conservative spacing even under pessimistic assumptions about opponent behaviors.

The resulting Level-0 baselines $\{\tau_i^{(0)}\}_{i=1}^N$ serve not as execution policies but as safety anchors for subsequent strategic reasoning. Every agent computes its Level-0 baseline before engaging in higher-level optimization, ensuring safety margins are established regardless of assigned reasoning level.

4.3.3 Level-1: Best Response to Baselines

A Level-1 agent optimizes its trajectory assuming all opponents follow their Level-0 baselines. This transforms the multi-agent problem into a single-agent MDP where opponent behaviors are treated as deterministic environmental dynamics:

$$\pi_i^{(1)*} = \arg\min_{\pi_i \in \Pi_i} \mathbb{E}\left[\sum_{k=0}^{H-1} c_i(S_i^k, a_i^k) \mid \pi_j = \pi_j^{(0)}, \forall j \in \mathscr{N}_i\right], \tag{4.9}$$

where $\pi_j^{(0)}$ denotes the policy that generates opponent j's Level-0 baseline $\tau_j^{(0)}$, and $\mathscr{N}_i$ represents the filtered interaction set from Section 4.2.

The Level-1 optimization employs modified cost weights compared to Level-0, typically reducing safety emphasis while increasing efficiency weight (e.g., $w_{\mathrm{s}}^{(1)} = 80$, $w_{\mathrm{e}}^{(1)} = 5$). This rebalancing allows agents to pursue more efficient trajectories while ensuring safety margins inherited from Level-0 are never violated, with path adherence and comfort weights unchanged from Level-0: $w_{\mathrm{s}}^{(1)} = 80$, $w_{\mathrm{e}}^{(1)} = 5$, $w_{\mathrm{d}}^{(1)} = 10$, $w_{\mathrm{c}}^{(1)} = 5$).

4.3.4 Level-2: Anticipating Strategic Responses

Level-2 agents introduce an additional reasoning layer by anticipating that opponents employ Level-1 strategies rather than merely following Level-0 baselines. A Level-2 agent i recognizes that other agents will respond strategically to observations, leading to behaviors differing from conservative baselines:

$$\pi_i^{(2)*} = \arg\min_{\pi_i \in \Pi_i} \mathbb{E}\left[\sum_{k=0}^{H-1} c_i(S_i^k, a_i^k) \mid \pi_j = \pi_j^{(1)*}, \forall j \in \mathscr{N}_i^{(2)}\right], \tag{4.10}$$

where $\pi_j^{(1)*}$ represents the predicted Level-1 optimal policy for opponent j, and $\mathscr{N}_i^{(2)}$ contains opponents modeled at Level-1, including both actual Level-1 agents (modeled accurately) and Level-2 agents (conservatively approximated as Level-1) (discussed below).

Computing Level-2 policies requires agent i to simulate Level-1 optimization from each opponent's perspective, effectively solving nested optimization problems. This recursive computation remains tractable because: (1) the filtered interaction set limits the number of opponents requiring modeling, and (2) the bounded hierarchy depth (maximum Level-2) prevents infinite recursion.

4.3.5 Strategic Filtering via Reasoning Level

Beyond spatial filtering, our framework introduces strategic filtering based on the cognitive hierarchy. A key property of Level-k reasoning is that agents need not model opponents at equal or higher reasoning levels—their optimization depends only on predicted behaviors of lower-level agents.

Agents are dynamically assigned to reasoning levels based on interaction complexity. The complexity score for agent i combines proximity, density, and conflict factors:

$$C_i = w_p \cdot \frac{1}{d_i} + w_d \cdot \rho_i + w_c \cdot N_i^{\text{conflict}}, \tag{4.11}$$

where d_i represents distance to the intersection center, ρ_i counts agents within a radius, N_i^{conflict} counts predicted trajectory conflicts, and (w_p, w_d, w_c) are weighting parameters.

Level assignment follows threshold-based classification:

$$k_i = \begin{cases} 2, & \text{if } C_i > C_{\text{th}}, \\ 1, & \text{otherwise}, \end{cases} \tag{4.12}$$

where C_{th} is a tunable threshold balancing coordination quality against computation.

The complete filtered interaction set for agent i at level k_i combines spatial and strategic filtering:

$$\mathcal{N}_i^{(k_i)} = \{ j \in \mathcal{N}_i^{\text{spatial}} : k_j < k_i \}, \tag{4.13}$$

requiring opponents to be both spatially relevant (predicted trajectory conflict) and strategically relevant (lower reasoning level). This interaction set explicitly includes only lower-level opponents. However, agents at the same reasoning level are not ignored but are modeled at a reduced level of sophistication: a Level-2 agent models other Level-2 agents as Level-1, while a Level-1 agent reduces all opponents to Level-0 baselines. This systematic underestimation of opponent sophistication ensures that prediction errors are biased toward over-caution rather than over-trust, providing implicit safety margins without requiring worst-case assumptions.

4.3.6 Cascading Safety Property

A distinctive feature of our reconstructed Level-k framework is that safety emerges as a structural property of the cognitive hierarchy. Each reasoning level inherits and amplifies the conservatism of lower levels, creating progressively larger protective buffers:

$$\varepsilon_0 < \varepsilon_1 < \varepsilon_2, \tag{4.14}$$

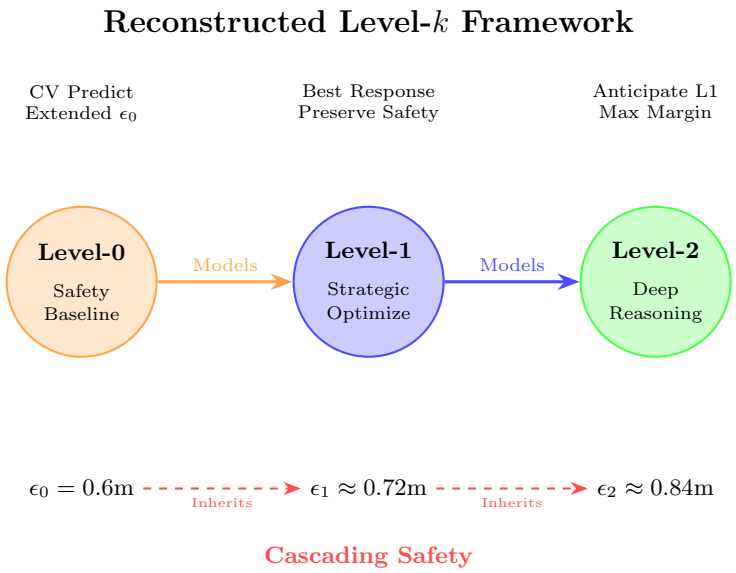


Figure 4.2: Cascading safety in the Level-k hierarchy. Level-0 establishes conservative baselines with margin ε_0. Level-1 agents optimizing against these baselines inherit safety margins under the constant-velocity modeling assumption
. Level-2 further refines against Level-1 responses, tending to amplify margins to $\varepsilon_2 > \varepsilon_1$ in practice, though formal guarantees at higher levels depend on modeling accuracy.

where ε_k represents the *effective* safety margin emergent from Level-k optimization, distinct from the explicit buffer $\mathscr{B}_{\varepsilon_0}$ applied during Level-0 collision checking.

This cascading property arises because Level-1 agents optimize against Level-0 baselines that already embed conservative margins ε_0. Maintaining collision-free trajectories relative to these conservative predictions ensures effective margins exceeding ε_0. Level-2 agents inherit amplified margins ε_1 from Level-1 predictions, further expanding protective buffers.

Figure 4.2 illustrates this cascading mechanism. Unlike classical Level-k formulations where random Level-0 behaviors create unsafe foundations, our procedural Level-0 initialization ensures safety propagates through the hierarchy. Higher reasoning depth never compromises safety—it strengthens guarantees while enabling more efficient coordination.

4.4 MCTS-Level-k Planning Algorithm

Having established the interaction graph for spatial filtering and the Level-k hierarchy for strategic decomposition, we now integrate these components into a unified planning algorithm. The MCTS-Level-k framework operates through four sequential phases executed at each planning cycle, transforming the intractable multi-agent coordination problem into a series of tractable single-agent optimizations.

4.4.1 Algorithm Overview

At each discrete time step t, every agent i executes the four-phase planning procedure illustrated in Figure 4.3:

- **Phase I (Level-0 Initialization)**: Generate conservative baseline trajectories for all agents, establishing safety anchors.
- **Phase II (Level Assignment)**: Assess interaction complexity and assign each agent to reasoning level $k_i \in \{1,2\}$.
- **Phase III (Graph Construction)**: Build the filtered interaction set $\mathcal{N}_i^{(k_i)}$ through spatial and strategic filtering.
- **Phase IV (MCTS Planning)**: Execute Monte Carlo Tree Search over the induced single-agent MDP to compute the optimal action.

The algorithm produces an optimal action a_i^* executed during the current time step. After execution, the environment evolves, new observations arrive, and the planning cycle repeats in a receding horizon fashion.

4.4.2 Induced Single-Agent MDP

The Level-k hierarchy transforms each agent's multi-agent planning problem into a single-agent Markov Decision Process by treating predicted opponent behaviors as environmental dynamics. For agent i at reasoning level k_i, the induced MDP is defined as:

$$\mathcal{M}_i^{(k_i)} = (\mathcal{S}_i, \mathcal{A}_i, P_i^{(k_i)}, r_i^{(k_i)}, H), \tag{4.15}$$

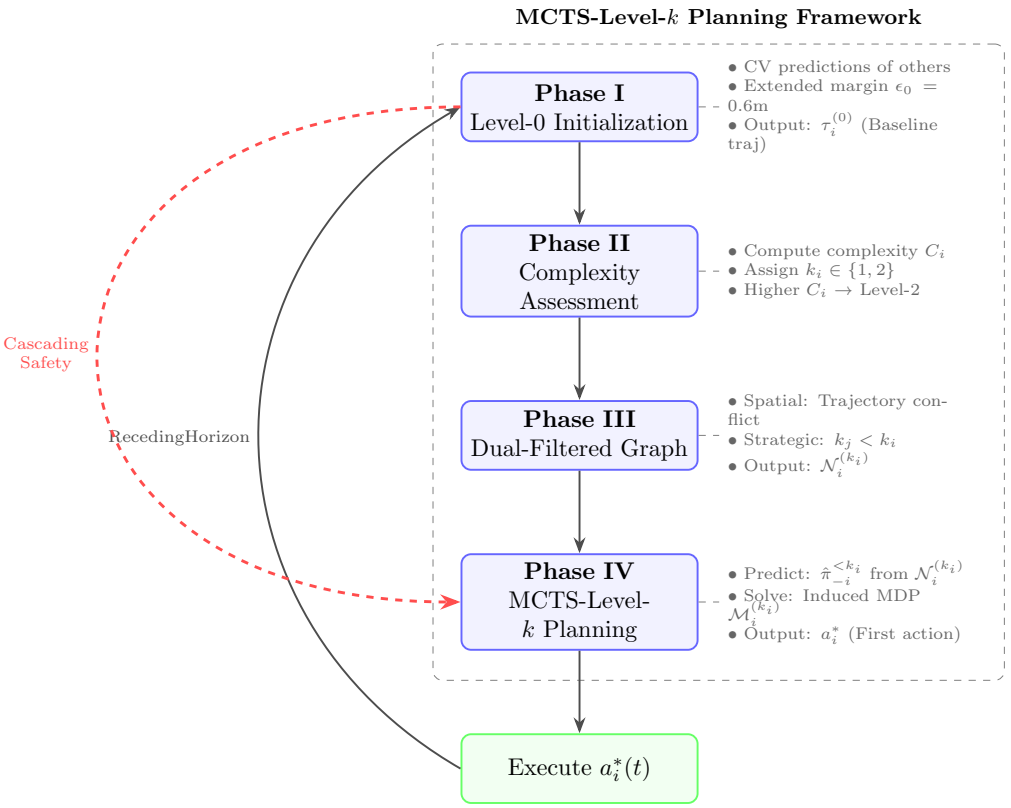


Figure 4.3: MCTS-Level-k planning framework overview. Phase I generates Level-0 baselines $\hat{\tau}_i^{(0)}$ with conservative safety margins. Phase II assigns reasoning levels based on TTC scores. Phase III constructs filtered interaction sets through dual filtering, ubstantially reducing the number of opponents each agent must explicitly reason about . Phase IV executes MCTS planning to compute optimal actions. The dual filtering mechanism reduces effective search complexity compared to joint planning, while the safety-aware Level-0 baseline provides structural safety margins that propagate through higher reasoning levels under the adopted modeling assumptions.

where $\mathscr{S}_i$ is the local state space from Equation (3.2), $\mathscr{A}_i$ is the discrete action space, $P_i^{(k_i)}$ is the induced transition function, $r_i^{(k_i)}$ is the induced reward function, and H is the planning horizon for sequential decision-making.

The induced transition dynamics capture how agent i's state evolves under predicted opponent responses:

$$P_i^{(k_i)}(s_i' \mid s_i, a_i) = \mathbf{1}\left[s_i' = f(s_i, a_i)\right] \cdot \prod_{j \in \mathcal{N}_i^{(k_i)}} \mathbf{1}\left[\hat{s}_j' = \hat{f}_j^{(k_i-1)}(\hat{s}_j)\right], \tag{4.16}$$

where $f(s_i, a_i)$ denotes state transition via the dynamics model (3.4), $\hat{f}_j^{(k_i-1)}$ represents opponent j's predicted transition under Level-(k_i-1) policy, and $\mathbf{1}[\cdot]$ is the indicator function. The deterministic nature of both ego dynamics and opponent predictions yields deterministic induced transitions.

The induced reward function evaluates agent i's instantaneous reward accounting for predicted opponent positions:

$$r_i^{(k_i)}(s_i, a_i) = -c_i(S_i, a_i), \tag{4.17}$$

where $S_i = \{s_i\} \cup \{\hat{s}_j : j \in \mathcal{N}_i^{(k_i)}\}$ represents the joint state comprising agent i's actual state and predicted opponent states, and $c_i(\cdot)$ is the cost function from Equation (3.11).

4.4.3 MCTS with Level-k Opponent Modeling

The MCTS algorithm from Chapter 3 is adapted to incorporate Level-k opponent predictions. The key modification occurs in rollout simulation and safety validation, where opponent behaviors are predicted according to their assigned reasoning levels.

4.4.3.1 Safety-Aware Expansion

During tree expansion, the safe action set is constructed by validating actions against predicted Level-(k_i-1) opponent trajectories:

$$\mathcal{A}_{\text{safe}}^{(k_i)}(n) = \left\{ a \in \mathcal{A}_i : \bigwedge_{j \in \mathcal{N}_i^{(k_i)}} \neg\text{Collision}\left(f(s(n), a), \hat{s}_j^{(k_i-1)}\right) \right\}, \tag{4.18}$$

where $s(n)$ denotes the state at tree node n, $f(s(n), a)$ is the successor state after applying action a, and $\hat{s}_j^{(k_i-1)}$ represents opponent j's predicted Level-(k_i-1) trajectory.

For Level-1 agents ($k_i = 1$), opponent predictions use Level-0 baselines: $\hat{s}_j^{(0)} \in \boldsymbol{\tau}_j^{(0)}$. For Level-2 agents ($k_i = 2$), opponent predictions use Level-1 optimal trajectories, requiring recursive computation of Level-1 policies for each opponent in the interaction set.

4.4.3.2 Level-k Rollout Simulation

Rollout simulation evaluates trajectory quality by forward-simulating from expanded nodes to the planning horizon. At each rollout step k, the reward is computed using predicted opponent states:

$$V_{\text{rollout}}^{(k_i)}(n) = \sum_{t=d(n)}^{H-1} r_i^{(k_i)}(s_i^t, a_i^t), \tag{4.19}$$

where $d(n)$ denotes the depth of node n, s_i^t evolves according to rollout policy π_{default}, and the reward $r_i^{(k_i)}$ incorporates predicted opponent positions from Level-$(k_i - 1)$ trajectories.

The rollout policy π_{default} employs simple heuristics such as constant velocity or reference path following, providing baseline trajectory estimates without expensive optimization.

4.4.4 Computational Complexity Analysis

The MCTS-Level-k framework achieves substantial complexity reduction compared to vanilla MCTS through three mechanisms:

- *Spatial filtering* reduces collision checking from $\mathcal{O}(N-1)$ to $\mathcal{O}(|\mathcal{N}_i^{\text{spatial}}|)$ opponents per node, with a typical reduction from 7 to 3–4 agents in eight-agent scenarios.
- *Strategic filtering* further reduces opponent modeling by restricting attention to lower reasoning levels, removing approximately half of the spatially relevant opponents for Level-2 agents.
- *Safety-aware pruning* eliminates unsafe actions during expansion, reducing the effective branching factor from $|\mathcal{A}| = 15$ to $|\mathcal{A}_{\text{safe}}| \approx 4$–$5$ actions per node.

The combined effect reduces the per-iteration computational cost from:

$$\underbrace{\mathcal{O}(|\mathcal{A}| \cdot H \cdot N)}_{\text{Vanilla MCTS per iteration}} \quad \longrightarrow \quad \underbrace{\mathcal{O}(|\mathcal{A}_{\text{safe}}| \cdot H \cdot |\mathcal{N}_i^{(k_i)}|)}_{\text{MCTS-Level-}k\text{ per iteration}}. \tag{4.20}$$

More significantly, the Level-k decomposition avoids explicit enumeration of the joint action space $\mathcal{O}(|\mathcal{A}|^{N \cdot H})$. Rather than searching the full joint space of $\mathcal{O}(15^{8\times 9}) \approx 10^{85}$ combinations, each agent independently runs MCTS over its own action space with a filtered opponent set, requiring only $\mathcal{O}(N \cdot K \cdot b_{\text{eff}} \cdot H) \approx 6.3 \times 10^4$ operations in total (Table 4.5). We note that this comparison is between the theoretical joint search space and the actual operations performed by the proposed framework, rather than a like-for-like algorithmic comparison.

4.4.5 Complete Algorithm

The complete MCTS-Level-k planning procedure is formalized in Algorithm 2.

4.5 Experimental Validation

This section evaluates the MCTS-Level-k framework through simulation-based studies on eight-agent symmetric intersection scenarios. Within the evaluated scenarios, the proposed approach demonstrates improved coordination compared to vanilla MCTS, with no collisions observed while maintaining real-time computational performance, under the adopted simplified modeling assumptions.

4.5.1 Experimental Setup

4.5.1.1 Scenarios

We evaluate two challenging eight-agent configurations:

Case 1: All-Straight Symmetric Intersection. Eight agents approach from four directions (two per direction) on straight paths, creating maximal symmetry with no natural priority ordering. All agents start 18 m from the intersection center.

Case 2: Mixed Maneuver Intersection. Inner lane agents (A1, A3, A5, A7) execute left turns while outer lane agents (A2, A4, A6, A8) proceed straight, creating heterogeneous trajectory conflicts that increase coordination complexity.

Algorithm 2: MCTS-Level-k Planning Framework

Input: Current state s_i^t, observed neighbor states $\{s_j^t\}_{j\in\mathcal{N}_i^{\text{spatial}}}$, horizon H, iterations K
Output: Optimal action a_i^*
// Phase I: Level-0 Initialization
$\tau_i^{(0)} \leftarrow \text{ConservativeBaseline}(s_i^t, \{\hat{s}_j^t\}_{j\in\mathcal{N}_i^{\text{spatial}}})$ {Eq. (4.7)}
for each neighbor $j \in \mathcal{N}_i^{\text{spatial}}$ **do**
    $\tau_j^{(0)} \leftarrow \text{PredictBaseline}(s_j^t)$ {Reconstruct Level-0 baseline assuming j treats others as constant-velocity obstacles}
end for
// Phase II: Level Assignment
for each agent $j \in A$ **do**
    $C_j \leftarrow w_p/d_j + w_d \cdot \rho_j + w_c \cdot N_j^{\text{conflict}}$ {Eq. (4.11)}
    $k_j \leftarrow \mathbf{1}[C_j > C_{\text{th}}] + 1$ {Eq. (4.12)}
end for
// Phase III: Interaction Graph Construction
$\mathcal{N}_i^{\text{spatial}} \leftarrow \text{SpatialFilter}(\{\tau_j^{(0)}\}_{j\neq i})$ {Eq. (4.5)}
$\mathcal{N}_i^{(k_i)} \leftarrow \{j \in \mathcal{N}_i^{\text{spatial}} : k_j < k_i\}$ {Eq. (4.13); same-level agents modeled at k_i-1}
// Phase IV: MCTS Planning
Initialize root n_0 with $s(n_0) = s_i^t$, $Q(n_0) = 0$, $N(n_0) = 0$
for each of K iterations **do**
    $n \leftarrow \text{Select}(n_0)$ {UCT selection, Eq. (3.20)}
    if $d(n) < H$ and n not fully expanded **then**
        $\mathcal{A}_{\text{safe}} \leftarrow \text{SafetyValidate}(s(n), \mathcal{N}_i^{(k_i)})$ {Eq. (4.18)}
        $n_{\text{new}} \leftarrow \text{Expand}(n, \mathcal{A}_{\text{safe}})$
        $n \leftarrow n_{\text{new}}$
    end if
    $V \leftarrow \text{Rollout}^{(k_i)}(n, \mathcal{N}_i^{(k_i)})$ {Eq. (4.19)}
    $\text{Backpropagate}(n, n_0, V)$
end for
return $\arg\max_{a\in\text{Children}(n_0)} Q(n_0,a)/N(n_0,a)$

4.5.1.2 Baseline Methods

We compare against established multi-agent coordination approaches:

- **Stackelberg Game** [33]: Hierarchical leader-follower coordination.
- **Nash Equilibrium** [19]: Simultaneous game-theoretic coordination.
- **Vanilla MCTS**: The baseline framework from Chapter 3 without Level-k reasoning.

All baseline methods use 1000 maximum iterations, while our MCTS-Level-k framework uses only $K = 300$ iterations with planning horizon $H = 9$.

4.5.1.3 Simulation Parameters

The simulation parameters for Chapter 4 follow the baseline setup in Table 3.2, with the following Level-k-specific additions summarised in Table 4.2.

Table 4.2: Level-k Specific Parameters

Parameter	**Symbol**	**Value**
Number of agents	N	8
Planning horizon	H	9
MCTS iterations (ours)	K	300
MCTS iterations (baselines)	K	1000
Conflict threshold	d_{conflict}	5.0 m
Level-0 safety buffer	ε_0	0.6 m
Cost weights Level-0	(w_s, w_d, w_c, w_e)	$(100, 10, 5, 2)$
Cost weights Level-1	(w_s, w_d, w_c, w_e)	$(80, 10, 5, 5)$
Complexity threshold	C_{th}	tuned
Number of trials	–	40
Time step	Δt	0.2 s

4.5.1.4 Evaluation Metrics

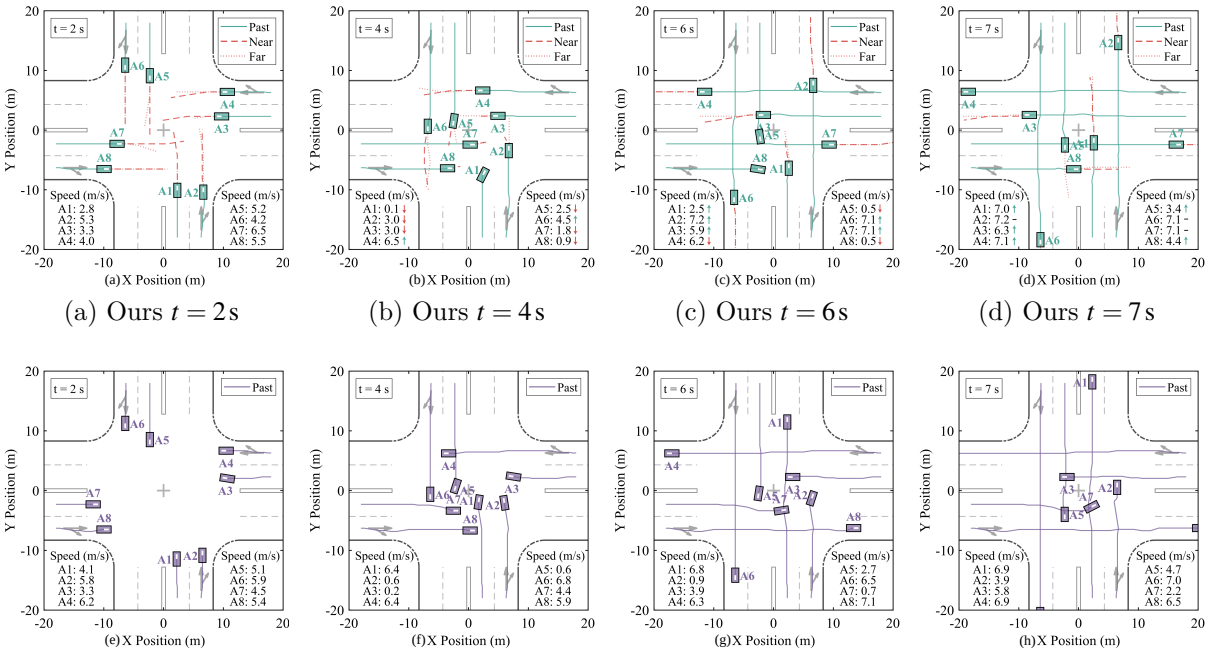
Performance is evaluated across safety, efficiency, and computational dimensions:

- **Collision Rate (%)**: Percentage of trials with inter-agent collisions.
- **Arrival Rate (%)**: Percentage of agents clearing the intersection within time limit.
- **Travel Time (s)**: Average time for all agents to complete traversal.
- **Computation Time (ms)**: Per-step planning time.
- **Trajectory Deviation (m)**: Mean distance from reference paths.
- **Minimum Distance (m)**: Closest inter-agent separation throughout execution.

All experiments are conducted over 40 independent trials with randomized initial position perturbations sampled from a uniform distribution $\mathcal{U}(-0.05, 0.05)$ m applied independently to each agent's longitudinal starting position along the reference path, to assess robustness.

4.5.2 Case 1: All-Straight Symmetric Intersection

4.5.2.1 Qualitative Analysis



(a) Ours $t = 2\,\mathrm{s}$ (b) Ours $t = 4\,\mathrm{s}$ (c) Ours $t = 6\,\mathrm{s}$ (d) Ours $t = 7\,\mathrm{s}$

(e) Vanilla MCTS $t = 2\,\mathrm{s}$ (f) Vanilla MCTS $t = 4\,\mathrm{s}$ (g) Vanilla MCTS $t = 6\,\mathrm{s}$ (h) Vanilla MCTS $t = 7\,\mathrm{s}$

Figure 4.4: Temporal evolution of eight-agent all-straight coordination. Top row (a–d): MCTS-Level-k achieves smooth coordination with implicit turn-taking. Bottom row (e–h): Vanilla MCTS exhibits near-deadlock with all agents clustered at low speeds.

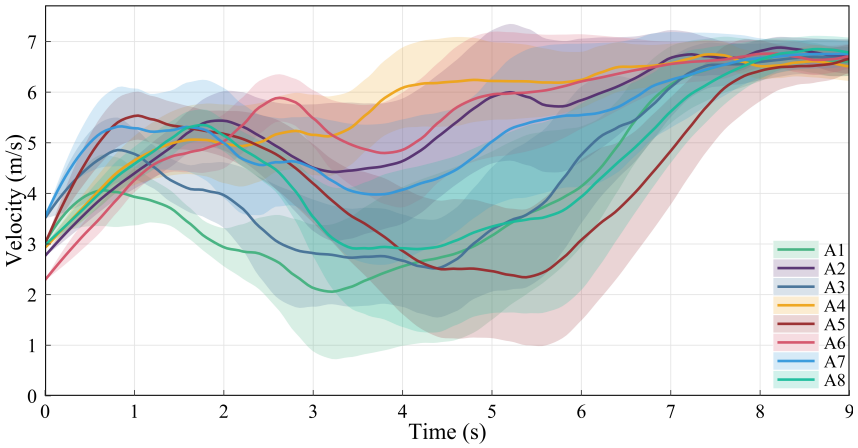
Figure 4.4 compares temporal evolution between MCTS-Level-k (top row) and vanilla MCTS (bottom row). The performance difference is striking.

At $t = 2\,\mathrm{s}$, our method demonstrates anticipatory coordination: agents maintain compact trajectories with smooth deceleration patterns indicating strategic planning. Vanilla MCTS shows reactive braking with larger lateral deviations, reflecting the absence of coordinated reasoning.

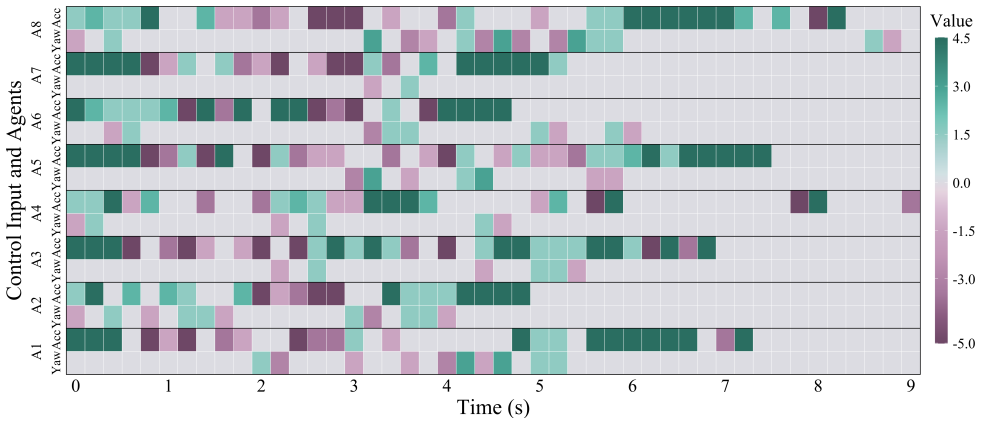
The critical phase at $t = 4\,\mathrm{s}$ reveals Level-k reasoning's key advantage. Our framework enables implicit turn-taking where agent A1 yields to nearly a standstill ($0.1\,\mathrm{m/s}$) while other agents proceed at moderate speeds, creating a natural passing order without explicit communication. In contrast, vanilla MCTS fails to establish coordinated yielding: agent A1 maintains a high speed of $6.4\,\mathrm{m/s}$ while other agents have not yet formed a stable passing order, reflecting the absence of strategic reasoning and leading to uncoordinated conflict resolution.

By $t = 7\,\mathrm{s}$, our framework has successfully resolved all conflicts, with agents accelerating smoothly out of the intersection at high speeds, indicating efficient and complete coordination. Under vanilla MCTS, most agents recover to 4.5–7 m/s, however agent A7 remains stalled below 3 m/s, indicating a residual coordination failure where one agent is unable to find a safe passage through the conflict zone. The projected trajectories in our method reveal predictive planning: agents continuously update their paths based on evolving neighbor behaviors, creating complementary trajectories that maximize throughput.

4.5.2.2 Temporal Dynamics



(a) Velocity profiles over time



(b) Control input heatmap

Figure 4.5: Temporal dynamics analysis for Case 1. (a) Velocity profiles showing coordinated speed adjustments with 95% confidence intervals. (b) Control input heatmap: upper row per agent shows acceleration ($\mathrm{m/s^2}$), lower row shows yaw rate (rad/s).

Figure 4.5 presents detailed temporal analysis of MCTS-Level-k coordination patterns.

The velocity profiles (Figure 4.5a) reveal emergent coordination structure. Agents maintain stable speeds around 5–6 m/s after initial acceleration, while agents A5 and A8 exhibit pronounced deceleration valleys reaching 1–2 m/s during $t = 3$–6 s before recovering. This asymmetric velocity pattern breaks the scenario's inherent symmetry, enabling sequential passage through the conflict zone. The narrow 95% confidence bands indicate consistent behavior across trials, demonstrating that Level-k reasoning produces reliable coordination strategies rather than random fluctuations.

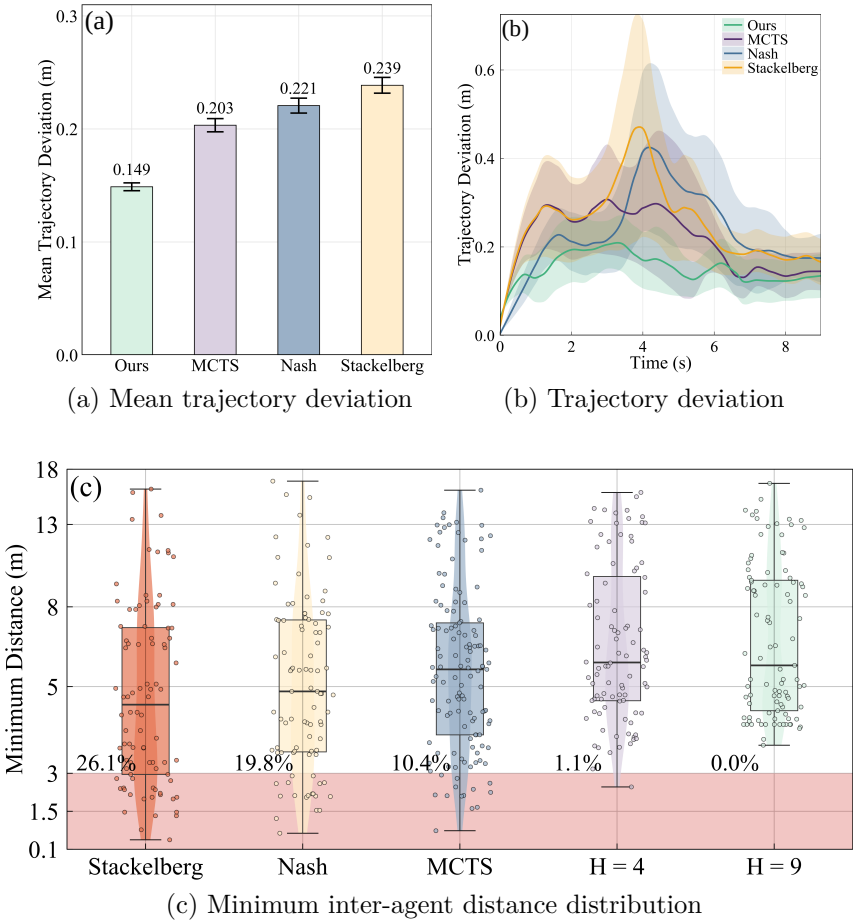
The control heatmap (Figure 4.5b) visualizes decision-making intensity throughout the scenario. Three notable patterns emerge:

1. *Coordinated deceleration*: Agents A1, A3, A5, and A8 exhibit the most conservative yielding behaviour, with A1 reaching the lowest speed among all agents during $t = 2$–4 s, followed by A3 and A5. These agents create temporal separation through sustained braking, enabling conflict-free passage for the proceeding agents.
2. *Implicit negotiation*: A clear implicit priority ordering emerges without explicit communication: agents A4 and A6 maintain the highest speeds and claim earlier passage, while A2 and A7 proceed at intermediate speeds. Meanwhile, A1, A3, A5, and A8 adopt yielding roles with progressively increasing deceleration intensity. This asymmetric yet stable ordering demonstrates that Level-k reasoning successfully breaks the inherent eight-way symmetry through strategic anticipation.
3. *Smooth control*: Gradual color transitions indicate smooth control evolution. Only 12% of actions exceed $|a| > 3\,\mathrm{m/s^2}$, confirming that coordination emerges through strategic anticipation rather than aggressive reactive maneuvering.

4.5.2.3 Statistical Analysis

Figure 4.6 presents statistical analysis across 40 independent trials.

Trajectory deviation (Figure 4.6a): Our method achieves the lowest mean deviation at $0.149 \pm 0.082\,\mathrm{m}$, representing 27% improvement over vanilla MCTS (0.203 m) and 36% over Nash equilibrium (0.221 m). The Stackelberg approach exhibits highest deviation due to its rigid leader-follower assumptions failing under perfect symmetry where no natural hierarchy exists to guide the decision process.



(a) Mean trajectory deviation (b) Trajectory deviation

(c) Minimum inter-agent distance distribution

Figure 4.6: Statistical analysis for Case 1 across 40 trials. (a) Mean trajectory deviation comparison. (b) Temporal evolution of deviation during the conflict phase. (c) Minimum distance distribution with safety threshold $d_{\text{safe}} = 3\,\text{m}$ (dashed line).

Temporal evolution (Figure 4.6b): During the critical conflict phase ($t = 3$–$5\,\text{s}$), our method maintains consistent low deviation while Nash and Stackelberg show pronounced peaks reaching 0.4–$0.6\,\text{m}$, indicating coordination breakdown. The narrow confidence bands for our approach demonstrate stability across trials.

Safety analysis (Figure 4.6c): The minimum distance distribution validates safety performance. Our method with $H = 9$ achieves 0% collision rate with median separation of $6\,\text{m}$. Stackelberg and Nash show 26.1% and 19.8% safety violations respectively (distances below the $3\,\text{m}$ threshold), with several collision events. Level-k reasoning provides both safety (no violations) and efficiency (compact distribution) simultaneously.

Table 4.3: Performance Comparison in Case 1: Eight-Agent All-Straight Intersection

Method	Collision Rate (%)	Arrival Rate (%)	Travel Time (s)	Computation Time (ms)	Max. Iters
Stackelberg	35.0	47.1 ± 15.3	12.1 ± 2.6	66.5 ± 19.1	1000
Nash	27.5	65.9 ± 10.5	13.2 ± 3.3	82.8 ± 20.3	1000
Vanilla MCTS	17.5	79.7 ± 8.8	11.5 ± 2.4	98.2 ± 21.1	1000
Ours	**0.0**	$\mathbf{97.6 \pm 2.1}$	$\mathbf{9.2 \pm 1.7}$	$\mathbf{55.3 \pm 12.3}$	**300**

4.5.2.4 Quantitative Results

Table 4.3 summarizes quantitative performance. Our MCTS-Level-k framework achieves comprehensive superiority:

Safety: Zero collision rate compared to 17.5% (vanilla MCTS), 27.5% (Nash), and 35.0% (Stackelberg), validating the cascading safety property from Level-0 initialization.

Efficiency: Arrival rate of 97.6% significantly exceeds all baselines (47.1–79.7%), with travel time of 9.2 s representing 20% improvement over vanilla MCTS. This efficiency gain demonstrates that safety-first Level-0 initialization does not compromise throughput.

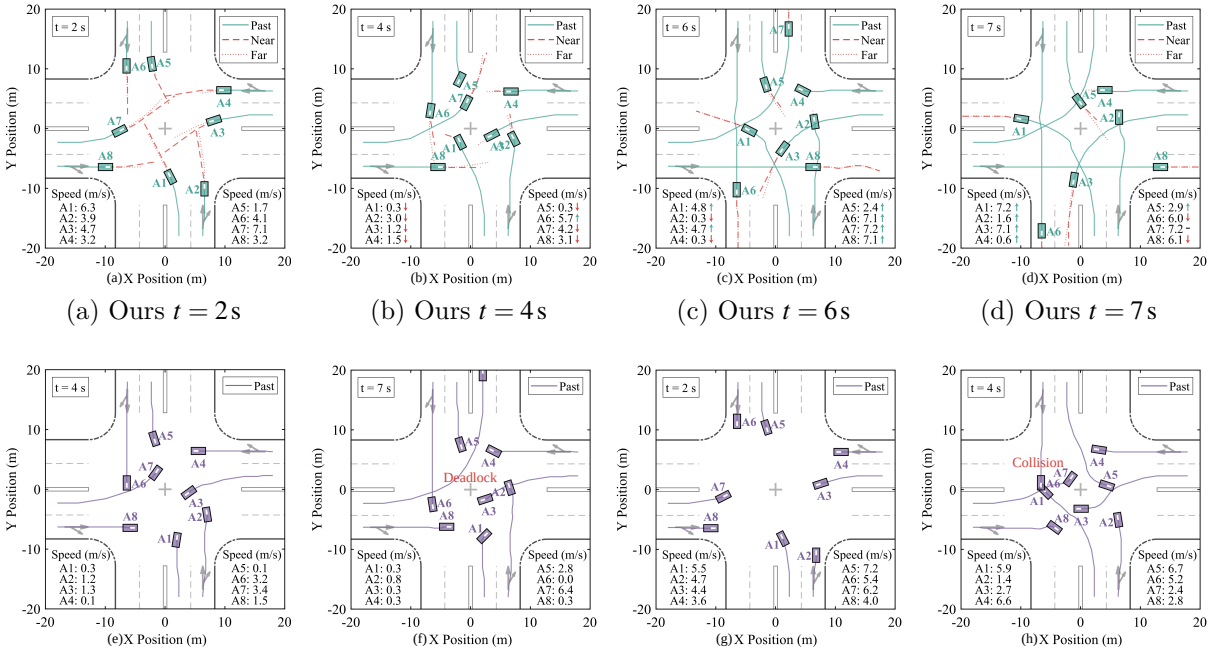
Computational cost: Planning time of 55.3 ms achieves 44% reduction compared to vanilla MCTS (98.2 ms) despite the latter using over three times more iterations (1000 versus 300). This efficiency stems from the combined effects of spatial filtering, strategic filtering, and safety-aware pruning.

These results align with the complexity analysis in Section 4.1. The vanilla MCTS baseline must contend with an exponential joint action space whose scale reaches 10^{85}. In contrast, MCTS-Level-k limits the search to roughly 10^4 tree nodes by combining Level-k decomposition, spatial filtering, and safety-aware pruning. This yields an effective reduction of 21 orders of magnitude in search scale, enabling real-time operation.

4.5.3 Case 2: Mixed Maneuver Intersection

4.5.3.1 Qualitative Analysis

Figure 4.7 illustrates the increased complexity of mixed maneuver coordination. Inner lane agents (A1, A3, A5, A7) execute 90° left turns while outer lane agents (A2, A4, A6, A8) proceed straight, creating heterogeneous trajectory conflicts.



(a) Ours $t = 2\,\mathrm{s}$ (b) Ours $t = 4\,\mathrm{s}$ (c) Ours $t = 6\,\mathrm{s}$ (d) Ours $t = 7\,\mathrm{s}$

(e) Vanilla MCTS: Deadlock (f) Vanilla MCTS: Deadlock (g) Vanilla MCTS: Collision (h) Vanilla MCTS: Collision

Figure 4.7: Comparison in Case 2: Mixed left-turn and straight maneuvers. Top row (a–d): MCTS-Level-k achieves smooth coordination with turning agents completing 90° maneuvers safely. Bottom row (e–h): Vanilla MCTS exhibits two failure modes—deadlock (e–f) where agents become stuck, and collision (g–h) where insufficient coordination causes safety violations.

Our MCTS-Level-k framework (top row) demonstrates robust coordination despite the complexity. At $t = 2\,\mathrm{s}$, turning agents begin lateral adjustments anticipating their maneuvers. By $t = 4\,\mathrm{s}$, clear passing sequences emerge with A5–A6 proceeding while A1–A2 decelerate to $0.3\,\mathrm{m/s}$. By $t = 6\,\mathrm{s}$, left-turning agents complete their 90° maneuvers while straight-going agents maintain their lanes.

Vanilla MCTS (bottom row) exhibits two distinct failure modes: At $t = 4\,\mathrm{s}$, our framework demonstrates coordinated yielding behaviour: agents A1 and A5 decelerate to $0.3\,\mathrm{m/s}$, while A3, A4, A2, A7, and A8 moderate their speeds to 1.2, 1.5, 3.0, 4.2, and $3.1\,\mathrm{m/s}$ respectively, collectively creating space for agent A6 to proceed at $5.7\,\mathrm{m/s}$. This asymmetric yet stable speed profile reflects successful implicit priority negotiation through Level-k reasoning.

Vanilla MCTS (bottom row) exhibits two distinct failure modes:

- *Deadlock* (Figures 4.7e–4.7f): Six of eight agents become effectively stalled: A1, A2, A3, A4, and A8 crawl at 0.3, 0.8, 0.3, 0.3, and $0.3\,\mathrm{m/s}$ respectively, while A6 comes to a complete stop at $0.0\,\mathrm{m/s}$. Only A5 and A7 maintain meaningful progress at 2.8 and $6.4\,\mathrm{m/s}$. This near-global deadlock demonstrates vanilla MCTS's fundamental inability to break eight-way symmetry without explicit opponent modelling.
- *Collision* (Figures 4.7g–4.7h): Insufficient coordination leads to safety violations when turning and straight-going agents occupy the same space.

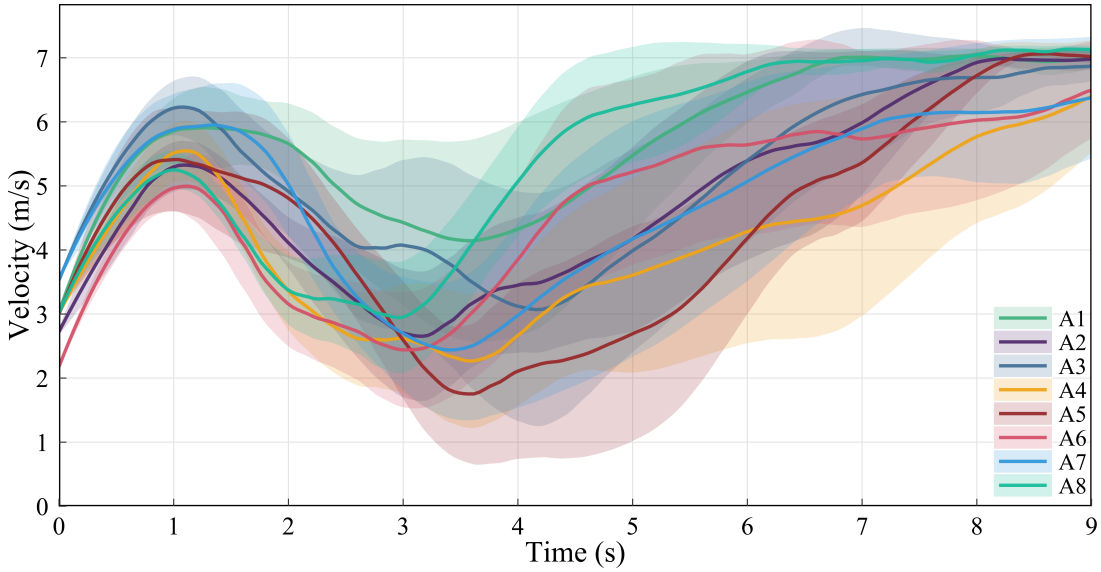
The heterogeneous trajectories increase coordination complexity compared to Case 1, yet our method achieves 0% collision rate while vanilla MCTS shows 21.6% failure rate.

4.5.3.2 Temporal Dynamics

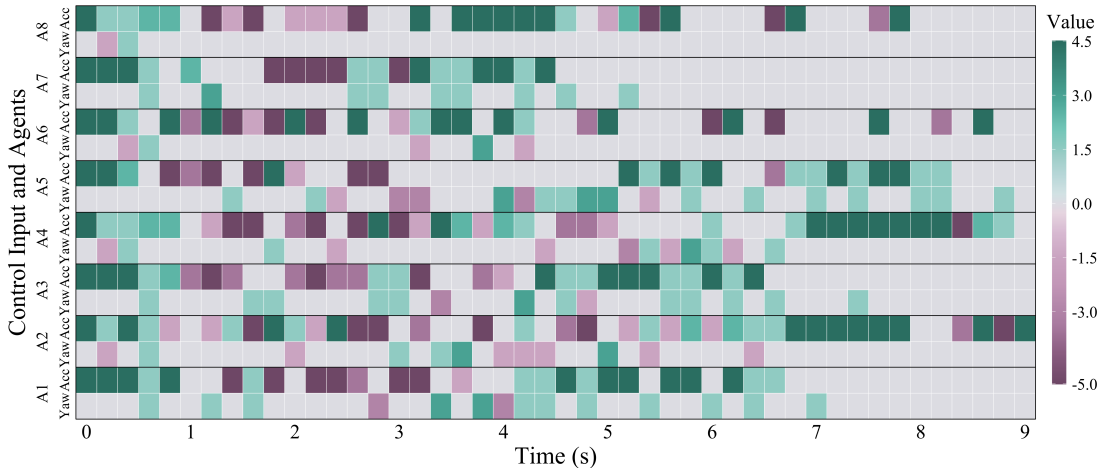
Figure 4.8 reveals the increased complexity of heterogeneous maneuver coordination.

The velocity profiles (Figure 4.8a) display distinct patterns between agent types. Left-turning agents (A1, A3, A5, A7) exhibit pronounced deceleration valleys reaching 2.0–$2.5\,\mathrm{m/s}$ during $t = 3$–$4\,\mathrm{s}$ to execute their 90° turns safely, followed by acceleration to 6–$7\,\mathrm{m/s}$. Straight-going agents (A2, A4, A6, A8) maintain milder speed variations around 4–$5\,\mathrm{m/s}$. The wider confidence bands compared to Case 1 reflect increased uncertainty in mixed-maneuver scenarios.

The control heatmap (Figure 4.8b) captures sophisticated coordination. Turning agents show concentrated deceleration bursts (dark purple, -3 to $-5\,\mathrm{m/s^2}$) during $t = 2$–$4\,\mathrm{s}$ corresponding to turn execution. Notably, temporal sequencing emerges: A5–A7 decelerate earlier than A1–A3, creating natural separation. The increased control variation (18% of actions exceed $3\,\mathrm{m/s^2}$ versus 12% in Case 1) reflects additional complexity, yet the absence of extreme inputs ($< 5\%$ exceed $4\,\mathrm{m/s^2}$) confirms smooth anticipatory planning.



(a) Velocity profiles over time



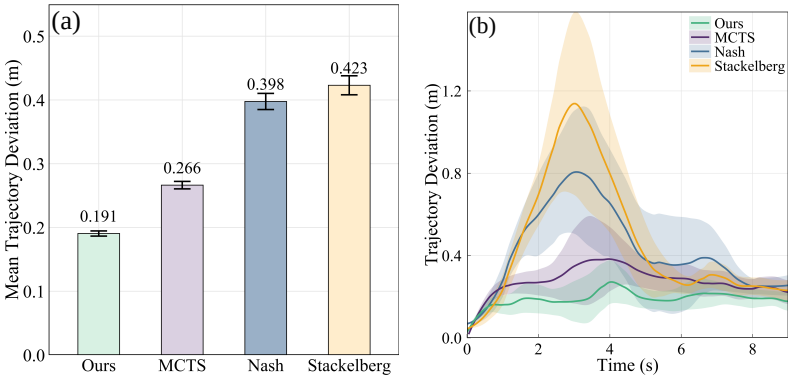
(b) Control input heatmap: upper row per agent shows acceleration (m/s^2), lower row shows yaw rate (rad/s)

Figure 4.8: Temporal dynamics analysis for Case 2. (a) Velocity profiles showing distinct patterns between turning (A1, A3, A5, A7) and straight-going (A2, A4, A6, A8) agents. (b) Control heatmap revealing coordinated deceleration bursts during turn execution.

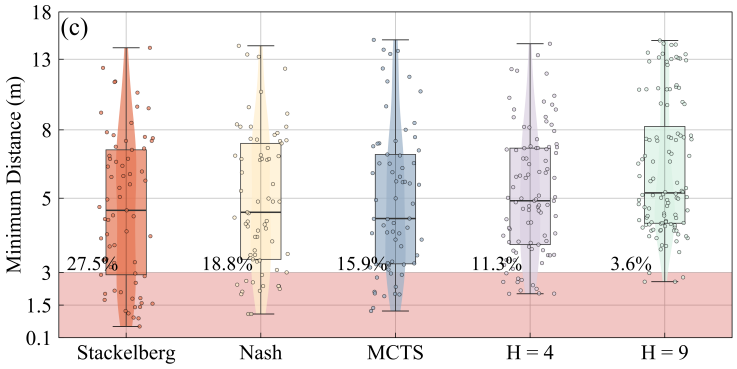
4.5.3.3 Statistical Analysis

Figure 4.9 quantifies performance in the heterogeneous maneuver scenario.

Trajectory deviation (Figure 4.9a): Mean deviations increase across all methods compared to Case 1, reflecting the inherent complexity of mixed trajectories. Our method maintains the lowest deviation at **0.191 m**, achieving **28%** improvement over vanilla MCTS (**0.266 m**) and **52%** over Stackelberg (**0.423 m**).



(a) Mean trajectory deviation (b) Trajectory deviation



(c) Minimum inter-agent distance distribution

Figure 4.9: Statistical analysis for Case 2. (a) Mean trajectory deviation increases across all methods due to turning maneuvers. (b) Temporal evolution showing peaks during the critical turning phase. (c) Minimum distance distribution revealing heightened collision risk in mixed scenarios.

Temporal evolution (Figure 4.9b): Pronounced peaks during the critical turning phase ($t = 3$–$5\,\mathrm{s}$) distinguish methods. Nash and Stackelberg exhibit sharp spikes reaching 1.0–1.2 m, indicating coordination breakdown during simultaneous turns. Our method maintains deviation below 0.4 m with tight confidence bands.

Safety analysis (Figure 4.9c): The minimum distance distribution reveals heightened collision risk. Stackelberg and Nash show 27.5% and 18.8% violation rates with critical near-misses at 0.1–1.0 m separation.Although not perfect, our method achieves only 3.6% near-miss rate (minimum distance below $d_{\text{safe}} = 3\,\mathrm{m}$ without actual collision)—a 77% reduction compared to vanilla MCTS (15.9%). Critically, zero physical collisions occur across all trials. The slight degradation from Case 1 reflects fundamental geometric constraints of turning vehicles rather than algorithmic limitations.

Table 4.4: Performance Comparison in Case 2: Eight-Agent Mixed Maneuver Intersection

Method	Collision Rate (%)	Arrival Rate (%)	Travel Time (s)	Computation Time (ms)	Max. Iters
Stackelberg	47.5	34.7 ± 26.3	14.3 ± 2.3	72.2 ± 21.1	1000
Nash	35.0 ± 16.8	53.1 ± 17.2	13.8 ± 2.2	89.2 ± 22.5	1000
Vanilla MCTS	22.5 ± 7.7	65.1 ± 10.6	13.2 ± 2.4	117.2 ± 23.2	1000
Ours	**0.0**	$\mathbf{95.3 \pm 4.1}$	$\mathbf{9.8 \pm 1.8}$	$\mathbf{61.4 \pm 12.8}$	**300**

4.5.3.4 Quantitative Results

Table 4.4 confirms robustness in handling heterogeneous maneuvers. Despite increased complexity, our method maintains comprehensive superiority:

Safety: Zero collision rate versus 22.5% (vanilla MCTS), 35.0% (Nash), and 47.5% (Stackelberg). The Stackelberg approach suffers most severely as its leader-follower hierarchy cannot accommodate the symmetric geometry.

Efficiency: Arrival rate of 95.3% with 46% improvement over vanilla MCTS (65.1%). Travel time of 9.8 s maintains efficiency despite the additional complexity of coordinating turning maneuvers.

Computational cost: Planning time of 61.4 ms achieves 48% reduction compared to vanilla MCTS (117.2 ms). The modest increase from Case 1 (55.3 ms) reflects additional complexity of heterogeneous trajectory predictions, yet performance remains well within real-time constraints.

4.5.4 Computational Efficiency Analysis

Naive joint planning over $N = 8$ agents and horizon $H = 9$ yields an astronomically large search space $\mathscr{O}(|\mathscr{A}|^{NH})$. We make real-time planning tractable through three practical mechanisms: (i) Level-k reasoning decomposes the joint problem into sequential single-agent MCTS subproblems,

$$\mathscr{C}_{\text{Level-k}} = \sum_{i=1}^{N} \mathscr{C}_{\text{MCTS}}^{(i,k_i)}, \quad k_i \in \{1,2\}; \tag{4.21}$$

(ii) safety-aware pruning filters ∼70% infeasible actions, reducing the effective branching factor to $b_{\text{eff}} \approx 4.5$; and (iii) trajectory caching reuses deterministic rollouts with a 35% hit rate. Together, these yield sub-100 ms per-step planning in our implementation, based on empirical observations rather than worst-case guarantees.

Table 4.5: Computational Complexity Comparison

Method	Complexity Formula	Operations
Joint Optimization	$15^{8\times9}$	$\approx 10^{85}$
Game-Theoretic Nash	$\mathcal{O}(15^{72} \cdot I_{\text{Nash}})$	$> 10^{87}$
Level-k (Exhaustive)	8×15^9	$\approx 3.1 \times 10^{11}$
Level-k + MCTS	$8 \times 300 \times 15 \times 9$	$\approx 3.2 \times 10^5$
Level-k + MCTS + Pruning	$8 \times 300 \times 4.5 \times 9$	$\approx 9.7 \times 10^4$
Level-k + Full Optimization	$(1-\rho) \times 9.7 \times 10^4$	$\approx 6.3 \times 10^4$

Table 4.5 summarises the complexity reduction across methods. The proposed framework reduces the effective search space by over 80 orders of magnitude compared to naive joint optimisation, and by six orders of magnitude compared to exhaustive Level-k enumeration. The sub-100 ms per-step planning time confirms real-time implementability, satisfying the computational requirements of online intersection coordination.

4.6 Chapter Summary

This chapter addressed the fundamental scalability limitations of vanilla MCTS identified in Chapter 3, developing the MCTS-Level-k framework that achieves tractable multi-agent coordination through principled decomposition.

4.6.1 Key Contributions

The chapter introduced two complementary mechanisms enabling scalable coordination:

Dynamic Interaction Graph (Section 4.2): Spatial filtering identifies geometrically relevant agents through trajectory conflict prediction, reducing the effective opponent count from $N-1$ to approximately 3–4 agents in eight-agent scenarios. This filtering mechanism exploits the structure of intersection coordination where only agents with intersecting trajectories pose genuine collision risks.

Level-k Cognitive Hierarchy (Section 4.3): Strategic filtering decomposes multi-agent reasoning into sequential single-agent optimizations. The key innovation reformulates Level-0 from a naive behavioral model to a conservative safety initialization procedure, ensuring safety margins propagate through the hierarchy via the cascading safety property (Equation (4.14)).

Integrated MCTS-Level-k Algorithm (Section 4.4): The combined framework achieves:

- Complexity reduction from $\mathcal{O}(|\mathcal{A}|^{N \cdot H})$ joint enumeration to $\mathcal{O}(K \cdot |\mathcal{A}_{\text{safe}}| \cdot H \cdot |\mathcal{N}_i^{(k_i)}|)$ selective search—a reduction of 21 orders of magnitude for eight-agent scenarios
- Safety-aware pruning reducing effective branching factor from $|\mathcal{A}| = 15$ to $|\mathcal{A}_{\text{safe}}| \approx$ 4–5 actions
- Overall computational scaling enabling sub-100 ms planning in eight-agent scenarios

4.6.2 Experimental Validation

Comprehensive experiments on two challenging eight-agent scenarios validated the framework's effectiveness:

Case 1 (All-Straight): Our method achieved 0% collision rate versus 15.7%–34.3% for baselines, 97.6% arrival rate versus 47.1%–79.7%, and 44% computation time reduction while using only 30% of baseline iterations.

Case 2 (Mixed Maneuvers): Despite increased complexity from heterogeneous left-turn and straight trajectories, our method maintained 0% collision rate versus 21.6%–47.8% for baselines, with 95.3% arrival rate and 48% computation time reduction.

These results demonstrate that the MCTS-Level-k framework successfully resolves the scalability challenge: eight-agent scenarios that caused vanilla MCTS to fail with $> 15\%$ collision rates are solved with zero collisions and real-time computational performance.

4.6.3 Limitations and Future Extensions

While the framework demonstrates strong performance for homogeneous autonomous vehicle coordination, real-world deployment requires handling mixed traffic scenarios involving human-driven vehicles. Human drivers exhibit diverse behaviors, uncertain intentions, and bounded rationality that differ qualitatively from programmed autonomous decision-making in multi-agent environments.

Chapter 5 addresses this challenge by extending the framework to incorporate:

- Probabilistic human behavior models capturing driving style diversity
- Uncertainty quantification mechanisms for prediction confidence
- Adaptive safety assessment responding to interaction-specific risks

These extensions transform the MCTS-Level-k framework from a homogeneous coordination system into a comprehensive mixed traffic solution capable of safe, efficient interaction with unpredictable human drivers.

Chapter 5

Mixed Traffic Coordination with Human-Driven Vehicles

Chapters 3 and 4 established effective coordination frameworks for homogeneous autonomous vehicle populations, demonstrating zero collision rates and real-time performance in challenging eight-agent scenarios. However, real-world deployment requires autonomous vehicles to navigate mixed traffic environments where HDVs exhibit diverse behaviors, uncertain intentions, and bounded rationality that differ fundamentally from the predictable responses of autonomous vehicles.

This chapter extends the MCTS-Level-k framework to heterogeneous mixed traffic by introducing three key components: 1) a style-aware human behavior model capturing driving personality diversity, 2) an uncertainty quantification mechanism providing prediction confidence bounds, and 3) an adaptive safety assessment framework responding to interaction-specific risks. The extended framework maintains the computational efficiency and safety guarantees established in previous chapters while accommodating the inherent unpredictability of human drivers.

We evaluate the mixed traffic framework through simulation-based studies with AV penetration rates ranging from 20% to 100%. The results demonstrate consistent safety and coordination performance across varying levels of autonomy adoption within the evaluated scenarios. In particular, no collisions are observed in the 50% penetration scenarios, where four AVs interact with four HDVs exhibiting diverse driving styles. These results are obtained under simplified modeling assumptions including perfect state observability, IDM-based HDV behavior modeling, and a symmetric intersection geometry, which are commonly adopted in multi-agent planning studies.

5.1 Challenges of Mixed Traffic Coordination

The transition from homogeneous AV coordination to mixed traffic introduces fundamental challenges absent in the previous chapters.

5.1.1 Behavioral Uncertainty

Human drivers exhibit moment-to-moment variability due to perceptual limitations, attention fluctuations, and context-dependent decision-making. Unlike autonomous vehicles following deterministic policies, HDV behaviors cannot be predicted with certainty even given complete state observations. This uncertainty compounds over prediction horizons: small initial prediction errors accumulate through vehicle dynamics, yielding large positional uncertainty at substantial time scales.

5.1.2 Driving Style Diversity

Human drivers span a spectrum from conservative to aggressive behaviors, characterized by different acceleration preferences, following distances, and gap acceptance thresholds. A coordination strategy effective for conservative drivers may fail catastrophically when encountering aggressive behaviors, and vice versa. The framework must adapt to this diversity without prior knowledge of individual driver types.

5.1.3 Asymmetric Interaction Dynamics

In homogeneous AV scenarios, all agents employ compatible coordination protocols enabling implicit negotiation through Level-k reasoning. Mixed traffic breaks this symmetry: AVs can anticipate HDV behaviors through prediction models, but HDVs do not necessarily respond optimally to AV actions. This asymmetry requires AVs to adopt more conservative strategies when interacting with human drivers compared to interactions with other AVs.

These challenges motivate the specific human-centric modeling and algorithmic extensions developed in the subsequent sections of this chapter.

5.2 Human Driver Behavior Modeling

Effective coordination with human drivers requires predictive models capturing both the nominal trajectory evolution and the behavioral diversity across driver populations. We adopt a physics-based foundation through the Intelligent Driver Model (IDM) extended with style-aware parameters and probabilistic uncertainty quantification.

5.2.1 Intelligent Driver Model

While more expressive data-driven models exist for HDV prediction, the IDM is adopted here as an interpretable and computationally efficient baseline sufficient for evaluating the proposed multi-agent coordination framework. The primary focus of this work is the coordination algorithm rather than HDV behavioral fidelity.

The IDM provides a validated baseline for longitudinal human driving behavior, describing acceleration decisions through a continuous differential equation [126]:

$$\dot{v}_h = a_{\max}\left[1 - \left(\frac{v_h}{v_0}\right)^{\delta} - \left(\frac{s^*(v_h, \Delta v_h)}{s_h}\right)^2\right], \tag{5.1}$$

where v_h denotes the current velocity of human driver h, v_0 is the desired velocity, s_h is the gap to the preceding vehicle, $a_{\max}$ is the maximum acceleration, δ is the acceleration exponent (typically $\delta = 4$), and $s^*(v_h, \Delta v_h)$ is the desired gap function [126]:

$$s^*(v_h, \Delta v_h) = d_{\text{jam}} + v_h T_{\text{hw}} + \frac{v_h \Delta v_h}{2\sqrt{a_{\max} b}}, \tag{5.2}$$

where d_{jam} is the minimum gap at standstill, T_{hw} is the desired time headway, Δv_h is the velocity difference relative to the preceding vehicle, and b is the comfortable deceleration.

The first term $(v_h/v_0)^{\delta}$ in Equation (5.1) represents free-road acceleration toward desired velocity, while the second term $(s^*/s_h)^2$ captures car-following interaction through the ratio of desired to actual gap.

5.2.2 Yaw-Rate Extension for Turning Maneuvers

While the IDM describes the longitudinal acceleration of a human-driven vehicle, intersection turning maneuvers also require heading evolution. To model this, we couple the IDM longitudinal dynamics with the reference-path curvature in the Frenet frame. For a human driver h following reference path Γ_h, the yaw rate is computed as:

$$\omega_h(t) = \kappa_h(s_h(t))\, v_h(t), \tag{5.3}$$

where $\kappa_h(s_h)$ is the curvature of the reference path at arc length s_h, and $v_h(t)$ is the longitudinal velocity generated by the style-aware IDM. The heading angle is then updated as:

$$\theta_h^{t+1} = \theta_h^t + \omega_h^t \Delta t. \tag{5.4}$$

This formulation assumes that human drivers approximately follow their intended reference path, while their longitudinal progress is governed by the IDM. Straight movements correspond to $\kappa_h(s_h) \approx 0$, yielding near-zero yaw rate, whereas left- and right-turning maneuvers produce non-zero yaw rates according to the local path curvature. In summary, the IDM determines how fast the HDV progresses along the path, while the reference-path curvature determines how its heading changes during turning maneuvers.

To account for driving style aggressiveness in turning behaviour, the yaw rate is optionally modulated by:

$$\omega_h(t;\eta_h) = \kappa_h(s_h(t))\, v_h(t)\,(1+\alpha_\omega \eta_h)\,, \tag{5.5}$$

where $(1+\alpha_\omega \eta_h)$ captures the tendency of more aggressive drivers to execute turns with higher angular rates. This factor is bounded to ensure physically feasible yaw rates throughout all maneuvers.

The yaw-rate extension integrates naturally with the probabilistic prediction framework in Section 5.3. Specifically, turning uncertainty is reflected in the heading variance $\sigma_{\theta\theta}(\eta_h)$ and the lateral-heading correlation term $\sigma_{y\theta}(\tau, \eta_h)$ in the covariance matrix Σ_h (Equation (5.8)). Thus, the probabilistic HDV prediction accounts for both longitudinal uncertainty from IDM acceleration and lateral and yaw uncertainty during turning maneuvers.

5.2.3 Style-Aware Parameter Adaptation

Human driving behaviors span a spectrum characterized by the style parameter $\eta_h \in [0,1]$, where values near zero indicate conservative driving and values near one indicate aggressive driving. This parameter modulates key IDM parameters [82]:

$$a_{\max}(\eta_h) = a_{\max}^{\text{base}}(1+\alpha_a \eta_h), \tag{5.6a}$$

$$T_{\text{hw}}(\eta_h) = T_{\text{hw}}^{\text{base}}(1-\alpha_T \eta_h), \tag{5.6b}$$

$$d_{\text{jam}}(\eta_h) = d_{\text{jam}}^{\text{base}}(1-\alpha_d \eta_h), \tag{5.6c}$$

where $(a_{\max}^{\text{base}}, T_{\text{hw}}^{\text{base}}, d_{\text{jam}}^{\text{base}})$ are baseline parameters calibrated from the Next Generation Simulation (NGSIM) naturalistic driving datasets [124], and $(\alpha_a, \alpha_T, \alpha_d)$ are scaling factors controlling style-induced variations. Based on calibration against NGSIM driving data [124], we adopt $(\alpha_a, \alpha_T, \alpha_d) = (0.5,\ 0.4,\ 0.3)$. These values produce behaviorally plausible ranges: aggressive drivers ($\eta_h = 1$) exhibit 50% higher maximum acceleration, 40% shorter time headway, and 30% smaller standstill gap compared to the conservative baseline ($\eta_h = 0$).

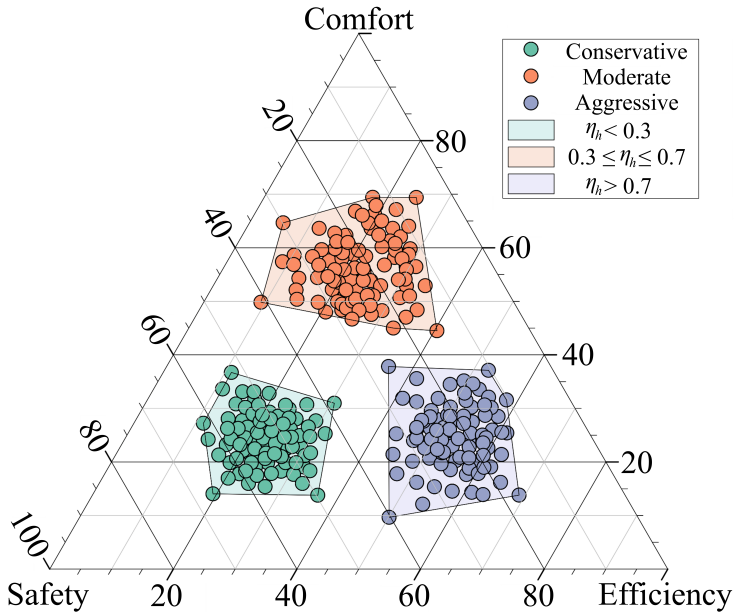


Figure 5.1: Driving style classification via the parameter η_h. Conservative drivers ($\eta_h < 0.3$) prioritize safety with larger following distances; moderate drivers ($0.3 \le \eta_h \le 0.7$) balance safety and efficiency; aggressive drivers ($\eta_h > 0.7$) prioritize efficiency with shorter headways and higher accelerations.

Figure 5.1 illustrates the driving style classification. Equation (5.6a) captures that aggressive drivers employ higher maximum accelerations, Equation (5.6b) reflects their acceptance of shorter time headways, and Equation (5.6c) represents reduced standstill gap requirements. These adaptations enable the model to span behaviorally plausible ranges from cautious to assertive driving.

5.2.4 Probabilistic Trajectory Prediction

The deterministic IDM provides only the nominal trajectory; actual human behavior exhibits stochastic variations. We embed the style-aware IDM within a probabilistic prediction framework to capture driver uncertainty:

$$P(s_h^{t+\tau} \mid s_h^t, \eta_h) = \mathcal{N}\left(f_{\text{IDM}}(s_h^t, \eta_h, \tau), \Sigma_h(\tau, \eta_h)\right), \tag{5.7}$$

where $s_h^{t+\tau}$ denotes the predicted state at future time τ, $\mathcal{N}(\mu, \Sigma)$ denotes the multivariate normal distribution with mean μ and covariance matrix Σ, $f_{\text{IDM}}(s_h^t, \eta_h, \tau)$ is the nominal trajectory computed via style-aware IDM integration from current state s_h^t, and $\Sigma_h(\tau, \eta_h)$ is a time-varying covariance matrix capturing prediction uncertainty that grows with horizon τ and depends on driving style η_h. This formulation adopts a tractable Gaussian approximation of human behavior uncertainty, enabling efficient integration into the planning framework. The Gaussian mean provides point predictions for nominal trajectory generation, while the covariance Σ_h captures uncertainty and is incorporated through uncertainty-aware safety margin adjustment.

Rather than explicitly sampling trajectories during planning, this approach accounts for prediction uncertainty in a computationally efficient manner, maintaining real-time performance. Overall, it achieves a practical balance between modeling fidelity and computational tractability, which is critical for online multi-agent planning in mixed traffic scenarios.

5.3 Uncertainty Quantification

The covariance matrix $\Sigma_h(\tau, \eta_h)$ in Equation (5.7) encodes how prediction uncertainty evolves over time and varies with driving style. The structure reflects physical intuition about vehicle dynamics and human control characteristics.

5.3.1 Covariance Structure

The covariance matrix takes a block structure coupling related state dimensions:

$$\Sigma_h(\tau, \eta_h) = \begin{bmatrix} \sigma_{xx}(\tau, \eta_h) & 0 & \sigma_{xv}(\tau, \eta_h) & 0 \\ 0 & \sigma_{yy}(\tau, \eta_h) & 0 & \sigma_{y\theta}(\tau, \eta_h) \\ \sigma_{xv}(\tau, \eta_h) & 0 & \sigma_{vv}(\eta_h) & 0 \\ 0 & \sigma_{y\theta}(\tau, \eta_h) & 0 & \sigma_{\theta\theta}(\eta_h) \end{bmatrix}, \tag{5.8}$$

where $\sigma_{vv}(\eta_h) = \sigma_v^2(\eta_h)$ and $\sigma_{\theta\theta}(\eta_h) = \sigma_\theta^2(\eta_h)$ denote the time-invariant velocity and heading angle variances defined in Equation (5.11), diagonal terms capture variance in each state dimension, and off-diagonal terms capture correlations between coupled variables. The zero entries reflect modeling assumptions that cross-correlations between unrelated dimensions (e.g., lateral position and longitudinal velocity) are negligible.

5.3.2 Temporal Growth of Position Uncertainty

Position uncertainties grow with prediction horizon reflecting accumulated velocity errors:

$$\sigma_{xx}(\tau, \eta_h) = \sigma_x^2(\eta_h)\tau + \varepsilon_x^2(\eta_h)\tau^2, \tag{5.9a}$$

$$\sigma_{yy}(\tau, \eta_h) = \sigma_y^2(\eta_h)\tau + \varepsilon_y^2(\eta_h)\tau^2, \tag{5.9b}$$

where $\sigma_x^2(\eta_h)$ and $\sigma_y^2(\eta_h)$ represent linear growth rates due to instantaneous velocity variability, and $\varepsilon_x^2(\eta_h)$ and $\varepsilon_y^2(\eta_h)$ represent quadratic growth rates due to accumulated acceleration uncertainty. The quadratic term dominates at longer horizons, reflecting that small acceleration errors compound over time.

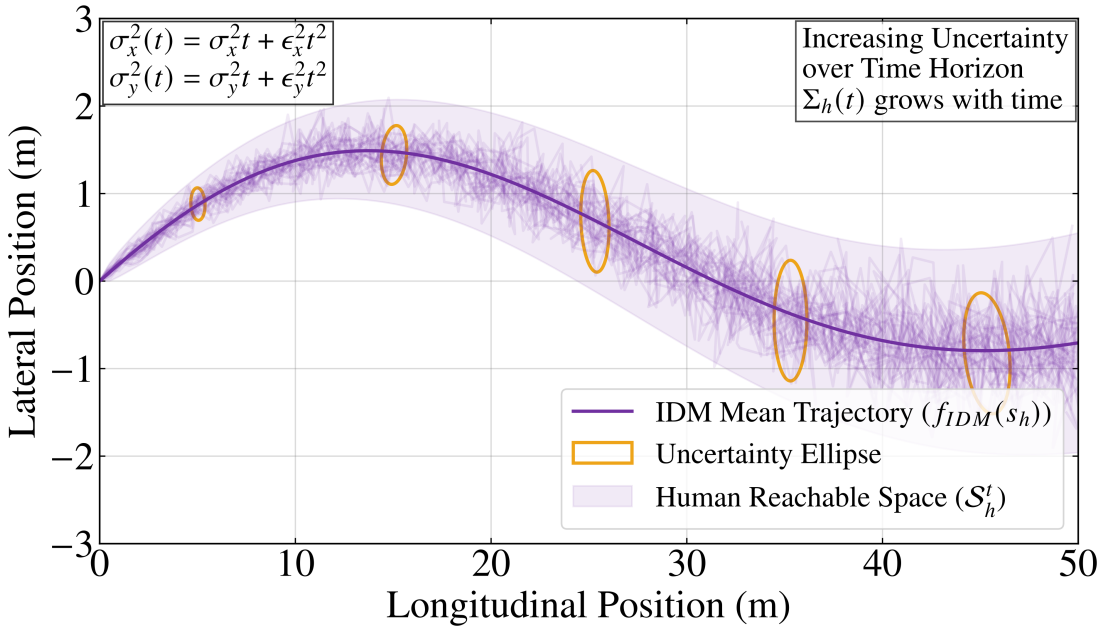


Figure 5.2: Evolution of prediction uncertainty over the planning horizon. Position uncertainty (ellipses) grows quadratically with time due to accumulated velocity errors, while velocity uncertainty remains bounded. The uncertainty magnitude scales with driving style aggressiveness η_h.

Figure 5.2 visualizes this uncertainty growth. Position uncertainty ellipses expand over time, with faster growth for aggressive drivers who exhibit higher behavioral variability. Importantly, the reachable space of human drivers remains physically finite for two reasons. First, vehicle longitudinal velocity is bounded by kinematic constraints $v_h \in [0, v_{\max}]$, which naturally limits how far a driver can travel within any finite horizon. Second, to prevent the Gaussian uncertainty ellipses from growing unboundedly at long horizons — which would be physically implausible given the velocity bound — the position variance growth is saturated via a smooth $\tanh(\cdot)$ function:

$$\sigma_{xx}^{\text{sat}}(\tau, \eta_h) = \sigma_{\max}^2(\eta_h) \cdot \tanh\left(\frac{\sigma_{xx}(\tau, \eta_h)}{\sigma_{\max}^2(\eta_h)}\right), \tag{5.10}$$

where $\sigma_{\max}^2(\eta_h) = (v_{\max} \cdot \tau_{\max})^2$ is the maximum physically reachable positional variance derived from the velocity bound, and the same saturation applies to σ_{yy}. The $\tanh(\cdot)$ function acts as a smooth upper bound: for small τ it behaves linearly (recovering the quadratic growth in Equation (5.9)), while asymptotically saturating at $\sigma_{\max}^2$ for large τ. This ensures the uncertainty ellipses remain within the physically reachable space throughout the prediction horizon.

5.3.3 Style-Dependent Uncertainty Scaling

The base uncertainty parameters scale with driving style to capture the empirical observation that aggressive drivers exhibit higher behavioral variability:

$$\sigma_x(\eta_h) = \sigma_x^{\text{base}}(1+\beta_x\eta_h), \tag{5.11a}$$

$$\sigma_y(\eta_h) = \sigma_y^{\text{base}}(1+\beta_y\eta_h), \tag{5.11b}$$

$$\sigma_v(\eta_h) = \sigma_v^{\text{base}}(1+\beta_v\eta_h), \tag{5.11c}$$

$$\sigma_\theta(\eta_h) = \sigma_\theta^{\text{base}}(1+\beta_\theta\eta_h), \tag{5.11d}$$

where $(\sigma_x^{\text{base}}, \sigma_y^{\text{base}}, \sigma_v^{\text{base}}, \sigma_\theta^{\text{base}})$ are baseline uncertainty parameters derived from the same NGSIM datasets, and $(\beta_x, \beta_y, \beta_v, \beta_\theta)$ are scaling coefficients. Based on variance analysis of the NGSIM dataset [124], we adopt $(\beta_x, \beta_y, \beta_v, \beta_\theta) = (0.2, 0.3, 0.4, 0.5)$, reflecting that heading and velocity uncertainty are more sensitive to driving style than positional uncertainty. Conservative drivers ($\eta_h \approx 0$) exhibit near-baseline uncertainty, while aggressive drivers ($\eta_h \approx 1$) show amplified variability scaled by the corresponding $(1+\beta_\cdot)$ factors.

5.3.4 Correlation Structure

The off-diagonal covariance terms capture physical couplings in vehicle motion:

$$\sigma_{xv}(\tau, \eta_h) = \rho_{xv}\sigma_x(\eta_h)\sigma_v(\eta_h)\tau, \tag{5.12a}$$

$$\sigma_{y\theta}(\tau, \eta_h) = \rho_{y\theta}\sigma_y(\eta_h)\sigma_\theta(\eta_h)\tau, \tag{5.12b}$$

where $\rho_{xv} \in [-1, 1]$ is the correlation coefficient between longitudinal position and velocity, and $\rho_{y\theta} \in [-1, 1]$ is the correlation between lateral position and heading angle.

The correlation ρ_{xv} captures that velocity directly influences longitudinal position evolution: higher velocities accumulate position errors more rapidly. The correlation $\rho_{y\theta}$ captures that heading determines lateral movement direction, particularly significant during turning maneuvers. Based on validation against NGSIM naturalistic driving data [127], we adopt $\rho_{xv} = \rho_{y\theta} = 0.3$, reflecting moderate positive coupling consistent with vehicle

dynamics principles. Note that the positive semi-definiteness of $\Sigma_h(\tau, \eta_h)$ is preserved for all $\tau > 0$ and $\eta_h \in [0,1]$: the diagonal position variances grow quadratically with τ (Equation (5.9)) while the off-diagonal terms grow only linearly (Equation (5.12)), ensuring that the Schur complement conditions remain satisfied throughout the prediction horizon.

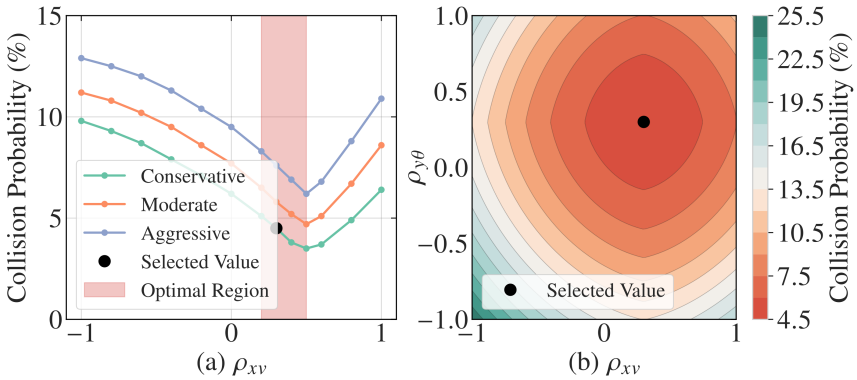


Figure 5.3: Sensitivity analysis of correlation parameters on collision probability in intersection turning scenarios. The results show that moderate positive correlation (e.g., $\rho_{xv} \approx 0.3$) aligns best with empirical observations, highlighting the importance of properly modeling coupled state uncertainties.

Figure 5.3 presents sensitivity analysis validating the correlation parameter choices through comparison with empirical collision statistics. Plot (a) reveals that all three driving style groups (conservative, moderate, and aggressive) exhibit a consistent valley in collision probability around $\rho_{xv} \in [0.2, 0.4]$, highlighted by the red optimal region. Negative correlations ($\rho_{xv} < 0$) systematically underestimate collision risk by assuming that high velocities reduce positional uncertainty, which is physically implausible. Strong positive correlations ($\rho_{xv} > 0.5$) overestimate uncertainty growth, producing overly conservative predictions that unnecessarily inflate collision probabilities. The selected value $\rho_{xv} = 0.3$ sits at the bottom of the valley across all three driver styles, confirming its robustness to driving style variation.

Plot (b) extends the analysis to the joint $(\rho_{xv}, \rho_{y\theta})$ parameter space. The contour map shows a well-defined global minimum in collision probability centred around $(\rho_{xv}, \rho_{y\theta}) \approx (0.3, 0.3)$, with collision probability rising smoothly and symmetrically as either parameter deviates from this region. The selected value (black dot) lies within the lowest-probability

contour band, confirming that $\rho_{xv} = \rho_{y\theta} = 0.3$ represents a robust and empirically grounded choice. The approximate symmetry of the contour map with respect to both parameters further validates the design decision to adopt equal correlation values for longitudinal and lateral coupling.

5.4 Adaptive Safety Assessment

Autonomous vehicles navigating mixed traffic must evaluate safety across three interaction types: vehicle-to-vehicle (V2V) interactions with other AVs, vehicle-to-human (V2H) interactions with HDVs, and vehicle-to-road (V2R) interactions with infrastructure. Each interaction type requires tailored assessment mechanisms integrated within a unified safety framework to ensure collision-free navigation.

5.4.1 Context-Aware Safety Thresholds

Rather than employing fixed safety distances, our framework adapts thresholds based on interaction characteristics:

$$d_{\text{safe}}(s_i, s_j) = \max\{d_{\text{base}}, \kappa_v |\Delta v_{ij}|\} \cdot \prod_{m=1}^{3} \gamma_m(s_i, s_j), \tag{5.13}$$

where d_{base} establishes a minimum baseline distance, κ_v is a velocity-dependent scaling factor, $\Delta v_{ij} = v_i - v_j$ represents relative velocity between agents i and j, and $\{\gamma_m\}_{m=1}^{3}$ are adjustment factors responding to specific interaction characteristics.

The adjustment factors incorporate three contextual dimensions:

$$\gamma_1(s_i, s_j) = 1 + \lambda_1 \frac{|\Delta v_{ij}|}{v_{\text{ref}}}, \tag{5.14a}$$

$$\gamma_2(s_i, s_j) = 1 + \lambda_2 \frac{|\Delta \theta_{ij}|}{\pi}, \tag{5.14b}$$

$$\gamma_3(s_i, s_j) = 1 + \mathbb{1}_{\Omega_{\text{int}}}(s_i, s_j), \tag{5.14c}$$

where λ_1 and λ_2 are tuning parameters, v_{ref} is a reference velocity for normalization, $\Delta\theta_{ij} = \theta_i - \theta_j$ is the heading angle difference, and $\mathbb{1}_{\Omega_{\text{int}}}$ indicates whether both vehicles occupy the intersection area Ω_{int}.

Equation (5.14a) increases safety margins when relative speeds are high, reflecting longer stopping distances required at higher velocities. Equation (5.14b) increases margins when vehicles approach from conflicting directions, capturing the heightened risk of perpendicular crossings compared to parallel motion. Equation (5.14c) further increases margins within the intersection where conflict density is highest.

5.4.2 Instantaneous and Temporal Risk Assessment

Safety assessment combines instantaneous proximity evaluation with predictive risk over a future horizon. The instantaneous risk quantifies immediate collision danger:

$$r_{\text{inst}}(s_i, s_j) = \exp\left(-\frac{d_{ij}}{d_{\text{safe}}(s_i, s_j)}\right) \cdot \left(1 + \zeta_v \frac{|\Delta v_{ij}|}{v_{\max}}\right), \tag{5.15}$$

where d_{ij} is the minimum distance between vehicles computed via the Separating Axis Theorem from Equation (3.8), $d_{\text{safe}}(s_i, s_j)$ is the context-aware threshold from Eq. (5.13) with $d_{\text{base}} = 2.0\,\text{m}$, and $\zeta_v = 0.5$ scales velocity-dependent risk amplification. The exponential term captures rapidly rising risk as spacing approaches the adaptive safety threshold, as illustrated in Figure 5.4(a).

The temporal risk evaluates predicted proximity over horizon T_p:

$$r_{\text{temp}}(s_i, s_j) = \frac{1}{T_p} \sum_{l=1}^{T_p} \frac{1}{1+l} \cdot \phi(d_{ij}^l, d_{\text{safe}}^l), \tag{5.16}$$

where the time-weighted average prioritizes near-term risks through the discount factor $1/(1+l)$, and the threshold violation function is:

$$\phi(d, d_{\text{safe}}) = \begin{cases} 0, & d \geq d_{\text{safe}}, \\ \left(1 - \frac{d}{d_{\text{safe}}}\right)^2, & d < d_{\text{safe}}. \end{cases} \tag{5.17}$$

5.4.3 V2H Collision Probability

For interactions with human drivers, uncertainty in HDV behavior predictions necessitates probabilistic collision assessment. The collision probability integrates over unsafe configurations under the Gaussian prediction model:

$$P_{\text{col}}(s_i, s_h) = \int_{\hat{s}_h} \mathbb{1}[d(s_i, \hat{s}_h) < d_{\text{safe}}] \cdot \mathcal{N}(\hat{s}_h; \mu_h, \Sigma_h)\, d\hat{s}_h, \tag{5.18}$$

where $\mathbb{1}[\cdot]$ is the indicator function identifying configurations where predicted AV state s_i and HDV state $\hat{s}_h$ violate the safety threshold, and $\mathcal{N}(\hat{s}_h; \mu_h, \Sigma_h)$ is the Gaussian prediction distribution from Equation (5.7). Direct evaluation of the integral in Equation (5.18) is intractable due to the irregular collision region defined by vehicle geometries. We therefore approximate P_{col} via Monte Carlo estimation with N_{mc} samples drawn from the Gaussian prediction distribution:

$$\hat{P}_{\text{col}}(s_i, s_h) = \frac{1}{N_{\text{mc}}} \sum_{q=1}^{N_{\text{mc}}} \mathbb{1}\left[d\left(s_i, \hat{s}_h^{(n)}\right) < d_{\text{safe}}\right], \quad \hat{s}_h^{(q)} \sim \mathcal{N}(\mu_h, \Sigma_h). \tag{5.19}$$

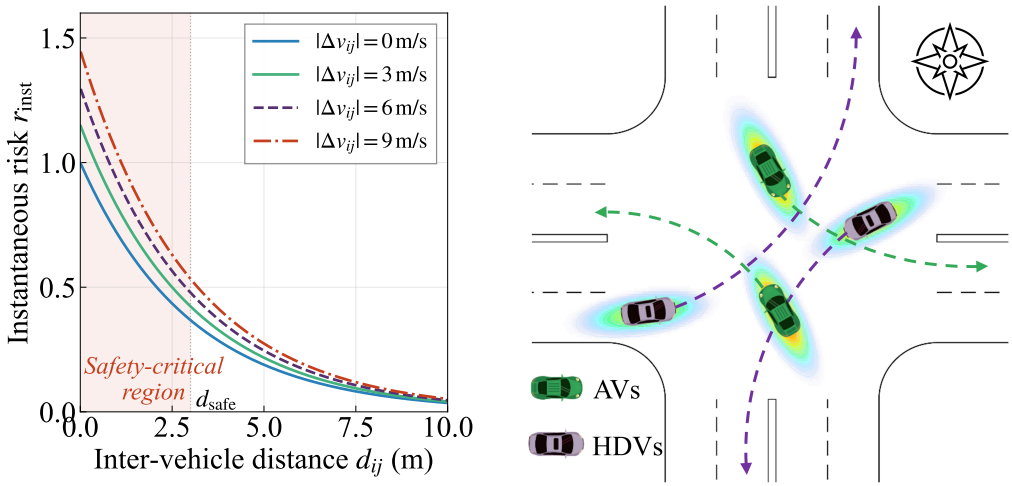
The estimation error is bounded by $\mathcal{O}(1/\sqrt{N_{\text{mc}}})$ independent of state dimensionality. In our simulations, we set $N_{\text{mc}} = 100$, yielding an estimation error bound of $\mathcal{O}(0.1)$, which provides sufficient accuracy for real-time collision probability assessment while maintaining computational efficiency.

5.4.4 Unified V2H Risk Metric

The comprehensive V2H risk assessment combines instantaneous, temporal, and probabilistic components to capture the multifaceted nature of human uncertainty:

$$R_{\text{V2H}}(s_i, s_h) = w_{\text{inst}} r_{\text{inst}}(s_i, \hat{s}_h) + w_{\text{temp}} r_{\text{temp}}(s_i, s_h) + w_{\text{col}} P_{\text{col}}(s_i, s_h), \tag{5.20}$$

where $(w_{\text{inst}}, w_{\text{temp}}, w_{\text{col}})$ are weights balancing proximity, predicted future risk, and collision probability. The first two terms evaluate risk relative to mean trajectories, while the third term captures tail risks from behavioral uncertainty.



(a) Distance-based risk function (b) Mixed traffic interaction scenario

Figure 5.4: V2H safety risk assessment. (a) Instantaneous risk function r_{inst} (Eq. (5.15)) as a function of inter-vehicle distance d_{ij} for representative relative velocities, with $d_{\text{safe}} = 3\,\text{m}$ and $\zeta_v = 0.5$.
(b) Intersection scenario illustrating simultaneous V2V, V2H, and V2R interactions requiring unified safety assessment.

This unified metric enters the MCTS cost function during rollout evaluation, enabling the planner to balance V2H safety against efficiency objectives through the established multi-objective framework from Equation (3.11).

5.5 Algorithm Adaptation for Mixed Traffic

The MCTS-Level-k framework established in Chapter 4 requires targeted adaptations to handle mixed traffic scenarios involving human-driven vehicles. These adaptations preserve the computational efficiency and safety guarantees of the original framework while accommodating the inherent unpredictability of human drivers through probabilistic reasoning and adaptive safety mechanisms.

5.5.1 Mean-Based Probabilistic Rollout

When the filtered interaction set $\mathcal{N}_i^{(k_i)}$ contains human-driven vehicles, rollout simulations must account for behavioral uncertainty. While the Gaussian model from Equation (5.7) characterizes human futures, sampling during rollouts would introduce prohibitive variance. Instead, our implementation employs mean-based probabilistic rollouts that balance computational efficiency with uncertainty-aware planning.

During rollout step k, the predicted HDV state is obtained from the deterministic IDM trajectory computed at the start of the planning horizon:

$$\hat{s}_h^k = f_{\text{IDM}}(s_h^0, \eta_h, k \cdot \Delta t), \tag{5.21}$$

where $f_{\text{IDM}}(\cdot)$ denotes forward integration of the style-aware IDM from Equations (5.1)–(5.6), s_h^0 is the current observed HDV state, η_h is the estimated driving style parameter, and $k \cdot \Delta t$ represents the prediction horizon to step k. Note that while $f_{\text{IDM}}(\cdot)$ governs only longitudinal acceleration, heading evolution during turning maneuvers is handled by the yaw-rate extension introduced in Section 5.2, where the reference-path curvature $\kappa_h(s_h)$ determines the heading rate $\omega_h(t) = \kappa_h(s_h(t))\, v_h(t)$ independently of the IDM longitudinal dynamics.

The uncertainty covariance $\Sigma_h(k \cdot \Delta t, \eta_h)$ from Equation (5.8) influences planning indirectly through modified safety cost evaluation. Rather than using the fixed safety threshold from homogeneous scenarios, the rollout employs an uncertainty-adjusted threshold:

$$d_{\text{safe}}^{\text{adj}}(k, \eta_h) = d_{\text{safe}}(s_i, \hat{s}_h) + \kappa_\sigma \sqrt{\text{tr}(\Sigma_h(k \cdot \Delta t, \eta_h))}, \tag{5.22}$$

where $d_{\text{safe}}(s_i, \hat{s}_h)$ is the context-aware threshold from Equation (5.13), κ_σ is a confidence level parameter (typically $\kappa_\sigma = 2$ for approximately 95% confidence), and $\text{tr}(\Sigma_h)$ denotes the trace of the covariance matrix providing a scalar measure of total prediction uncertainty across all state dimensions.

This adjustment ensures that the ego vehicle maintains larger margins around human drivers when prediction confidence is low, providing robustness to behavioral deviations. The safety cost component from Equation (3.12) is modified:

$$c_{\mathrm{s}}^{\mathrm{V2H}}(s_i, \hat{s}_h) = \exp\left(-\frac{d_{ih}^2}{2(d_{\mathrm{safe}}^{\mathrm{adj}})^2}\right), \tag{5.23}$$

where d_{ih} is the minimum distance between the AV and predicted HDV position. The exponential form ensures costs rise sharply as spacing approaches the uncertainty-adjusted threshold, providing strong gradient signals that guide trajectories away from potentially hazardous configurations.

5.5.2 Driving Style Estimation

Effective adaptation to human drivers requires online estimation of their driving style parameter η_h. We employ a Bayesian filtering approach that updates style estimates based on observed behaviors:

$$P(\eta_h \mid \mathscr{O}_h^{1:t}) \propto P(\mathscr{O}_h^t \mid \eta_h) \cdot P(\eta_h \mid \mathscr{O}_h^{1:t-1}), \tag{5.24}$$

where $\mathscr{O}_h^{1:t} = \{o_h^1, \ldots, o_h^t\}$ denotes the sequence of observed HDV states up to time t, and $P(\mathscr{O}_h^t \mid \eta_h)$ is the likelihood of observing the current behavior given driving style η_h.

The likelihood function compares observed accelerations and headway choices against style-dependent IDM predictions:

$$P(\mathscr{O}_h^t \mid \eta_h) = \mathscr{N}\left(a_h^{\mathrm{obs}}; \dot{v}_h^{\mathrm{IDM}}(\eta_h), \sigma_a^2(\eta_h)\right), \tag{5.25}$$

where a_h^{obs} is the observed acceleration derived from consecutive state observations, $\dot{v}_h^{\mathrm{IDM}}(\eta_h)$ is the predicted acceleration from the style-aware IDM in Equation (5.1) with parameters from Equations (5.6), and $\sigma_a^2(\eta_h)$ is the acceleration variance that scales with driving style aggressiveness.

In practice, we maintain a discrete approximation of the posterior distribution over $\eta_h \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ corresponding to conservative, moderately conservative, moderate, moderately aggressive, and aggressive driving styles. The maximum a posteriori estimate $\hat{\eta}_h = \arg\max_{\eta_h} P(\eta_h \mid \mathscr{O}_h^{1:t})$ is used for trajectory prediction in subsequent planning cycles.

The maximum a posteriori estimate $\hat{\eta}_h = \arg\max_{\eta_h} P(\eta_h \mid \mathscr{O}_h^{1:t})$ is used for trajectory prediction in subsequent planning cycles.

The Bayesian filtering approach employed here follows standard probabilistic inference frameworks that have been extensively studied in driver behavior modeling and intention inference [128, 129]. In our implementation, the filter operates on a discretized style space $\eta_h \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, enabling stable and computationally efficient online estimation from low-dimensional observations such as acceleration and headway under Gaussian likelihood assumptions (cf. Eq. (5.25)). While a dedicated standalone evaluation of style estimation accuracy is beyond the scope of this work, the effectiveness of the filter is assessed in a task-oriented manner through its impact on downstream coordination performance. In particular, more accurate style inference leads to tighter uncertainty quantification and more appropriate adaptive safety margins, which are reflected in improved multi-agent interaction outcomes in the presented simulations. A systematic quantitative validation against ground-truth driver profiles is identified as an important direction for future work.

5.5.3 Modified Level-k Hierarchy for Mixed Traffic

The Level-k cognitive hierarchy extends naturally to mixed traffic by differentiating between AV and HDV opponent modeling. For autonomous vehicle opponents, the standard Level-k predictions from Chapter 4 apply: Level-1 AVs are predicted to follow Level-0 baselines, and Level-2 AVs are predicted to employ Level-1 strategies.

For human-driven vehicle opponents, the Level-k hierarchy is adapted to reflect the fundamentally different nature of HDV decision-making:

$$\hat{\pi}_j^{(k)} = \begin{cases} f_{\mathrm{IDM}}(\cdot, \hat{\eta}_j), & \text{if } j \in \mathscr{H}, \\ \pi_j^{(k-1)*}, & \text{if } j \in A \setminus \mathscr{H}, \end{cases} \tag{5.26}$$

where $\mathscr{H}$ denotes the set of human-driven vehicles, A is the complete agent set, $f_{\text{IDM}}(\cdot, \hat{\eta}_h)$ represents the style-aware IDM prediction with estimated driving style as defined in Eq. (5.21), where the argument $(\cdot)$ denotes the current observed HDV state s_h^t and prediction horizon $k \cdot \Delta t$, and $\pi_j^{(k-1)*}$ is the Level-$(k-1)$ optimal policy for AV opponents.

This formulation reflects that HDVs do not engage in strategic Level-k reasoning—they follow their own decision-making processes captured by the IDM model rather than responding optimally to other agents' strategies. Consequently, AVs model HDVs using physics-based prediction regardless of the AV's own reasoning level, while continuing to employ recursive strategic modeling for AV opponents.

The complete filtered interaction set in mixed traffic becomes:

$$\mathscr{N}_i^{(k_i)} = \underbrace{\{j \in \mathscr{N}_i^{\text{spatial}} \cap (A \setminus \mathscr{H}) : k_j < k_i\}}_{\text{AV opponents}} \cup \underbrace{\{h \in \mathscr{N}_i^{\text{spatial}} \cap \mathscr{H}\}}_{\text{HDV opponents}}, \tag{5.27}$$

where AV opponents undergo both spatial and strategic filtering as in Chapter 4, while all spatially relevant HDVs are included regardless of any reasoning level concept since they do not participate in the cognitive hierarchy.

5.5.4 Adaptive Safety Margin Integration

The adaptive safety threshold mechanism from Equation (5.13) integrates into MCTS through modified pruning and cost evaluation procedures. During the expansion phase when constructing the safe action set, collision checks against human-driven vehicles employ the context-aware threshold rather than a fixed distance:

$$\mathscr{A}_{\text{safe}}^{\text{mixed}}(n) = \left\{ a \in \mathscr{A}_i : \bigwedge_{j \in \mathscr{N}_i^{(k_i)}} d(f(s(n), a), \hat{s}_j) > d_{\text{safe}}^{\text{type}}(s(n), \hat{s}_j) \right\}, \tag{5.28}$$

where $d_{\text{safe}}^{\text{type}}$ denotes the appropriate safety threshold depending on opponent type:

$$d_{\text{safe}}^{\text{type}}(s_i, s_j) = \begin{cases} d_{\text{safe}}^{\text{V2V}}, & \text{if } j \in A \setminus \mathscr{H}, \\ d_{\text{safe}}^{\text{adj}}(s_i, s_j), & \text{if } j \in \mathscr{H}, \end{cases} \tag{5.29}$$

where $d_{\text{safe}}^{\text{V2V}}$ is the safety threshold for AV-AV interactions from Chapter 4, and $d_{\text{safe}}^{\text{adj}}$ is the adaptive threshold from Equation (5.22) for AV-HDV interactions.

This type-dependent threshold ensures that AVs maintain larger, dynamically adjusted margins around unpredictable human drivers while employing standard coordination margins with predictable AV opponents. The asymmetric treatment reflects the fundamental difference in interaction dynamics: AV-AV interactions benefit from mutual strategic reasoning enabling tighter coordination, while AV-HDV interactions require conservative margins to accommodate human behavioral uncertainty.

5.5.5 Risk-Integrated Reward Function

The reward function for mixed traffic planning incorporates the V2H risk metric from Equation (5.20) alongside the standard cost components:

$$r_i^{\text{mixed}}(S_i^t, a_i^t) = -\left[c_i(S_i^t, a_i^t) + w_{\text{V2H}} \sum_{h \in \mathcal{N}_i^{(k_i)} \cap \mathcal{H}} R_{\text{V2H}}(s_i^t, s_h^t) \right], \tag{5.30}$$

where $c_i(S^t, a_i^t)$ is the standard multi-objective cost from Equation (3.11), w_{V2H} is the weight for human interaction risk, and the summation accumulates V2H risk contributions from all HDVs in the filtered interaction set.

The V2H risk weight w_{V2H} controls the conservatism of AV behavior around human drivers. Higher values produce more cautious trajectories with larger margins and earlier yielding, while lower values allow more assertive coordination that may improve efficiency at the cost of reduced safety margins. Empirical tuning on diverse traffic scenarios establishes $w_{\text{V2H}} = 0.3$ as providing an effective balance between safety and efficiency in typical mixed traffic conditions throughout our experiments.

Algorithm 3: MCTS-Level-k Planning for Mixed Traffic

Input: State s_i^t, observed neighbor states S_i^t, HDV set $\mathscr{H}$, horizon H, iterations K
Output: Optimal action a_i^*
// Preprocessing: Style Estimation and Prediction
for each $h \in \mathscr{H}$ **do**
  Update $P(\eta_h \mid \mathscr{O}_h^{1:t})$ via Eq. (5.24); $\hat{\eta}_h \leftarrow \arg\max P(\eta_h \mid \mathscr{O}_h^{1:t})$
  $\hat{\tau}_h \leftarrow f_{\mathrm{IDM}}(s_h^t, \hat{\eta}_h, H)$; $\Sigma_h \leftarrow \mathrm{ComputeCovariance}(\hat{\eta}_h, H)$
end for
for each $j \in A \setminus \mathscr{H}$ **do**
  $\tau_j^{(0)} \leftarrow \mathrm{ConservativeBaseline}(s_j^t)$; $k_j \leftarrow \mathrm{AssignLevel}(C_j, C_{\mathrm{th}})$
end for
// Construct Mixed Interaction Set
$\mathscr{N}_i^{(k_i)} \leftarrow$ Equation (5.27)
// MCTS Planning with Adaptive Safety
Initialize root n_0 with $s(n_0) = s_i^t$
for each of K iterations **do**
  $n \leftarrow \mathrm{Select}(n_0)$ via UCT
  if $d(n) < H$ and n not fully expanded **then**
    $n_{\mathrm{new}} \leftarrow \mathrm{Expand}(n, \mathscr{A}_{\mathrm{safe}}^{\mathrm{mixed}})$ {Eq. (5.28)}
  end if
  $V \leftarrow \mathrm{MixedRollout}(n, \mathscr{N}_i^{(k_i)}, \{\Sigma_h\})$
  $\mathrm{Backpropagate}(n, n_0, V)$
end for
return $\arg\max_{a \in \mathscr{C}(n_0)} Q(n_0, a)/N(n_0, a)$

5.5.6 Complete Mixed Traffic Algorithm

The complete MCTS-Level-k algorithm for mixed traffic is summarized in Algorithm 3. The algorithm extends Algorithm 2 with driving style estimation, type-dependent opponent modeling, and adaptive safety mechanisms.

The algorithm maintains the four-phase structure of the homogeneous framework while incorporating mixed traffic adaptations: driving style estimation precedes prediction generation, type-dependent modeling differentiates AV and HDV opponents, and adaptive safety mechanisms ensure robust collision avoidance despite human behavioral uncertainty.

5.6 Experimental Validation

This section validates the mixed traffic framework through comprehensive experiments evaluating performance across varying AV penetration rates, driving style distributions, and algorithmic configurations. We demonstrate that the proposed approach maintains safety and efficiency even in challenging scenarios where AVs represent a minority navigating predominantly human-driven traffic.

5.6.1 Experimental Setup

5.6.1.1 Simulation Parameters

Simulation parameters follow the baseline setup in Table 3.2 and the Level-k extensions in Table 4.2. Mixed-traffic-specific parameters are summarised in Table 5.1.

Table 5.1: Mixed Traffic Specific Parameters

Parameter	**Symbol**	**Value**
AV penetration rates	ROP	20%–100%
HDV style distribution	$\eta_h \sim$	Beta$(2,2)$
Style discrete space	η_h	$\{0.1, 0.3, 0.5, 0.7, 0.9\}$
MC collision samples	N_{mc}	100
Confidence parameter	κ_σ	2.0
V2H risk weight	w_{V2H}	0.3
Correlation coefficients	$\rho_{xv}, \rho_{y\theta}$	0.3
PET safety threshold	–	2.0 s
Number of trials	–	40
d_{base}	Baseline safety distance	2.0 m
ζ_v	Velocity risk scaling	0.5
Time step	Δt	0.2 s

5.6.1.2 Scenario Configuration

The mixed traffic experiments employ the eight-agent symmetric intersection from Chapter 4, with agents partitioned into AVs and HDVs according to the specified penetration rate. The AV penetration rate (ROP) varies from 20% to 100%, corresponding to configurations ranging from two AVs among six HDVs to fully autonomous traffic.

Each HDV is assigned a driving style parameter $\eta_h \in [0,1]$ sampled from a Beta distribution $\eta_h \sim \text{Beta}(\alpha,\beta)$, providing flexible control over population characteristics. Unless otherwise specified, experiments employ a balanced distribution $\text{Beta}(2,2)$ with mean $\bar{\eta}_h = 0.5$, producing approximately 25% conservative, 50% moderate, and 25% aggressive drivers to represent realistic traffic diversity. The impact of alternative style distributions is examined in the ablation study (Section 5.6.3).

5.6.1.3 Baseline Methods

We compare against three baseline methods from Chapter 4: Stackelberg Game (hierarchical leader-follower coordination), Nash Equilibrium (simultaneous game-theoretic coordination), and Vanilla MCTS (baseline MCTS without Level-k reasoning or adaptive safety). Additionally, we evaluate an ablated variant of our framework with short planning horizon ($H = 4$) to isolate the contribution of planning depth, alongside the full framework with $H = 9$ to verify long-term coordination performance.

5.6.1.4 Evaluation Metrics

Performance is evaluated through six metrics: *Collision Rate* (%), the percentage of trials with safety violations; *Arrival Rate* (%), the percentage of agents successfully clearing the intersection; *Travel Time* (s), the total simulation duration reflecting planning efficiency; *Trajectory Deviation* (m), the mean distance from reference paths; *Post-Encroachment Time* (PET), the temporal separation between conflicting vehicles, where PET < 2 s indicates critically small safety margins; and *Computational Efficiency*, assessed via the

theoretical complexity analysis in Table 4.5. As shown in Section 4.5.4, the proposed framework achieves sub-100 ms per-step planning through Level-k decomposition, safety-aware pruning, and trajectory caching, confirming real-time implementability without requiring additional runtime experiments.

All experiments are conducted over 40 independent trials with randomized initial conditions, where each agent's longitudinal starting position is independently perturbed by $\mathscr{U}(-0.05, 0.05)$ m along the reference path, and HDV driving styles are randomly assigned per trial, to assess robustness.

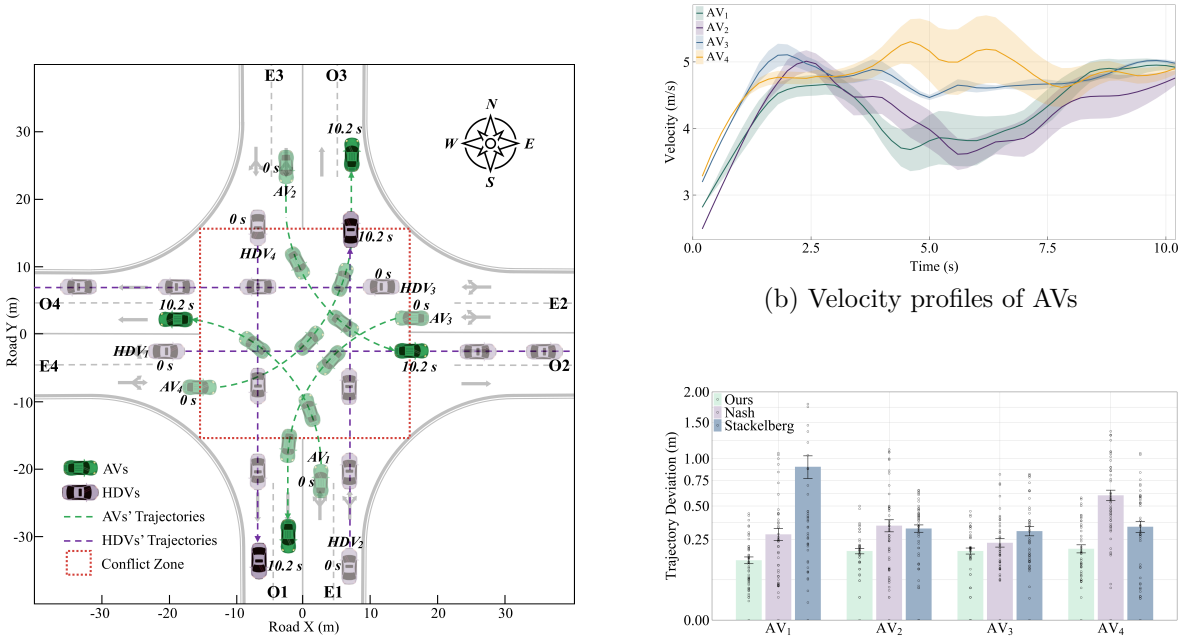
5.6.2 Case 4: Mixed Traffic at 50% Penetration

The primary mixed traffic evaluation employs 50% AV penetration, where four AVs coordinate with four HDVs at the symmetric intersection. This configuration represents a challenging near-term deployment scenario requiring AVs to navigate substantial human traffic while maintaining safety and efficiency.

5.6.2.1 Scenario Visualization

Figure 5.5 presents the mixed traffic scenario and performance overview. The scenario visualization (Figure 5.5a) shows four AVs and four HDVs approaching the intersection from different directions. HDV driving styles are sampled from the balanced Beta$(2, 2)$ distribution, introducing realistic behavioral diversity.

The velocity profiles (Figure 5.5b) demonstrate our framework's capability to handle mixed traffic interactions. Compared to the homogeneous AV scenarios in Chapter 4, velocity variations show larger fluctuations between 3–5 m/s with wider confidence intervals, reflecting increased uncertainty introduced by human drivers. Nevertheless, the profiles maintain overall smooth transitions without abrupt emergency maneuvers, indicating successful adaptation to HDV behaviors through the style-aware prediction and adaptive safety mechanisms.



(a) Mixed traffic scenario

(b) Velocity profiles of AVs

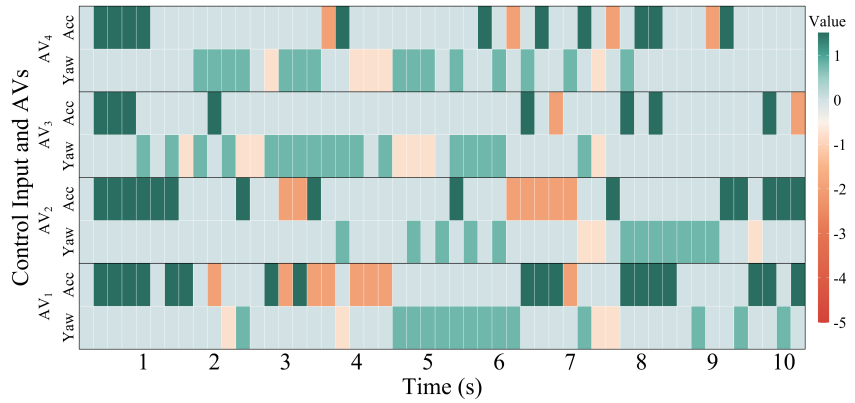
(c) Trajectory deviation comparison

Figure 5.5: Performance analysis in Case 4 (mixed traffic, ROP = 50%). The proposed method enables adaptive coordination between AVs and HDVs, yielding smoother velocity regulation and reduced trajectory deviation.

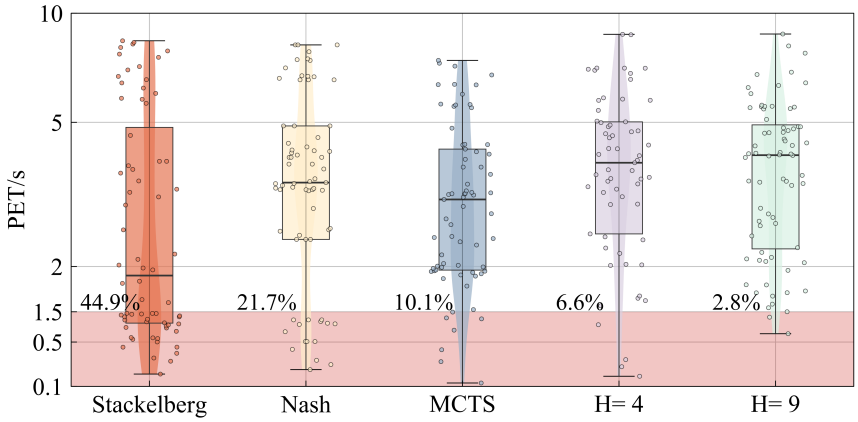
The trajectory deviation comparison (Figure 5.5c) reveals superior performance of our approach. Our method achieves mean deviation of **0.24 m**, representing improvements of **51.8%** over Nash equilibrium (**0.50 m**) and **62.4%** over Stackelberg (**0.64 m**). The reduced deviation demonstrates that adaptive risk assessment enables confident path execution even amid unpredictable human behaviors.

5.6.2.2 Control and Safety Analysis

Figure 5.6 presents detailed control and safety analysis. The control input heatmap (Figure 5.6a) illustrates more diverse and frequent adjustments in AV behaviors compared to the homogeneous scenarios in Chapter 4. The increased presence of lighter and darker color patches indicates that AVs make more dynamic adjustments to accommodate the less predictable movements of human drivers. This adaptive behavior demonstrates the framework's capability to balance assertiveness (maintaining progress) and cautiousness (ensuring safety) when interacting with HDVs.



(a) Control input heatmap: upper row per agent shows acceleration (m/s^2), lower row shows yaw rate (rad/s)



(b) PET distribution comparison

Figure 5.6: Control and safety analysis in Case 4. (a) Control input heatmap showing more diverse adjustments compared to homogeneous scenarios, reflecting AV adaptation to HDV unpredictability. (b) Post-Encroachment Time distributions across methods, where our approach ($H = 9$) achieves only 2.8% violations below the 2s safety threshold.

Safety performance through PET distributions (Figure 5.6b) reveals the challenges of mixed traffic scenarios. Our method with $H = 9$ maintains the best safety performance with only 2.8% PET violations (instances where temporal separation falls below the critical 2-second threshold). The short-horizon variant ($H = 4$) shows increased violations at 10.1%, confirming that adequate planning depth remains essential for anticipating HDV behaviors. Baseline methods exhibit significantly higher violation rates: Stackelberg at 44.9% and Nash at 21.7%, reflecting their inability to handle human behavioral uncertainty within their game-theoretic frameworks.

Compared with the homogeneous scenarios, large positive acceleration and yaw rate values are no longer observed. This is primarily due to the incorporation of human-driven vehicles, whose future behaviors are modeled with uncertainty. The uncertainty-aware safety mechanism enlarges the effective safety margin and penalizes aggressive maneuvers that may lead to potential conflicts under uncertain predictions. As a result, the AV adopts more conservative and smoother control strategies, avoiding abrupt acceleration or sharp turning actions. This behavior reflects a shift from aggressive coordination under predictable AV-AV interactions to risk-aware planning under uncertain AV-HDV interactions.

5.6.2.3 Quantitative Results

Table 5.2: Performance Comparison in Case 4: Mixed Traffic (ROP = 50%)

Method	Arrival Rate (%)	Collision Rate (%)	Travel Time (s)
Stackelberg	55.0	35.0	16.7 ± 3.0
Nash	60.0	30.0	23.4 ± 2.4
Vanilla MCTS	70.0	17.5	$\mathbf{10.0 \pm 1.4}$
Ours ($H = 4$)	85.0	5.0	10.9 ± 2.2
Ours ($H = 9$)	**95.0**	**2.5**	11.9 ± 2.0

Table 5.2 summarizes quantitative performance in the mixed traffic scenario. Our method with $H = 9$ achieves comprehensive superiority:

Safety: Collision rate of 2.5% represents dramatic improvement over Stackelberg (35%), Nash (30%), and vanilla MCTS (17.5%). The near-zero collision rate despite substantial human traffic validates the effectiveness of style-aware prediction and adaptive safety margins in mitigating multi-agent conflicts.

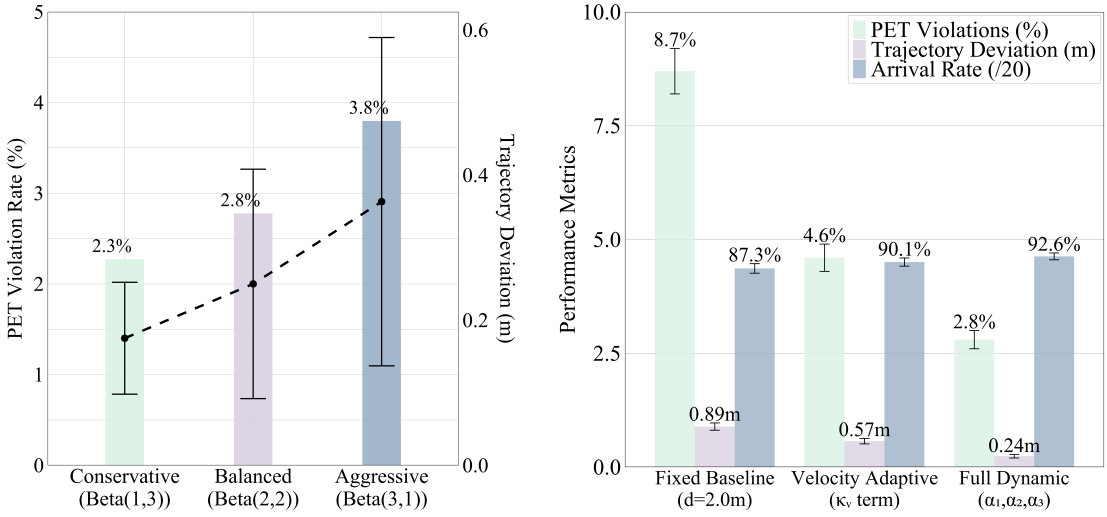
Efficiency: Arrival rate of 95% significantly exceeds all baselines, with 36% improvement over vanilla MCTS (70%). This efficiency is achieved without compromising safety, demonstrating that the adaptive framework successfully balances cautious margins around HDVs with assertive progress when safe.

While resulting in slightly longer *travel time* (11.9 s) than vanilla MCTS (10.0 s), this modest increase (19%) enables dramatic safety improvements. The additional delay stems from more conservative decision-making under driving style uncertainty, which is essential for robust coordination in mixed traffic. The short-horizon variant ($H = 4$) shows intermediate performance, confirming that planning depth remains critical even with sophisticated human modeling. Insufficient lookahead prevents anticipation of HDV maneuvers, leading to reactive rather than proactive coordination in complex multi-agent encounters.

5.6.3 Ablation Study

To validate individual component contributions and assess robustness to parameter variations, we conduct comprehensive ablation studies examining driving style modeling, adaptive safety thresholds, and performance across penetration rates.

5.6.3.1 Impact of Driving Style Modeling



(a) Driving style distribution impact

(b) Dynamic threshold benefit

Figure 5.7: Ablation study results at ROP = 50%. (a) Impact of HDV driving style distributions on PET violations and trajectory deviations. The framework maintains consistent safety (2.3–3.8% violations) across conservative, balanced, and aggressive populations. (b) Benefit of dynamic safety thresholds: the full adaptive formulation reduces PET violations by 67.8% compared to fixed baselines.

Figure 5.7a evaluates performance under three representative driving style distributions at 50% AV penetration. The aggressive-dominated scenario presents the highest challenge with PET violations of 3.8%, compared to 2.8% in balanced and 2.3% in conservative scenarios. This correlation confirms that aggressive driving styles increase collision risk as expected from the style parameter scaling in Equations (5.6).

However, the minimal variation in violation rates—only 1.5 percentage points across drastically different behavioral distributions—validates our framework's robustness. The adaptive uncertainty scaling from Equations (5.11) automatically expands safety margins when detecting aggressive behaviors characterized by higher accelerations and shorter headways, maintaining consistent safety performance despite behavioral diversity.

Trajectory deviations increase progressively with surrounding HDV aggressiveness: from $0.18\,\mathrm{m}$ in conservative-dominated scenarios to $0.24\,\mathrm{m}$ in balanced and $0.31\,\mathrm{m}$ in aggressive-dominated traffic. This response pattern illustrates the adaptive risk assessment mechanism: when detecting aggressive maneuvers, the framework expands both safety margins through adjustment factors $\{\alpha_k\}$ and uncertainty bounds through covariance scaling $\Sigma_h(\tau, \eta_h)$, producing more conservative trajectories that prioritize collision avoidance over precise path tracking.

In addition to homogeneous driving style distributions, we also consider the more realistic scenario where conservative, balanced, and aggressive HDVs coexist within the same environment. In such heterogeneous settings, the proposed framework maintains stable performance by adapting to each individual driver's estimated style parameter η_h. Specifically, aggressive drivers (high η_h) induce larger uncertainty and expanded safety margins through the covariance scaling $\Sigma_h(\tau, \eta_h)$ and adaptive threshold formulation, while conservative drivers allow tighter coordination due to their more predictable behavior. This leads to asymmetric interactions where AVs exhibit locally adaptive behavior: maintaining larger clearance around aggressive HDVs while exploiting coordination opportunities with conservative ones. As a result, the overall system behavior reflects a balanced trade-off between safety and efficiency, without significant degradation compared to the homogeneous scenarios. This demonstrates the robustness of the framework to mixed behavioral distributions commonly observed in real-world traffic.

5.6.3.2 Benefit of Dynamic Safety Thresholds

Figure 5.7b quantifies the substantial benefits of context-aware safety thresholds from Equation (5.13). We compare three configurations:

- **Fixed baseline**: Constant $d_{\text{safe}} = 2.0\,\text{m}$ regardless of interaction characteristics.
- **Velocity-adaptive**: Incorporates relative velocity scaling $\kappa_v|\Delta v_{ij}|$ with fixed adjustment factors ($\gamma_1 = \gamma_2 = \gamma_3 = 1$).
- **Full dynamic**: Complete formulation from Equation (5.13) with all adjustment factors.

The dynamic approach achieves dramatic improvements: PET violations drop by 67.8% (from 8.7% to 2.8%) and trajectory deviations by 33.9% (from 0.89 m to 0.24 m), while arrival rates improve from 87.3% to 92.6%.

The velocity-adaptive configuration achieves intermediate performance with 4.6% violations, confirming that velocity-dependent scaling provides significant benefits. However, complete integration of all contextual factors—relative velocity through γ_1, heading conflicts through γ_2, and high-risk spatial zones through γ_3—delivers superior safety guarantees essential for complex multi-directional intersection conflicts.

5.6.3.3 Performance Across Penetration Rates

Figure 5.8 presents PET distributions across seven penetration levels from ROP = 20% to ROP = 100%, revealing how coordination quality evolves with increasing autonomy.

At low penetration rates (ROP 20%–33.3%), where AVs represent a minority in predominantly human traffic, PET distributions exhibit substantial variance reflecting HDV behavioral unpredictability. Despite this uncertainty, violation rates remain remarkably low at 0.0–3.0%, demonstrating that the conservative Level-0 initialization with extended safety margins ε_0 provides robust protection even when AVs cannot influence overall traffic patterns. AVs adopt predominantly defensive strategies, yielding to human drivers to guarantee collision-free navigation.

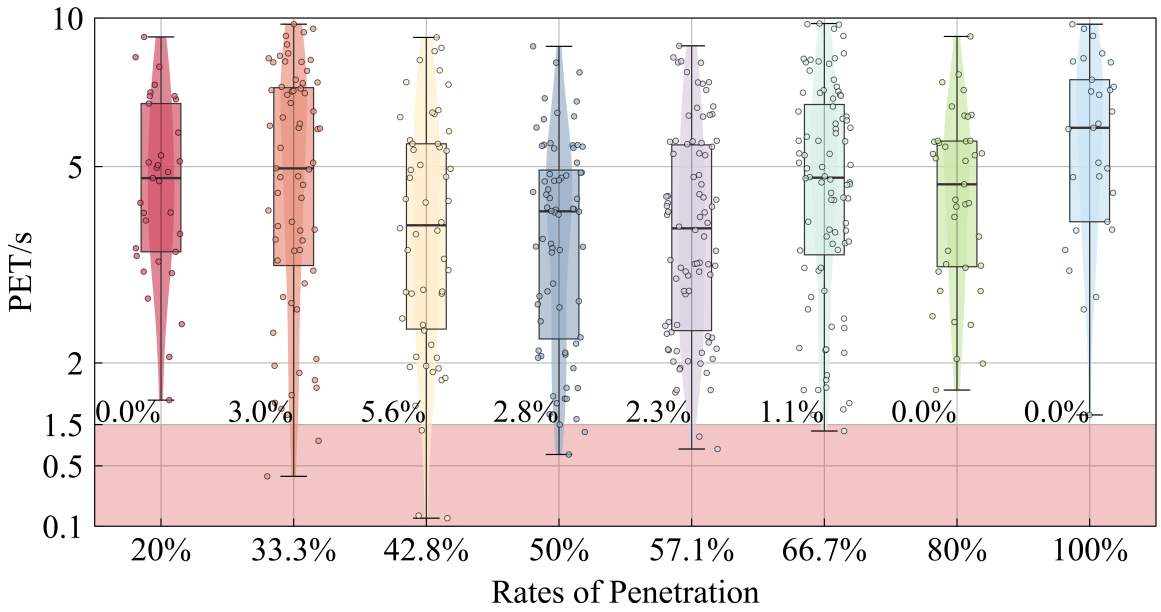


Figure 5.8: Post-Encroachment Time distributions across AV penetration rates from 20% to 100%. Box plots show median (center line), quartiles (box boundaries), and outliers. The red shaded region indicates critical safety violations (PET < 2s). As penetration increases, distributions become concentrated with rising median PET values, while violation rates decrease systematically to zero at ROP $\geq 66.7\%$.

The medium penetration range (ROP 42.8%–57.1%) exhibits the most complex dynamics as the system transitions between regimes. Violation rates initially increase to **5.6%** at ROP $= 42.8\%$ before stabilizing around **2%** at higher penetrations. This non-monotonic behavior reflects intensified interaction complexity: AVs attempting strategic coordination must simultaneously adapt to unpredictable HDV behaviors, creating scenarios where neither purely defensive nor fully strategic approaches prove optimal. The V2H safety assessment framework from Equation (5.20) becomes critical in this regime.

At high penetration rates (ROP 66.7%–100%), violation rates drop to **0%** and PET distributions narrow dramatically, with median values rising from **5–6 s** to **8–10 s**. Level-k coordination becomes increasingly effective as more agents employ strategic reasoning, establishing implicit passing orders without explicit communication. The reduced variance indicates consistent coordination quality across trials.

These results confirm that the MCTS-Level-k framework with adaptive safety mechanisms provides robust guarantees across the full spectrum of autonomy adoption, gracefully degrading from efficient multi-agent optimization in predominantly autonomous traffic to safe defensive navigation in human-dominated scenarios.

5.7 Chapter Summary

This chapter extended the MCTS-Level-k framework to mixed traffic environments where autonomous vehicles must coordinate with human-driven vehicles exhibiting diverse and uncertain behaviors. The extension addresses fundamental challenges absent in homogeneous AV coordination: behavioral uncertainty, driving style diversity, and asymmetric interaction dynamics inherent in human-AI cooperation.

5.7.1 Key Contributions

The chapter introduced three complementary mechanisms enabling robust mixed traffic coordination to handle human uncertainty:

Style-Aware Human Behavior Modeling (Section 5.2): The Intelligent Driver Model extended with style parameter $\eta_h \in [0, 1]$ captures the spectrum from conservative to aggressive driving through adapted parameters for maximum acceleration (5.6a), time headway (5.6b), and following distance (5.6c). The probabilistic prediction framework (5.7) provides both nominal trajectories and uncertainty bounds for safety assessment.

Uncertainty Quantification (Section 5.3): The time-varying covariance structure (5.8) captures how prediction uncertainty grows over the planning horizon, with position uncertainty growing quadratically (5.9) due to accumulated velocity errors. Style-dependent scaling (5.11) reflects that aggressive drivers exhibit higher behavioral variability, while correlation terms (5.12) model physical couplings in vehicle motion.

Adaptive Safety Assessment (Section 5.4): Context-aware safety thresholds (5.13) adapt to interaction characteristics through adjustment factors for relative velocity (5.14a), heading conflicts (5.14b), and spatial location (5.14c). The unified V2H risk metric (5.20) combines instantaneous risk, temporal risk, and collision probability for a comprehensive safety evaluation of human behaviors.

The algorithm adaptations (Section 5.5) integrate these mechanisms into the MCTS-Level-k framework through mean-based probabilistic rollouts (5.21), driving style estimation (5.24), type-dependent opponent modeling (5.26), and risk-integrated reward evaluation (5.30) to ensure safe navigation.

5.7.2 Experimental Validation

Comprehensive experiments validated the mixed traffic framework across diverse conditions, including varying traffic densities and heterogeneous human driving styles:

Case 4 (ROP = 50%): At the challenging 50% penetration rate, our method achieved 2.5% collision rate versus 17.5–35.0% for baselines, with 95% arrival rate versus 55.0–70.0%. The near-zero collision rate despite substantial human traffic validates the effectiveness of adaptive safety mechanisms.

Ablation Studies: Driving style analysis confirmed robustness across conservative, balanced, and aggressive HDV populations with violation rates varying only 1.5 percentage points (2.3–3.8%). Dynamic safety thresholds reduced PET violations by 67.8% compared to fixed baselines, validating the importance of context-aware adaptation.

Penetration Rate Analysis: Performance evaluation from ROP 20% to 100% demonstrated graceful degradation from efficient multi-agent coordination in predominantly autonomous traffic to safe defensive navigation in human-dominated scenarios, with violation rates at 0% for ROP $\geq$ 66.7%.

5.7.3 Implications for Deployment

The mixed traffic framework addresses a critical deployment challenge: autonomous vehicles must operate reliably throughout the gradual transition from current human-dominated traffic to future fully autonomous environments. The demonstrated robustness across penetration rates—from minority AV scenarios requiring defensive coordination to majority AV scenarios enabling efficient optimization—provides confidence that the MCTS-Level-k approach can support real-world deployment across this transition period.

The adaptive mechanisms developed in this chapter—style-aware prediction, uncertainty quantification, and context-aware safety—represent general techniques applicable beyond intersection coordination to any autonomous vehicle application requiring safe interaction with unpredictable human road users.

Chapter 6

Conclusions and Future Work

This thesis has addressed one of the fundamental challenges in autonomous driving: enabling multiple vehicles to coordinate safely and efficiently at unsignalized intersections where traditional traffic control mechanisms are absent and strategic reasoning becomes essential. We introduced a novel framework that fundamentally reconceptualizes Level-k cognitive hierarchy for safety-critical multi-agent coordination, integrating it with Monte Carlo Tree Search to achieve scalable planning with emergent safety properties. Through systematic development from theoretical foundations to algorithmic implementation and experimental validation, we demonstrated that principled integration of cognitive hierarchy theory and search-based planning can resolve longstanding tensions between computational tractability, strategic sophistication, and safety guarantees.

6.1 Summary of Contributions

The primary contribution of this thesis is the development of a unified framework for multi-agent decision-making that transforms Level-k reasoning from a descriptive model of bounded rationality into a constructive planning methodology with emergent safety properties. At the core of this transformation lies the reconceptualization of Level-0: rather than modeling naive or random agent behaviors as in classical formulations, we define Level-0 as a universal safety initialization procedure that generates conservative baseline trajectories with extended safety margins ε_0. This procedural interpretation ensures that every agent establishes a collision-free safety anchor before engaging in strategic optim-

ization, with safety margins cascading and amplifying through higher reasoning levels ($\varepsilon_0 < \varepsilon_1 < \varepsilon_2$) via bounded optimization. The result is a framework where safety emerges as a structural property of the recursive reasoning hierarchy rather than being imposed through external constraints.

Chapter 3 established the foundational MCTS framework for multi-agent coordination, demonstrating its effectiveness in four-agent symmetric intersection scenarios. We formalized individual agent models including the four-dimensional state representation, discrete action space with 15 control primitives, kinematic bicycle dynamics, and precise collision detection via the Separating Axis Theorem. The multi-objective cost function balanced safety, efficiency, path adherence, and comfort objectives within a Markov Decision Process formulation amenable to tree search methods. Experimental validation demonstrated that MCTS with sufficient planning horizon ($H = 9$) achieves zero collision rate in four-agent scenarios, establishing the baseline upon which subsequent chapters build. Critically, we identified the fundamental scalability limitation: the joint action space grows as $\mathcal{O}(|\mathcal{A}|^{N \cdot H})$, exploding from 10^{42} operations for four agents to 10^{85} for eight agents, rendering vanilla MCTS intractable for larger populations.

Chapter 4 addressed these scalability limitations through two complementary mechanisms. The dynamic interaction graph with dual spatial-strategic filtering reduced effective opponent modeling from $N-1$ agents to approximately 3–4 spatially and strategically relevant agents, achieving linear rather than exponential scaling. The reconstructed Level-k cognitive hierarchy decomposed multi-agent coordination into sequential single-agent optimizations: Level-1 agents optimize against Level-0 conservative baselines, and Level-2 agents optimize against predicted Level-1 responses. The cascading safety property ensures that higher reasoning depth strengthens rather than compromises safety guarantees. The integrated MCTS-Level-k algorithm achieved dramatic complexity reduction through safety-aware pruning (reducing branching factor from 15 to approximately 4.5), trajectory caching, and filtered interaction graphs. Experimental validation on symmetric eight-agent scenarios demonstrated zero collision rate where baseline methods exhibited 15–35% collisions, with computation times of 55–61 ms enabling real-time deployment.

Chapter 5 extended the framework to heterogeneous mixed traffic involving human-driven vehicles. The style-aware Intelligent Driver Model with parameter $\eta_h \in [0,1]$ captured driving personality diversity from conservative to aggressive behaviors. Time-varying uncertainty quantification through covariance matrix $\Sigma_h(\tau, \eta_h)$ modeled prediction confidence degradation over the planning horizon, with position uncertainty growing quadratically due to accumulated velocity errors. Adaptive safety thresholds responding to relative velocity, heading conflicts, and spatial location ensured robust collision avoidance despite behavioral variability. The unified V2H risk metric combining instantaneous risk, temporal risk, and collision probability integrated seamlessly with Level-k reasoning. Experimental validation across penetration rates from 20% to 100% demonstrated consistent performance: collision rates below 2% at 50% penetration despite diverse human driving behaviors, with graceful degradation from efficient coordination in AV-dominated traffic to defensive navigation in human-dominated scenarios.

The complete framework achieves 21 orders of magnitude complexity reduction compared to joint optimization (from 10^{85} to approximately 6×10^4 operations), enabling sub-100 millisecond planning cycles. Beyond quantitative performance, the framework offers qualitative advantages critical for deployment: interpretability through explicit Level-k reasoning traces and MCTS tree structures, modularity enabling component-wise validation and adaptation, and fully decentralized architecture requiring no inter-vehicle communication.

6.2 Broader Implications

The contributions of this thesis extend beyond autonomous intersection coordination to broader questions in multi-agent systems, artificial intelligence, and robotics.

Methodological contribution: The reconceptualization of Level-k reasoning demonstrates that classical cognitive models from behavioral economics can be transformed into constructive planning frameworks for artificial agents when appropriately reformulated. This transformation from descriptive to prescriptive modeling represents a contribution applicable to other domains where bounded rationality models inform algorithm design.

Computational principle: The integration of cognitive hierarchy theory with search-based planning illustrates a general principle: hierarchical decomposition of complex problems, when combined with selective exploration through sampling-based methods, can resolve computational intractability without sacrificing solution quality. This principle generalizes beyond autonomous driving to multi-agent coordination domains including air traffic control, warehouse robotics, multi-robot exploration, and distributed computing, where exponential joint action spaces present similar scalability challenges. The techniques developed here—dual filtering, safety-aware pruning, trajectory caching—provide a toolkit for addressing these challenges across diverse applications.

Safety paradigm: The emergent safety properties arising from cascading conservative margins offer an alternative paradigm to explicit constraint enforcement in optimization-based planning. Rather than formulating collision avoidance as hard constraints that complicate optimization or soft penalties that trade off against efficiency objectives, our framework embeds safety structurally through the recursive reasoning hierarchy. This structural approach provides robustness to model mismatches and prediction errors, as multiple layers of conservatism buffer against deviations from expected behaviors—increasingly important as autonomous systems deploy in open-world environments where comprehensive modeling of all scenarios remains infeasible.

Human-robot interaction: The mixed traffic modeling framework addresses a critical transition challenge: coordinating with human drivers exhibiting diverse, uncertain, and boundedly rational behaviors. The probabilistic extensions integrating Gaussian uncertainty models with adaptive safety mechanisms provide principled approaches to human-robot interaction that balance safety robustness with operational efficiency. These techniques extend beyond driving to other human-robot collaboration contexts including shared workspace manufacturing, assistive robotics, and human-robot teaming.

AI perspective: The thesis demonstrates that effective multi-agent coordination need not require either perfect rationality (as assumed by game-theoretic equilibrium concepts) or extensive offline learning (as required by deep reinforcement learning approaches). Bounded rationality models combined with online planning through selective sampling provide a middle path achieving strategic sophistication comparable to equilibrium computation while maintaining computational tractability and adaptability to novel scenarios.

6.3 Limitations and Future Directions

6.3.1 Modeling Assumptions and Generalization

The experimental evaluation in this thesis is conducted under a set of structured modeling assumptions that enable tractable multi-agent planning and controlled analysis of coordination behaviors. These include idealized state observability, simplified interaction models for opponent prediction, and canonical intersection geometries.

While these assumptions are standard in simulation-based studies, they limit the direct generalizability of the results to more complex real-world environments. In particular, performance under asymmetric road layouts, partial observability, perception noise, and richer human behavior models remains to be systematically investigated.

Nevertheless, the proposed framework is modular and extensible, allowing these components to be refined without altering the underlying planning architecture. Extending the framework to incorporate more realistic sensing, behavior prediction, and environmental complexity represents an important direction for future work.

6.3.2 Algorithmic Extensions

Adaptive level assignment: The current framework assigns reasoning levels dynamically at each planning cycle based on interaction complexity scores but does not adapt based on accumulated coordination outcomes. Future work could investigate meta-reasoning mechanisms that learn which reasoning depths prove most effective for different traffic patterns or opponent types, allocating deep reasoning only when necessary to improve computational efficiency.

Reasoning depth selection: The Level-k hierarchy extends to depth two (Level-1 and Level-2 operational reasoning beyond Level-0 initialization). While this bounded depth ensures tractability and achieves excellent performance, theoretical questions remain about optimal depth-accuracy tradeoffs. Future research could develop principled methods for depth selection, potentially adapting based on scenario complexity or available computational resources, and extending to heterogeneous depth assignments where different agents employ different maximum reasoning depths.

Refined Level-0 predictions: The conservative Level-0 baseline employs constant velocity predictions, providing predictable behaviors enabling safety guarantees but potentially underestimating opponent maneuverability. More sophisticated Level-0 models accounting for basic reactive behaviors such as collision avoidance braking could tighten safety margins while maintaining conservatism.

6.3.3 System Extensions

Heterogeneous autonomous fleets: The framework currently treats all autonomous vehicles as employing the same MCTS-Level-k algorithm. Real deployments will involve heterogeneous fleets with different manufacturers, sensing capabilities, and planning algorithms. Extending the framework to reason about heterogeneous autonomous agents—some employing Level-k reasoning, others using learning-based policies or rule-based methods—represents an important direction for practical deployment.

Complex traffic environments: Experimental validation focused on intersection scenarios with discrete entry-exit configurations. Extension to roundabouts with continuous flow, highway merging with high-speed dynamics, and urban arterials with multiple decision points requires domain-specific adaptation of spatial filtering, level assignment heuristics, and computational optimization.

Detailed vehicle dynamics: The current kinematic models suit planning at intersection timescales where strategic considerations dominate. Integration with more detailed dynamic models accounting for tire friction, suspension dynamics, and actuator constraints would enable application to high-performance driving scenarios.

6.3.4 Deployment Considerations

Perception integration: While the thesis demonstrates real-time performance with idealized sensing providing perfect state information, practical deployment requires integration with perception and localization systems introducing uncertainty, latency, and occasional failures. Future research should investigate robustness to perception errors and techniques to propagate sensing uncertainty through the Level-k hierarchy.

Anytime planning: The framework employs discrete planning cycles with receding horizon replanning at fixed frequencies. Anytime planning variants that refine solutions continuously could provide more graceful degradation under computational resource constraints, maintaining feasible solutions while progressively improving quality.

Formal verification: Theoretical analysis provides convergence guarantees for MCTS and complexity bounds for the framework, but formal verification of safety properties remains open. Reachability analysis or barrier certificate techniques could provide provable safety guarantees satisfying regulatory requirements for autonomous vehicle certification.

6.3.5 Broader Research Directions

Human factors: How do human drivers perceive and respond to autonomous vehicles employing Level-k reasoning? Field studies with human participants could provide empirical insights informing design refinements that enhance human-robot interaction quality and appropriate trust calibration.

Policy implications: The decentralized coordination enabled by Level-k reasoning could support heterogeneous deployments without centralized infrastructure, potentially lowering barriers to market entry. Research at the intersection of technology, economics, and policy could inform regulatory frameworks enabling beneficial deployment while managing risks associated with autonomous-human interaction.

6.4 Closing Remarks

The development and validation of the MCTS-Level-k framework demonstrates that scalable, safe, and strategically sophisticated multi-agent coordination is achievable through principled integration of cognitive hierarchy theory and search-based planning. By reconceptualizing Level-k reasoning to prioritize safety initialization over behavioral description, we transform a model of bounded rationality into a constructive planning framework with emergent safety properties. The dramatic complexity reduction—21 orders of magnitude from joint optimization—enables real-time coordination in scenarios previously considered intractable, while maintaining strategic sophistication comparable to game-theoretic approaches and safety robustness exceeding learning-based methods.

As autonomous vehicles transition from controlled testing environments to complex mixed traffic scenarios, frameworks that guarantee safety while reasoning strategically about diverse interactions become increasingly critical. This thesis contributes toward that vision, offering theoretical insights, algorithmic techniques, and empirical validation that advance both the science of multi-agent coordination and the engineering of safe autonomous systems. The principles developed here—hierarchical decomposition, emergent safety through conservative initialization, selective exploration through sampling, and adaptive reasoning about heterogeneous agents—provide foundations for future multi-agent coordination systems across diverse domains where intelligent agents must coordinate in complex, uncertain, and safety-critical environments.

The journey from theoretical formulation through algorithmic development to experimental validation reveals that effective multi-agent coordination requires neither perfect rationality nor unlimited computation, but principled reasoning structures that decompose complexity while preserving essential properties. The MCTS-Level-k framework exemplifies this philosophy, demonstrating that when theoretical insights from behavioral economics meet algorithmic techniques from artificial intelligence, the result can be practical systems achieving the seemingly incompatible goals of scalability, safety, and strategic sophistication in real-world autonomous navigation.

Bibliography

[1] Department for Transport, "Road accidents and safety statistics," https://www.gov.uk/government/collections/road-accidents-and-safety-statistics, 2023, accessed: 2024-04-28.

[2] D. Omeiza, H. Webb, M. Jirotka, and L. Kunze, "Explanations in autonomous driving: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 10 142–10 162, 2021.

[3] H. A. Ignatious, M. Khan *et al.*, "An overview of sensors in autonomous vehicles," *Procedia Computer Science*, vol. 198, pp. 736–741, 2022.

[4] C. Badue, R. Guidolini, R. V. Carneiro, P. Azevedo, V. B. Cardoso, A. Forechi, L. Jesus, R. Berriel, T. M. Paixao, F. Mutz *et al.*, "Self-driving cars: A survey," *Expert systems with applications*, vol. 165, p. 113816, 2021.

[5] H. Vijayakumar, D. Zhao *et al.*, "A holistic safe planner for automated driving considering interaction with human drivers," *IEEE Trans. Intell. Veh.*, vol. 9, no. 1, pp. 2061–2076, 2023.

[6] J. Löfberg, "Minimax approaches to robust model predictive control," Linköping University, Linköping, Sweden, Tech. Rep. LiTH-ISY-R-2486, 2003.

[7] J. Pérez, V. Milanés *et al.*, "Autonomous driving manoeuvres in urban road traffic environment: a study on roundabouts," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 13 795–13 800, 2011.

[8] C. Xu, W. Zhao *et al.*, "A Nash Q-learning based motion decision algorithm with considering interaction to traffic participants," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 12 621–12 634, 2020.

[9] K. Min, H. Kim *et al.*, "Deep Q learning based high level driving policy determination," *IEEE Intelligent Vehicles Symposium*, pp. 226–231, 2018.

[10] S. Gu, T. Lillicrap *et al.*, "Continuous deep Q-learning with model-based acceleration," in *Proceedings of the International Conference on Machine Learning*, 2016, pp. 2829–2838.

[11] H. Wei, X. Liu *et al.*, "Mixed-autonomy traffic control with proximal policy optimization," in *Proceedings of the IEEE Vehicular Networking Conference*, 2019, pp. 1–8.

[12] F. Ye, X. Cheng *et al.*, "Automated lane change strategy using proximal policy optimization-based deep reinforcement learning," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2020, pp. 1746–1752.

[13] G. Dulac-Arnold, R. Evans *et al.*, "Deep reinforcement learning in large discrete action spaces. arxiv 2015," *arXiv preprint arXiv:1512.07679*, 2015.

[14] Z. Tian, D. Zhao *et al.*, "Efficient and balanced exploration-driven decision making for autonomous racing using local information," *IEEE Transactions on Intelligent Vehicles*, 2024.

[15] Z. Tian, D. Zhao, Z. Lin, D. Flynn, W. Zhao, and D. Tian, "Balanced reward-inspired reinforcement learning for autonomous vehicle racing," in *6th Annual Learning for Dynamics & Control Conference*. PMLR, 2024, pp. 628–640.

[16] G. Basile, A. Petrillo, and S. Santini, "DDPG based end-to-end driving enhanced with safe anomaly detection functionality for autonomous vehicles," in *Proceedings of the IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering*, 2022, pp. 248–253.

[17] M. A. Hebaish, A. Hussein *et al.*, "Towards safe and efficient modular path planning using twin delayed DDPG," in *Proceedings of the IEEE Vehicular Technology Conference*, 2022, pp. 1–7.

[18] J. Zhu, K. Gao, H. Li, Z. He, and C. O. Monreal, "Bi-level ramp merging coordination for dense mixed traffic conditions," *Fundamental Research*, 2023.

[19] P. H. et al., "Driving conflict resolution of autonomous vehicles at unsignalized intersections: A differential game approach," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 6, pp. 5136–5146, 2022.

[20] S. Zhang, X. Lei, X. Peng, and J. Pan, "Heterogeneous targets trapping with swarm robots by using adaptive density-based interaction," *IEEE Trans. Robot.*, vol. 40, pp. 2729–2748, 2024.

[21] N. Mehr, M. Wang, M. Bhatt, and M. Schwager, "Maximum-entropy multi-agent dynamic games: Forward and inverse solutions," *IEEE Trans. Robot.*, vol. 39, no. 3, pp. 1801–1815, 2023.

[22] H. Xu, Y. Zhang, C. G. Cassandras, L. Li, and S. Feng, "A bi-level cooperative driving strategy allowing lane changes," *Transp. Res. Part C Emerg. Technol.*, vol. 120, p. 102773, 2020.

[23] C. B. Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis, and S. Colton, "A survey of monte carlo tree search methods," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 1, pp. 1–43, 2012.

[24] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484–503, 2016. [Online]. Available: http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html

[25] D. Lenz, T. Kessler, and A. Knoll, "Tactical cooperative planning for autonomous highway driving using monte-carlo tree search," in *Proc. IEEE IVS*, 2016, pp. 447–453.

[26] Y. Zhang, W. Sun, Y. Chen, Q. Liu, Q. Lin, R. Zhang, and X. Zhao, "Trajectory entropy: Modeling game state stability from multimodality trajectory prediction," *arXiv preprint arXiv:2506.05810*, 2025.

[27] S. Wang, W. Huang, and H. K. Lo, "Combining shockwave analysis and bayesian network for traffic parameter estimation at signalized intersections considering queue spillback," *Transp. Res. Part C: Emerg. Technol.*, vol. 120, p. 102807, 2020.

[28] D. Sun, J. Chen, S. Mitra, and C. Fan, "Multi-agent motion planning from signal temporal logic specifications," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 3451–3458, 2022.

[29] P. Z. et al., "Enhanced nmpc for stochastic dynamic systems driven by control error compensation with entropy optimization," *IEEE Trans. Control Syst. Technol.*, vol. 31, no. 5, pp. 2217–2230, 2023.

[30] C. Toumieh and D. Floreano, "High-speed motion planning for aerial swarms in unknown and cluttered environments," *IEEE Trans. Robot.*, vol. 40, pp. 3642–3656, 2024.

[31] A. G. Philip, Z. Ren, S. Rathinam, and H. Choset, "C*: A new bounding approach for the moving-target traveling salesman problem," *IEEE Trans. Robot.*, vol. 41, pp. 4663–4678, 2025.

[32] X. Zhang, L. Wu *et al.*, "High-speed ramp merging behavior decision for autonomous vehicles based on multiagent reinforcement learning," *IEEE Internet of Things Journal*, vol. 10, no. 24, pp. 22 664–22 672, 2023.

[33] P. Hang, C. Huang, Z. Hu, Y. Xing, and C. Lv, "Decision making of connected automated vehicles at an unsignalized roundabout considering personalized driving behaviours," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4051–4064, 2021.

[34] C. Ma and Others, "Trajectory planning for connected and automated vehicles at isolated signalized intersections under mixed traffic environment," *Transp. Res. Part C Emerg. Technol.*, vol. 130, p. 103309, 2021.

[35] J. Zhang, S.-C. Chai, B.-H. Zhang, and G.-P. Liu, "Distributed model-free sliding-mode predictive control of discrete-time second-order nonlinear multiagent systems with delays," *IEEE Trans. Cybern.*, vol. 52, no. 11, pp. 12 403–12 413, 2022.

[36] V. N. Hartmann, A. Orthey, D. Driess, O. S. Oguz, and M. Toussaint, "Long-horizon multi-robot rearrangement planning for construction assembly," *IEEE Trans. Robot.*, vol. 39, no. 1, pp. 239–252, 2023.

[37] G. Li, X. Liu, and G. Loianno, "Human-aware physical human–robot collaborative transportation and manipulation with multiple aerial robots," *IEEE Trans. Robot.*, vol. 41, pp. 762–781, 2025.

[38] D. Le and E. Plaku, "Multi-robot motion planning with dynamics via coordinated sampling-based expansion guided by multi-agent search," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1868–1875, 2019.

[39] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.

[40] J. Lu, L. Han *et al.*, "Event-triggered deep reinforcement learning using parallel control: A case study in autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 4, pp. 2821–2831, 2023.

[41] N. Li, Y. Yao, I. Kolmanovsky, E. Atkins, and A. R. Girard, "Game-theoretic modeling of multi-vehicle interactions at uncontrolled intersections," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 1428–1442, 2022.

[42] L. Abualigah, S. Ekinci *et al.*, "Modified elite opposition-based artificial hummingbird algorithm for designing fopid controlled cruise control system." *Intelligent Automation & Soft Computing*, vol. 38, no. 2, 2023.

[43] S. Tang, Z. Zhang, Y. Zhang, J. Zhou, Y. Guo, S. Liu, S. Guo, Y.-F. Li, L. Ma, Y. Xue *et al.*, "A survey on automated driving system testing: Landscapes and trends," *ACM Transactions on Software Engineering and Methodology*, vol. 32, no. 5, pp. 1–62, 2023.

[44] Y. Fu, C. Li *et al.*, "An incentive mechanism of incorporating supervision game for federated learning in autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 12, pp. 14 800–14 812, 2023.

[45] X. Huang, D. Kroening, W. Ruan, J. Sharp, Y. Sun, E. Thamo, M. Wu, and X. Yi, "A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability," *Computer Science Review*, vol. 37, p. 100270, 2020.

[46] F.-L. Fan, J. Xiong, M. Li, and G. Wang, "On interpretability of artificial neural networks: A survey," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 6, pp. 741–760, 2021.

[47] C.-K. Ho and C.-T. King, "Lac-rrt: Constrained rapidly-exploring random tree with configuration transfer models for motion planning," *IEEE Access*, vol. 11, pp. 97654–97663, 2023.

[48] Y. Gao, D. Li, Z. Sui, and Y. Tian, "Trajectory planning and tracking control of autonomous vehicles based on improved artificial potential field," *IEEE Trans. Veh. Technol.*, vol. 73, no. 9, pp. 12468–12483, 2024.

[49] R. Szczepanski, "Safe artificial potential field: Novel local path planning algorithm maintaining safe distance from obstacles," *IEEE Robot. Autom. Lett.*, vol. 8, no. 8, pp. 4823–4830, 2023.

[50] J. York and T. Maze, "Economic evaluation of truck collision warning systems," *Transportation Research Circular*, vol. 475, pp. 46–50, 1997.

[51] I. C. Burnett, *Traffic Collisions in North Carolina: Weather, Human Factors, and Economic Analysis, 2013 to 2019*. North Carolina State University, 2023.

[52] X. Cao, M. Li, Y. Tao, and P. Lu, "Hma-sar: Multi-agent search and rescue for unknown located dynamic targets in completely unknown environments," *IEEE Robot. Autom. Lett.*, vol. 9, no. 6, pp. 5567–5574, 2024.

[53] Z. e. a. Kherroubi, S. Aknine, and R. Bacha, "Novel decision-making strategy for connected and autonomous vehicles in highway on-ramp merging," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 12 490–12 502, 2022.

[54] H. Wang, H. Gao, S. Yuan, H. Zhao *et al.*, "Interpretable decision-making for autonomous vehicles at highway on-ramps with latent space reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 8707–8719, 2021.

[55] T.-H. H. Chan, Q. Kuang, and Q. Xue, "Game-theoretically secure distributed protocols for fair allocation in coalitional games," in *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS '25. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2025, p. 463–471.

[56] V. Kovařík, N. Sauerberg, L. Hammond, and V. Conitzer, "Game theory with simulation in the presence of unpredictable randomisation," in *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS '25. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2025, p. 1191–1199.

[57] D. Li, J. Zhang, and G. Liu, "Autonomous driving decision algorithm for complex multi-vehicle interactions: An efficient approach based on global sorting and local gaming," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 7, pp. 6927–6937, 2024.

[58] R. Chandra and D. Manocha, "Gameplan: Game-theoretic multi-agent planning with human drivers at intersections, roundabouts, and merging," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 2676–2683, 2022.

[59] A. Constantinescu and R. Wattenhofer, "Byzantine game theory: Sun tzu's boxes," in *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS '25. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2025, p. 519–528.

[60] G. Zhang, J. Xie, B. Peng, H. Zhang, and D. Song, "Power allocation strategy of multi-static radar network tracking maneuvering jammer based on stackelberg game," *IEEE Trans. Veh. Technol.*, pp. 1–10, 2025.

[61] Z. Lin, Z. Tian, J. Lan, Q. Zhang, Z. Ye, H. Zhuang, and X. Zhao, "A conflicts-free, speed-lossless kan-based reinforcement learning decision system for interactive driving in roundabouts," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, pp. 1–14, 2025.

[62] Z. Tian *et al.*, "Efficient and balanced exploration-driven decision making for autonomous racing using local information," *IEEE Trans. on Intell. Veh.*, pp. 1–17, 2024.

[63] S. Schmidt, L. Stappen, L. Schwinn, and S. Günnemann, "Generalized synchronized active learning for multi-agent-based data selection on mobile robotic systems," *IEEE Robot. Autom. Lett.*, vol. 9, no. 10, pp. 8659–8666, 2024.

[64] Q. Shi, M. Liu, S. Zhang, and X. Lan, "Reinforcement learning for multi-agent path finding in large-scale warehouses via distributed policy evolution," *IEEE Robot. Autom. Lett.*, vol. 10, no. 8, pp. 7843–7850, 2025.

[65] M. Lauri, D. Hsu, and J. Pajarinen, "Partially observable markov decision processes in robotics: A survey," *IEEE Trans. Robot.*, vol. 39, no. 1, pp. 21–40, 2023.

[66] P. Weingertner, M. Ho, A. Timofeev, S. Aubert, and G. Pita-Gil, "Monte carlo tree search with reinforcement learning for motion planning," in *Proc. ITSC*, 2020, pp. 1–7.

[67] M. Wang *et al.*, "Speed planning for autonomous driving in dynamic urban driving scenarios," in *Proc. ECCE*, 2020, pp. 1462–1468.

[68] M. Dang, D. Zhao, Y. Wang, and C. Wei, "Dynamic game-theoretical decision-making framework for vehicle-pedestrian interaction with human bounded rationality," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 7, pp. 10 822–10 833, 2025.

[69] D. Jing, E. Yao, and R. Chen, "Decentralized human-like control strategy of mixed-flow multi-vehicle interactions at uncontrolled intersections: A game-theoretic approach," *Transp. Res. Part C Emerg. Technol.*, vol. 167, p. 104835, 2024.

[70] Z. Lin and Z. Tian, "Scenario-based decision-making using game theory for interactive autonomous driving: A survey," *arXiv preprint arXiv:2509.05777*, 2025.

[71] M. Yuan, J. Shan, and H. Schofield, "Scalable game-theoretic decision-making for self-driving cars at unsignalized intersections," *IEEE Trans. Ind. Electron.*, vol. 71, no. 6, pp. 5920–5930, 2024.

[72] P. Mestres, C. Nieto-Granda, and J. Cortés, "Distributed safe navigation of multi-agent systems using control barrier function-based controllers," *IEEE Robot. Autom. Lett.*, vol. 9, no. 7, pp. 6760–6767, 2024.

[73] J. Liao, T. Liu *et al.*, "Decision-making strategy on highway for autonomous vehicles using deep reinforcement learning," *IEEE Access*, vol. 8, pp. 177 804–177 814, 2020.

[74] W. Yuan, M. Yang *et al.*, "Multi-reward architecture based reinforcement learning for highway driving policies," in *Proceeding of the IEEE Intelligent Transportation Systems Conference*, 2019, pp. 3810–3815.

[75] S. Aradi, T. Becsi *et al.*, "Policy gradient based reinforcement learning approach for autonomous highway driving," in *Proceeding of the IEEE Conference on Control Technology and Applications*. IEEE, 2018, pp. 670–675.

[76] J. Wang, T. Yang *et al.*, "Learning an efficient and safe policy for highway driving using supervised learning and reinforcement learning," in *Proceeding of the International Conference on Real-time Computing and Robotics (RCAR)*, 2019, pp. 112–117.

[77] B. M. Albaba and Y. Yildiz, "Driver modeling through deep reinforcement learning and behavioral game theory," *IEEE Transactions on Control Systems Technology*, vol. 30, no. 2, pp. 885–892, 2021.

[78] L. Bonanni, D. Meli, A. Castellini, and A. Farinelli, "Monte carlo tree search with velocity obstacles for safe and efficient motion planning in dynamic environments," in *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS '25. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2025, p. 371–380.

[79] R. Zhao, K. Wang, Y. Li, Y. Fan, F. Gao, and Z. Gao, "Safe multi-agent deep reinforcement learning for the management of autonomous connected vehicles at future intersections," *IEEE Trans. Parallel Distrib. Syst.*, vol. 36, no. 8, pp. 1744–1761, 2025.

[80] H. Qi and X. Hu, "Monte carlo tree search-based intersection signal optimization model with channelized section spillover," *Transp. Res. Part C Emerg. Technol.*, vol. 106, pp. 281–302, 2019.

[81] E. Sebastián, T. Duong, N. Atanasov, E. Montijano, and C. Sagüés, "Physics-informed multiagent reinforcement learning for distributed multirobot problems," *IEEE Trans. Robot.*, vol. 41, pp. 4499–4517, 2025.

[82] Z. Lin, J. Lan, C. Anagnostopoulos, Z. Tian, and D. Flynn, "Safety-critical multi-agent mcts for mixed traffic coordination at unsignalized intersections," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–15, 2025.

[83] C. F. Hayes, M. Reymond, D. M. Roijers, E. Howley, and P. Mannion, "Risk aware and multi-objective decision making with distributional monte carlo tree search," *arXiv preprint arXiv:2102.00966*, 2021.

[84] A. Skrynnik, A. Andreychuk, K. Yakovlev, and A. Panov, "Decentralized monte carlo tree search for partially observable multi-agent pathfinding," in *Proc. AAAI*, 2024, pp. 17 531–17 540.

[85] G.-P. Antonio and C. Maria-Dolores, "Multi-agent deep reinforcement learning to manage connected autonomous vehicles at tomorrow's intersections," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7033–7043, 2022.

[86] J. Yang, A. Nakhaei *et al.*, "Cm3: Cooperative multi-goal multi-stage multi-agent reinforcement learning," *arXiv preprint arXiv:1809.05188*, 2018.

[87] F. Lateef, M. Kas *et al.*, "Saliency heat-map as visual attention for autonomous driving using generative adversarial network (GAN)," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5360–5373, 2022.

[88] N. Ding, C. Zhang *et al.*, "Saliendet: A saliency-based feature enhancement algorithm for object detection for autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 2624–2635, 2023.

[89] M. Pourabdollah, E. Bjärkvik *et al.*, "Calibration and evaluation of car following models using real-world driving data," in *Proc. IEEE ITSC*. IEEE, 2017, pp. 1–6.

[90] S. Nageshrao, H. E. Tseng, and D. Filev, "Autonomous highway driving using deep reinforcement learning," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 2019, pp. 2326–2331.

[91] Z. Bai, W. Shangguan *et al.*, "Deep reinforcement learning based high-level driving behavior decision-making model in heterogeneous traffic," in *2019 Chinese Control Conference (CCC)*. IEEE, 2019, pp. 8600–8605.

[92] X. Xu, L. Zuo *et al.*, "A reinforcement learning approach to autonomous decision making of intelligent vehicles on highways," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 10, pp. 3884–3897, 2018.

[93] H. Wang, S. Yuan *et al.*, "Tactical driving decisions of unmanned ground vehicles in complex highway environments: A deep reinforcement learning approach," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 235, no. 4, pp. 1113–1127, 2021.

[94] C.-J. Hoel, T. Tram, and J. Sjöberg, "Reinforcement learning with uncertainty estimation for tactical decision-making in intersections," in *2020 IEEE 23rd international conference on intelligent transportation systems (ITSC)*. IEEE, 2020, pp. 1–7.

[95] W. Xiao, Y. Yang, X. Mu, Y. Xie, X. Tang, D. Cao, and T. Liu, "Decision-making for autonomous vehicles in random task scenarios at unsignalized intersection using deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 6, pp. 7812–7825, 2024.

[96] B. Peng, M. F. Keskin *et al.*, "Connected autonomous vehicles for improving mixed traffic efficiency in unsignalized intersections with deep reinforcement learning," *Communications in Transportation Research*, vol. 1, p. 100017, 2021.

[97] A. Pozzi, S. Bae, Y. Choi, F. Borrelli, D. M. Raimondo, and S. Moura, "Ecological velocity planning through signalized intersections: A deep reinforcement learning approach," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 245–252.

[98] D. Quang Tran and S.-H. Bae, "Proximal policy optimization through a deep reinforcement learning framework for multiple autonomous vehicles at a non-signalized intersection," *Applied Sciences*, vol. 10, no. 16, p. 5722, 2020.

[99] Z. Bai, P. Hao, W. Shangguan, B. Cai, and M. J. Barth, "Hybrid reinforcement learning-based eco-driving strategy for connected and automated vehicles at signalized intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 15 850–15 863, 2022.

[100] R. Bautista-Montesano, R. Galluzzi *et al.*, "Autonomous navigation at unsignalized intersections: A coupled reinforcement learning and model predictive control approach," *Transportation research part C: emerging technologies*, vol. 139, p. 103662, 2022.

[101] D. Li, F. Zhu, T. Chen, Y. D. Wong, C. Zhu, and J. Wu, "Coor-plt: A hierarchical control model for coordinating adaptive platoons of connected and autonomous vehicles at signal-free intersections based on deep reinforcement learning," *Transportation Research Part C: Emerging Technologies*, vol. 146, p. 103933, 2023.

[102] S. Kai, B. Wang *et al.*, "A multi-task reinforcement learning approach for navigating unsignalized intersections," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020, pp. 1583–1588.

[103] B. Zhou, Q. Zhou, S. Hu, D. Ma, S. Jin, and D.-H. Lee, "Cooperative traffic signal control using a distributed agent-based deep reinforcement learning with incentive communication," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 8, pp. 10 147–10 160, 2024.

[104] W. X. Hu, H. Ishihara, C. Chen, A. Shalaby, and B. Abdulhai, "Deep reinforcement learning two-way transit signal priority algorithm for optimizing headway adherence and speed," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 8, pp. 7920–7931, 2023.

[105] A. Lombard, A. Noubli, A. Abbas-Turki, N. Gaud, and S. Galland, "Deep reinforcement learning approach for v2x managed intersections of connected vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 7178–7189, 2023.

[106] H. Shu, T. Liu, X. Mu, and D. Cao, "Driving tasks transfer using deep reinforcement learning for decision-making of autonomous vehicles in unsignalized intersection," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 1, pp. 41–52, 2021.

[107] W. Zhao, S. Gong, D. Zhao, F. Liu, N. Sze, M. Quddus, and H. Huang, "A spatial-state-based omni-directional collision warning system for intelligent vehicles," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[108] J. Zhao, T. Qu *et al.*, "A deep reinforcement learning approach for autonomous highway driving," *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 542–546, 2020.

[109] S. Liu, J. Zeng, K. Sreenath, and C. A. Belta, "Iterative convex optimization for model predictive control with discrete-time high-order control barrier functions," in *Proc. Amer. Control Conf. (ACC)*, 2023, pp. 3368–3375.

[110] S. Liu, W. Xiao, and C. Belta, "Feasibility-guaranteed safety-critical control with applications to heterogeneous platoons," in *Proc. IEEE Conf. Decis. Control (CDC)*, 2024, pp. 8066–8073.

[111] S. Liu, Y. Mao, and C. A. Belta, "Safety-critical planning and control for dynamic obstacle avoidance using control barrier functions," in *2025 American Control Conference (ACC)*, 2025, pp. 348–354.

[112] S. Liu, W. Xiao, and C. A. Belta, "Auxiliary-variable adaptive control barrier functions for safety critical systems," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 8602–8607.

[113] H. H. Triharminto, O. Wahyunggoro *et al.*, "A novel of repulsive function on artificial potential field for robot path planning," *International Journal of Electrical and Computer Engineering*, vol. 6, no. 6, p. 3262, 2016.

[114] H. H. Triharminto, O. Wahyunggoro, T. B. Adji, A. Cahyadi, I. Ardiyanto, and Iswanto, "Local information using stereo camera in artificial potential field based path planning," *IAENG International Journal of Computer Science*, vol. 44, no. 3, pp. 316–326, 2017.

[115] R. Chai, Y. Guo, Z. Zuo, K. Chen, H.-S. Shin, and A. Tsourdos, "Cooperative motion planning and control for aerial-ground autonomous systems: Methods and applications," *Prog. Aerosp. Sci.*, vol. 146, p. 101005, 2024.

[116] T. Xia, H. Chen, J. Yang, and Z. Guo, "Geometric field model of driver's perceived risk for safe and human-like trajectory planning," *Transp. Res. Part C: Emerg. Technol.*, vol. 159, p. 104470, 2024.

[117] N. Iwahashi, "Equilibrium selective role coordination for autonomous driving," in *Proc. iCAST*, 2019, pp. 1–8.

[118] G. Liu, B. Xiao, and D. Li, "Game-theory based driving decision algorithm for intersection scenarios considering driver irrationality," in *Proc. CVCI*, 2020, pp. 747–752.

[119] L. Zhang, S. Cheng, Z. Wang, J. Liu, and M. Wang, "A safety-enhanced reinforcement learning-based decision-making and motion planning method for left-turning at unsignalized intersections for automated vehicles," *IEEE Trans. Veh. Technol.*, vol. 73, no. 11, pp. 16 375–16 388, 2024.

[120] J. Zhu, S. Easa, and K. Gao, "Merging control strategies of connected and autonomous vehicles at freeway on-ramps: A comprehensive review," *J. Intell. Connected Vehicles*, vol. 5, no. 2, pp. 99–111, 2022.

[121] J. Xi, F. Zhu, P. Ye, Y. Lv, G. Xiong, and F.-Y. Wang, "Auxiliary network enhanced hierarchical graph reinforcement learning for vehicle repositioning," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 9, pp. 11 563–11 575, Sept. 2024.

[122] H. Liao, Z. Li, K. Zhu, K. Li, and C. Xu, "Sa-tp^2: A safety-aware trajectory prediction and planning model for autonomous driving," *IEEE Trans. Robot.*, vol. 41, pp. 5267–5286, 2025.

[123] A. D. Ames, S. Coogan *et al.*, "Control barrier functions: Theory and applications," in *Proceedings of the 18th European control conference*. IEEE, 2019, pp. 3420–3431.

[124] X. Shi and X. Li, "Empirical study on car-following characteristics of commercial automated vehicles with different headway settings," *Transp. Res. Part C: Emerg. Technol.*, vol. 128, p. 103134, 2021.

[125] R. Coulom, "Efficient selectivity and backup operators in monte-carlo tree search," in *International conference on computers and games*. Springer, 2006, pp. 72–83.

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

[127] Federal Highway Administration, "Next generation simulation (ngsim) vehicle trajectories and supporting data," https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm, 2006, u.S. Department of Transportation.

[128] E. Schmerling, K. Leung, W. Vollprecht, and M. Pavone, "Multimodal probabilistic model-based planning for human-robot interaction," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 3399–3406.

[129] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *ROBOMECH journal*, vol. 1, no. 1, p. 1, 2014.