

Hierarchical Multi-Agent MCTS for Safety-Critical Coordination in Mixed-Autonomy Roundabouts

Zhihao Lin¹, Jianglin Lan^{1,§}, Shuo Liu², *Student Member, IEEE*, Zhen Tian¹, Dezong Zhao¹, *Senior Member, IEEE*, and Chongfeng Wei¹, *Senior Member, IEEE*

Abstract—Navigating unsignalized roundabouts in mixed-autonomy traffic presents significant challenges due to dense vehicle interactions, lane-changing complexities, and behavioral uncertainties of human-driven vehicles (HDVs). This paper proposes a safety-critical decision-making framework for connected and automated vehicles (CAVs) navigating dual-lane roundabouts alongside HDVs. We formulate the problem as a multi-agent Markov Decision Process and develop a hierarchical safety assessment mechanism that evaluates three critical interaction types: CAV-to-CAV (C2C), CAV-to-HDV (C2H), and CAV-to-Boundary (C2B). A key contribution is our lane-specific uncertainty model for HDVs, which captures distinct behavioral patterns between inner and outer lanes, with outer-lane vehicles exhibiting 2.3× higher uncertainty due to less constrained movements. We integrate this safety framework with a multi-agent Monte Carlo Tree Search (MCTS) algorithm that employs safety-aware pruning to eliminate high-risk trajectories while maintaining computational efficiency. The reward function incorporates Shapley value-based credit assignment to balance individual performance with group coordination. Extensive simulation results validate the effectiveness of the proposed approach under both fully autonomous (100% AVs) and mixed traffic (50% AVs + 50% HDVs) conditions. Compared to benchmark methods, our framework consistently reduces trajectory deviations across all AVs and significantly lowers the rate of Post-Encroachment Time (PET) violations, achieving only 1.0% in the fully autonomous scenario and 3.2% in the mixed traffic setting.

Index Terms—Autonomous vehicles, decision making, mixed traffic, Monte Carlo tree search, risk assessment

I. INTRODUCTION

NAVIGATION at roundabouts presents unique challenges for Connected and Autonomous Vehicles (CAVs) [1], particularly in mixed traffic where both CAVs and human-driven vehicles (HDVs) must safely coordinate through complex circular geometries and multiple merging points [2], [3]. The difficulty lies in managing several critical interactions simultaneously while ensuring safety and efficiency in a highly dynamic setting with continuous merging, lane-changing, and exit decisions [4]. This challenge is further amplified by intricate interaction patterns induced by roundabout-specific geometric constraints and yielding rules. Therefore, a comprehensive understanding of both deterministic CAV behaviors and uncertain HDV behaviors is essential [5], [6].

Traditional approaches to roundabout management often rely on rule-based decision-making methods, which generate

conflict-free navigation sequences through preset yielding regulations and lane assignment strategies [7], [8]. These methods struggle to capture the complex decision-making behaviors of human drivers, where yielding involves significant behavioral variability. Conventional priority-based strategies [9] ensure safety by enforcing strict yielding rules, but may substantially increase the requirement for onboard communication quality. More sophisticated rule-based approaches incorporating machine learning and static- and dynamic-constraint-based optimization have been proposed, yet their effectiveness has only been verified in simple single-lane roundabouts [10].

The application of machine learning to roundabout navigation has advanced significantly, yet critical gaps remain between theory and practical deployment [11], [12]. While deep reinforcement learning methods, including multi-agent deep deterministic policy gradient approaches, have shown success in controlled simulations [13]–[15], their performance degrades when faced with the geometric diversity and behavioral uncertainty of real-world roundabouts [16]. The core challenge lies not in learning capability itself, but in the mismatch between underlying assumptions and the requirements of safety-critical traffic coordination [17]–[19]. Specifically, neural network-based approaches exhibit three key limitations [20]. First, the exponential growth of the state-action space with vehicle density makes it computationally prohibitive to cover all safety-critical scenarios during training [11], [12]. Second, limited interpretability prevents real-time safety verification, a non-negotiable requirement in mixed autonomy [13], [15]. Third, the sim-to-real gap is severe in roundabouts due to the interplay between lane-changing decisions and continuous control, leading to unexpected behaviors when policies encounter novel traffic patterns.

In complex environments like roundabouts, vehicle interactions are often modeled using game-theoretic reasoning, aiming to capture both cooperative and competitive behaviors among drivers [21]–[24]. Frameworks such as Nash and Stackelberg games [25] enable multi-agent planning under strategic assumptions. However, real-world scenarios introduce significant deviations from these assumptions—human drivers may act irrationally, interpret gaps inconsistently, or delay exit decisions [26], leading to mismatches between model predictions and actual behavior. Furthermore, the non-uniqueness of game-theoretic solutions [27] can hinder reliable coordination, particularly in multi-lane roundabouts where rapid decision-making is crucial for maintaining safety and efficiency [28]. Beyond learning and game-theoretic approaches, safety-critical control methods have also been widely inves-

¹Zhihao Lin, Jianglin Lan, Zhen Tian, Dezong Zhao, and Chongfeng Wei are with James Watt School of Engineering, University of Glasgow, Glasgow G12 8QQ, United Kingdom.

²Shuo Liu is with Boston University, Brookline, MA, USA.

[§]Corresponding author. Jianglin Lan (e-mail: Jianglin.Lan@glasgow.ac.uk)

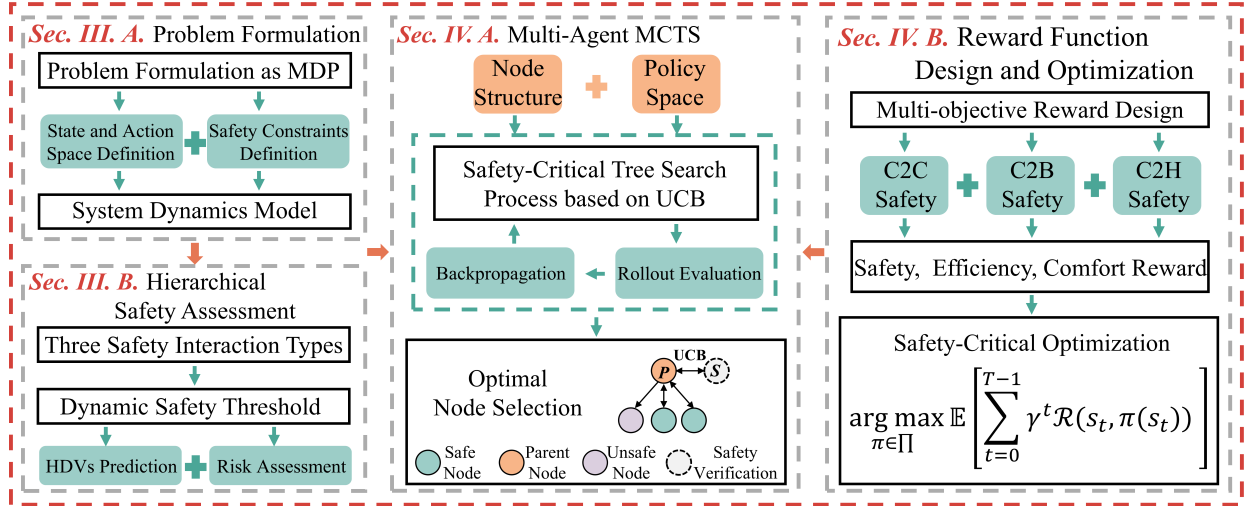


Fig. 1: Overview of the proposed safety-critical decision making framework based on MCTS for roundabout navigation.

tigated to provide formal safety guarantees. Representative examples include Control Barrier Functions (CBFs) [29], [30], reachability analysis [31], and formal methods. In contrast, our work addresses safety through a Monte Carlo Tree Search (MCTS) framework with hierarchical risk assessment, tailored to mixed traffic at unsignalized roundabouts.

MCTS [32] has emerged as a promising approach by marrying the learning-based and game-theoretic methods for interactive navigation [33]. Unlike traditional DRL which requires extensive offline training, MCTS can efficiently explore the action space through online planning, making it particularly suitable for the dynamic and geometry-dependent nature of roundabout environments [34], [35]. The algorithm's inherent ability to balance exploration and exploitation makes it suitable for handling the uncertainties in mixed roundabout traffic, where vehicles are required to continuously make decisions about lane positioning, gap acceptance, and exit timing [36], [37]. However, current MCTS implementations such as [38], [39] often fall short in addressing comprehensive safety considerations and face significant scalability challenges in multi-agent scenarios involving multiple lanes and exit options.

This paper introduces a safety-critical multi-agent MCTS framework for coordinating mixed traffic at dual-lane roundabouts. The framework addresses the unique challenges posed by roundabout geometry, including lane-specific uncertainty modeling, exit proximity effects, and complex interactions between inner and outer lane vehicles. Unlike prior MCTS-based planners that apply uniform safety checks or treat all traffic participants identically, our approach introduces a *structurally differentiated* design: safety assessment, uncertainty modeling, and reward allocation are tailored to the specific interaction type (CAV-to-CAV (C2C) [40], CAV-to-HDV (C2H) [41] and CAV-to-Boundary (C2B) [42]) and lane context, enabling these components to reinforce each other through hierarchical integration rather than operating as independent modules. The main contributions are summarized as follows:

- We propose a safety-critical multi-agent MCTS framework for dual-lane roundabouts, unifying safety pruning, uncertainty modeling, and reward shaping into an

interaction-aware tree search pipeline for mixed traffic.

- We develop a structurally differentiated safety assessment mechanism that handles various interactions via lane-dependent risk metrics and conservative thresholds for human uncertainty, validated through ablation studies.
- We introduce two key structural components for robust coordination: (i) an adaptive HDV prediction model capturing lane-specific uncertainty and exit proximity effects, and (ii) a Shapley value-based reward allocation scheme enabling contribution-aware credit assignment and reducing collision rates in cooperative roundabout navigation.

II. SAFETY-CRITICAL DECISION-MAKING FRAMEWORK

We consider unsignalized dual-lane roundabouts, where CAVs must navigate among other CAVs and HDVs without traffic signals. This task requires handling both predictable autonomous behaviors and uncertainty in human driving patterns.

To model this interaction-rich environment, we cast the problem as a multi-agent Markov Decision Process (MDP), formally defined as $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma \rangle$. Here, $\mathcal{S}$ represents the joint state space capturing critical vehicle-level information such as positions, velocities, and heading angles; $\mathcal{A}$ defines the joint action space, including bounded control inputs for acceleration and steering; $\mathcal{T}$ describes the transition dynamics; $\mathcal{R}$ is the reward function guiding agent behavior; and γ is the discount factor used for long-term planning.

The main complexity arises from three intertwined interaction types: C2C, C2H, and C2B interactions. Within this framework, the goal is to determine an optimal policy π^* that ensures both safety and efficiency in navigation. Denoting $\pi \in \Pi$ as a policy in the admissible space, s_t as the system state at time step t , and $\mathcal{R}(s_t, \pi(s_t))$ as the instantaneous reward, the optimization problem can be expressed as:

$$\pi^* = \arg \max_{\pi \in \Pi} \mathbb{E} \left[\sum_{t=0}^{T-1} \gamma^t \mathcal{R}(s_t, \pi(s_t)) \right]. \quad (1)$$

An overview of our proposed framework is depicted in Fig. 1. The architecture is composed of four key modules.

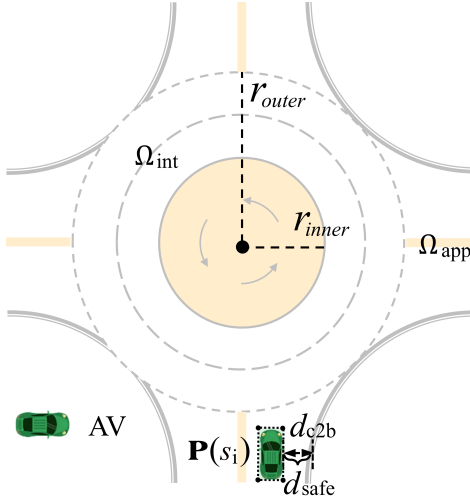


Fig. 2: Illustration of the interaction scenario.

First, the multi-agent MDP model is established, including formal definitions of states, actions, and dynamic models with embedded safety constraints (Sec. III-A). Next, we introduce a hierarchical safety mechanism that considers layered safety constraints across vehicle types and environmental boundaries, using dynamically updated safety thresholds and predictive risk assessment for HDVs (Sec. III-B).

Building on this foundation, we develop a safety-aware multi-agent MCTS algorithm with safety-encoded tree nodes, UCB-based exploration–exploitation balancing, and rollout-based policy extraction with backpropagation (Sec. IV-A). Finally, we formulate a multi-objective reward function that jointly accounts for A2A, A2H, and A2R safety, motion efficiency, and dynamic feasibility, yielding an integrated optimization scheme (Sec. IV-B).

Fig. 2 illustrates the interactive environment surrounding a CAV navigating a roundabout. The figure highlights key spatial regions that shape decision-making, including a central interaction zone where vehicles must account for potential conflicts and trajectory overlaps, and a broader approach area where speed and heading adjustments begin in anticipation of entering the roundabout. The inner circle represents the main circulation area, while the shaded zones near the exits denote heightened interaction regions where yielding, merging, and exiting are most critical. These regions collectively define the spatial context in which CAVs must coordinate with surrounding agents to ensure safe and efficient navigation, and the key notation is summarized in Table I.

III. SAFETY-CRITICAL DECISION MAKING SYSTEM

A. Multi-Agent MDP Formulation for Dual-Lane Roundabout

We formulate the dual-lane roundabout navigation problem involving N CAVs and M HDVs as a multi-agent MDP: $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma \rangle$. For the i -th vehicle, the state vector s_i consists of its polar coordinates (r_i, θ_i) , velocity v_i , heading angle ϕ_i , and lane index l_i . The control input u_i includes acceleration

TABLE I: SUMMARY OF KEY NOTATION

Symbol	Description
<i>State and Control</i>	
$s_i = [r_i, \theta_i, v_i, \phi_i, l_i]^\top$	State vector of vehicle i
$u_i = [a_i, \dot{\phi}_i, \delta_i]^\top$	Control input of vehicle i
$\mathcal{C}, \mathcal{H}$	Set of CAVs / HDVs
N, M	Number of CAVs / HDVs
<i>Safety Assessment</i>	
$d(s_i, s_j)$	Min. distance between vehicles i, j
$d_{\text{safe}}(s_i, s_j)$	Adaptive safe distance
d_{min}	Min. base safe distance
$d_{\text{c2b}}(s_i)$	Min. distance to roundabout boundary
$Q_{\text{risk}}^{\text{cc}}, Q_{\text{risk}}^{\text{ch}}$	C2C / C2H overall risk metric
$Q_{\text{th}}^{\text{cc}}, Q_{\text{th}}^{\text{ch}}$	Safety thresholds for C2C / C2H
C_{ih}	C2H collision probability
<i>HDV Uncertainty Modeling</i>	
$f_{\text{IDM}}(s_h)$	IDM nominal prediction for HDV h
Σ_h^{lane}	Lane-specific covariance matrix
$\mathbf{M}_{\text{lane}}(l_h)$	Lane multiplier matrix
$\mathbf{M}_{\text{exit}}(\mathbf{E}_h)$	Exit proximity multiplier
<i>MCTS and Reward</i>	
N_n, Q_n	Node visit count / cumulative reward
e_{exp}	UCB exploration constant
K	Number of MCTS iterations
ϕ_i	Shapley value of agent i
λ_i^t	Cooperation coefficient

command a_i , steering rate $\dot{\phi}_i$, and lane-changing decision δ_i . The joint state and action spaces are defined as:

$$\mathcal{S} = \prod_{i=1}^{N+M} \left\{ s_i = \begin{bmatrix} r_i \\ \theta_i \\ v_i \\ \phi_i \\ l_i \end{bmatrix} \in \mathbb{R}^5 \mid \begin{array}{l} r_i \in [r_{\text{inner}}, r_{\text{outer}}], \\ \theta_i \in [0, 2\pi), \\ v_i \in [0, v_{\text{max}}], \\ \phi_i \in [-\pi, \pi], \\ l_i \in \{0, 1\} \end{array} \right\}, \quad (2)$$

$$\mathcal{A}_i = \left\{ u_i = [a_i \quad \dot{\phi}_i \quad \delta_i]^\top \in \mathbb{R}^3 \mid \begin{array}{l} |a_i| \leq a_{\text{max}}, \\ |\dot{\phi}_i| \leq \dot{\phi}_{\text{max}}, \\ \delta_i \in \{-1, 0, 1\} \end{array} \right\}, \quad (3)$$

where r_{inner} and r_{outer} are the inner and outer lane radii, $l_i = 0$ represents the inner lane, $l_i = 1$ represents the outer lane, and δ_i denotes the lane-changing decision (-1: move inward, 0: maintain, 1: move outward). The roundabout navigation decisions must satisfy the following safety constraints on state transitions and inter-vehicle distances:

$$\mathcal{S}_{\text{safe}} = \{s \in \mathcal{S} \mid d(s_i, s_j) \geq d_{\text{safe}}, \forall i, j \in \mathcal{C} \cup \mathcal{H}\}, \quad (4)$$

$$d(s_i, s_j) = \min_{\mathbf{p}_i \in \mathbf{P}(s_i), \mathbf{p}_j \in \mathbf{P}(s_j)} \|\mathbf{p}_i - \mathbf{p}_j\|_2,$$

where $\mathbf{s} = [s_1, s_2, \dots, s_{N+M}]^T$, $d(s_i, s_j)$ is the minimum distance between the i -th and j -th vehicles, d_{safe} is the minimum safe distance, $\mathbf{P}(s_i)$ represents the four vertices of the i -th CAV as shown in Fig. 2, $\mathcal{C}$ represents the set of CAVs and $\mathcal{H}$ represents the set of HDVs. Let $\mathbf{s}_t = [s_{1,t}, s_{2,t}, \dots, s_{N+M,t}]^T$ be the joint state vector of all vehicles at time step t , with

TABLE II: SAFETY DISTANCE ADJUSTMENT FACTORS

Factor	Expression	Purpose
α_v	$1 + \beta_v \Delta v_{ij} / v_{\text{ref}}$	Velocity adjustment
α_ϕ	$1 + \beta_\phi \Delta \phi_{ij} / \pi$	Heading adjustment
α_l	$1 + \beta_l l_i - l_j $	Lane difference
α_z	$1 + \sum_{\Omega} \mathbb{I}_{\Omega}(s_i, s_j)$	Zone-based risk

$s_{i,t} = [r_{i,t}, \theta_{i,t}, v_{i,t}, \phi_{i,t}, l_{i,t}]^T$ being the state vector of vehicle $i \in \mathcal{C} \cup \mathcal{H}$.

Navigation decisions must also satisfy constraints over T :

$$st + 1 \in \mathcal{S}_{\text{safe}}, \forall t \in [0, T], |v_i| \leq v_{\text{max}}, d_{c2b}(s_i) \geq d_{\text{min}}, \quad (5)$$

where $d_{c2b}(s_i)$ is the minimum distance to roundabout boundaries, with the minimum allowable value d_{min} . The CAV state transitions can be precisely calculated using control inputs and dynamics as follows:

$$s_{t+1} = \Phi(s_t, \mathbf{u}_t), \quad (6)$$

with the transition function $\Phi(s_t, \mathbf{u}_t) = g_i(s_{i,t}, u_{i,t})_{i=1}^N$, $\mathbf{u}_t = [u_{1,t}, u_{2,t}, \dots, u_{N,t}]^T$ contains the control inputs for all CAVs at time t . $g_i(s_{i,t}, u_{i,t})$ contains the vehicle kinematic model of vehicle $i \in \mathcal{C}$ as follows:

$$g_i(s_{i,t}, u_{i,t}) = \begin{bmatrix} \text{sat}_{[r_{\text{inner}}, r_{\text{outer}}]} \left(r_{i,t} + \frac{v_{i,t} \sin(\phi_{i,t})}{r_{i,t}} \Delta t \right) \\ \text{wrap}_{[0, 2\pi]} \left(\theta_{i,t} + \frac{v_{i,t} \cos(\phi_{i,t})}{r_{i,t}} \Delta t \right) \\ \text{sat}_{[0, v_{\text{max}}]} (v_{i,t} + a_{i,t} \Delta t) \\ \text{wrap}_{[-\pi, \pi]} (\phi_{i,t} + \dot{\phi}_{i,t} \Delta t) \\ \text{LC}(l_{i,t}, \delta_{i,t}, \mathbf{Z}_{i,t}) \end{bmatrix}, \quad (7)$$

where Δt is the time step, $\text{sat}[a, b](\cdot)$ keeps values within bounds, $\text{wrap}[a, b](\cdot)$ handles angle continuity, and $\text{LC}(l_{i,t}, \delta_{i,t}, \mathbf{Z}_{i,t})$ is the lane-changing function that depends on current lane $l_{i,t}$, decision $\delta_{i,t}$, and safety conditions $\mathbf{Z}_{i,t}$. While CAV states are updated by (6), HDV behavior involves inherent uncertainties that grow over time and vary significantly between inner and outer lanes. These uncertainties must be properly characterized for reliable safety assessment in roundabout environments.

B. Hierarchical Safety Assessment for Roundabout Navigation

To ensure safe navigation in dual-lane roundabouts, we develop a hierarchical safety assessment framework that evaluates three critical interaction types. For C2C interactions, we directly use vehicle states $s_i, s_j \in \mathcal{S}$ since their evolution can be precisely characterized by (7). For C2H interactions involving HDVs (denoted by $h \in \mathcal{H}$), we need to predict their future states $\hat{s}_h$ due to behavioral uncertainties. Additionally, C2B interactions are considered for spatial safety constraints within the roundabout geometry.

1) *C2B safety assessment*: The C2B safety assessment focuses on spatial constraints by partitioning the roundabout environment into the interaction area Ω_{int} and approach/exit areas Ω_{app} :

$$\begin{aligned} \Omega_{\text{int}} &= \{s \in \mathcal{S} \mid r_{\text{inner}} \leq r \leq r_{\text{outer}}\}, \\ \Omega_{\text{app}} &= \{s \in \mathcal{S} \mid r < r_{\text{inner}} \vee r > r_{\text{outer}}\}. \end{aligned} \quad (8)$$

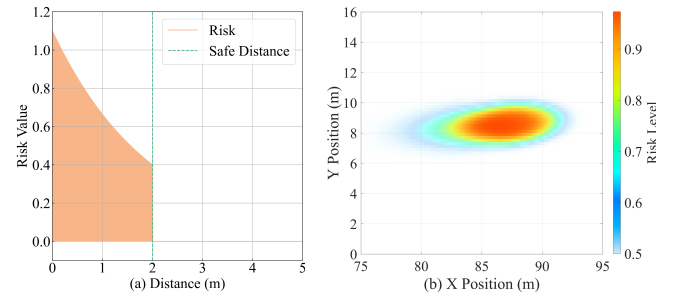


Fig. 3: Safety-critical risk assessment. (a) Distance-based risk. (b) safety risk visualization.

The safety level is evaluated through the minimum distance to roundabout boundaries $d_{c2b}(s_i)$ and its corresponding penalty function $\varphi_{c2b}(d)$, defined as:

$$\begin{aligned} d_{c2b}(s_i) &= \min_{p \in \mathbf{P}(s_i)} \text{distance}(p, \partial\Omega_{\text{circ}} \cup \partial\Omega_{\text{appr}}), \\ \varphi_{c2b}(d) &= \begin{cases} -\infty, & \text{if } d \leq d_{\text{min}}, \\ -\beta \left(\frac{d_{\text{min}}}{d} \right)^2, & \text{if } d_{\text{min}} < d \leq d_{\text{safe}}, \\ 0, & \text{if } d > d_{\text{safe}}, \end{cases} \end{aligned} \quad (9)$$

where ∂ denotes the area boundary and β is a scaling factor.

2) *C2C and C2H safety assessment*: The adaptive safety distance accounts for multiple risk factors via a base value and adjustment terms, as shown in Table II:

$$d_{\text{safe}}(s_i, s_j) = \max\{d_{\text{min}}, \kappa_v |\Delta v_{ij}|\} \cdot \prod_{k \in \mathcal{K}} \alpha_k(s_i, s_j), \quad (10)$$

where d_{min} is the minimum base distance, κ_v is the velocity scaling factor, $\Delta v_{ij} = v_i - v_j$ is the relative velocity, $\Delta \phi_{ij}$ is the relative heading angle, $l_i, l_j \in \{0, 1\}$ denote lane indices, v_{ref} is the reference velocity for normalization, β_k are weighting parameters, and $\mathbb{I}_{\Omega}(\cdot)$ is the indicator function for vehicles in zone Ω . $\mathcal{K} = \{v, \phi, l, z\}$ indexes the adjustment factors in Table II. Fig. 3 visualizes our safety-critical risk assessment framework for CAVs and HDVs navigating roundabouts. Fig. 3(a) presents the foundational distance-based risk function, characterized by an exponential decay curve, which quantifies collision urgency and informs immediate avoidance actions in MCTS decisions. The clearly defined safe distance threshold enables effective and computationally efficient binary safety decisions. Fig. 3(b) provides a spatial risk heat map for a representative roundabout scenario, highlighting high-risk zones at lane-changing and exit-merging points.

Based on the distance measure in (4) and safety threshold in (10), we define two complementary risk metrics. The instantaneous risk captures immediate collision threats:

$$r_{\text{inst}}(s_i, s_j) = \exp\left(\frac{d_{\text{safe}} - d_{\text{min}}}{d_{\text{safe}}}\right) \cdot \left(1 + \frac{|\Delta v_{ij}|}{v_{\text{max}}}\right), \quad (11)$$

where d_{min} is the minimum distance between vehicles and the exponential term penalizes proximity violations while the ve-

locity term accounts for relative speed risks. The temporal risk aggregates future collision probabilities with time discounting:

$$R_T(s_i, s_j) = \frac{1}{T} \sum_{t=1}^T \frac{1}{1+t} \cdot \rho(d_t, d_{\text{safe}}), \quad (12)$$

where $\rho(d, d_{\text{safe}}) = \max\{0, (1 - d/d_{\text{safe}})^2\}$ provides a quadratic penalty when $d < d_{\text{safe}}$. For C2C interactions, the overall safety level is quantified by:

$$Q_{\text{risk}}^{\text{cc}}(s_i, s_j) = w_1^{\text{cc}} r_{\text{inst}}(s_i, s_j) + w_2^{\text{cc}} R_T(s_i, s_j), \quad i, j \in \mathcal{C}, \quad (13)$$

where w_1^{cc} and w_2^{cc} are given weights.

3) *Lane-Specific HDV Uncertainty Modeling*: For C2H interactions, we adopt the IDM [43] as the nominal behavior predictor and embed it within a probabilistic framework to account for behavioral deviations. The lane-specific covariance Σ_h^{lane} captures the distinct uncertainty profiles of inner and outer lanes, where drivers exhibit different levels of yielding tendency and speed variability. The uncertainty model is:

$$P(\hat{s}_h | s_h) = \mathcal{N}(f_{\text{IDM}}(s_h), \Sigma_h^{\text{lane}}(t, l_h, \mathbf{E}_h)), \quad (14)$$

where $f_{\text{IDM}}(s_h)$ captures the nominal human driving behavior, s_h is the current state of HDVs, and $\Sigma_h^{\text{lane}}(t, l_h, \mathbf{E}_h)$ is the lane-specific time-varying covariance matrix:

$$\Sigma_h^{\text{lane}}(t, l_h, \mathbf{E}_h) = \Sigma_{\text{base}}(t) \cdot \mathbf{M}_{\text{lane}}(l_h) \cdot \mathbf{M}_{\text{exit}}(\mathbf{E}_h), \quad (15)$$

where the lane-specific multiplier matrix is:

$$\mathbf{M}_{\text{lane}}(l_h = 0) = \text{diag}(0.3, 0.5, 0.4, 0.6, 0.2), \quad (16)$$

$$\mathbf{M}_{\text{lane}}(l_h = 1) = \text{diag}(1.2, 1.5, 1.3, 1.4, 1.8). \quad (17)$$

The scaling factors are motivated by the geometric asymmetry of roundabout lanes: inner-lane vehicles are constrained by a tighter turning radius that naturally limits lateral and heading deviations, whereas outer-lane vehicles have greater freedom for radial displacement and exit maneuvers [11]. The effectiveness of this lane-specific design over uniform uncertainty is validated in the ablation study (Section V-C).

The exit proximity multiplier $\mathbf{M}_{\text{exit}}(\mathbf{E}_h)$ accounts for increased uncertainty near exits:

$$\mathbf{M}_{\text{exit}}(\mathbf{E}_h) = \mathbf{I} + \sum_{k=1}^{N_{\text{exit}}} \xi_k \exp\left(-\frac{|\theta_h - \theta_{\text{exit},k}|^2}{2\sigma_{\text{exit}}^2}\right) \mathbf{J}_k, \quad (18)$$

where N_{exit} is the number of exits, $\theta_{\text{exit},k}$ is the angular position of the k -th exit, ξ_k is the exit influence factor, σ_{exit} is the exit influence range, and $\mathbf{J}_k$ is a diagonal matrix containing state-specific uncertainty amplification factors for the k -th exit. The base covariance matrix $\Sigma_{\text{base}}(t)$ is defined as:

$$\Sigma_{\text{base}}(t) = \begin{bmatrix} \sigma_r^2 t + \epsilon_r^2 t^2 & 0 & \rho_{rv} \sigma_r \sigma_v t & 0 & 0 \\ 0 & \sigma_\theta^2 t + \epsilon_\theta^2 t^2 & 0 & \rho_{\theta\phi} \sigma_\theta \sigma_\phi t & 0 \\ \rho_{rv} \sigma_r \sigma_v t & 0 & \sigma_v^2 & 0 & 0 \\ 0 & \rho_{\theta\phi} \sigma_\theta \sigma_\phi t & 0 & \sigma_\phi^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_l^2 \end{bmatrix}, \quad (19)$$

where $(\sigma_r, \sigma_\theta, \epsilon_r, \epsilon_\theta)$ capture position uncertainty growth including the radial uncertainty measured by $\sigma_r^2 t + \epsilon_r^2 t^2$ and angular

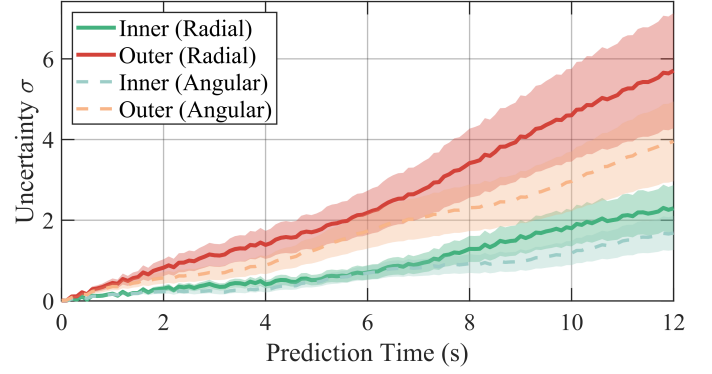


Fig. 4: Oscillating temporal uncertainty evolution showing the growth of prediction uncertainty over time for inner and outer lanes in both radial and angular dimensions.

uncertainty measured by $\sigma_\theta^2 t + \epsilon_\theta^2 t^2$, while $(\sigma_{v,\phi}, \rho_{rv}, \rho_{\theta\phi})$ model velocity, heading, and state correlations.

Fig. 4 illustrates the temporal evolution of prediction uncertainty for HDVs in inner and outer lanes. Over a 12-second horizon, outer-lane radial uncertainty (red line) reaches approximately 5.8σ compared to 2.5σ for the inner lane (green line), while angular uncertainties show similar disparities (4.0σ outer vs. 2.0σ inner). The shaded regions represent 75%–125% confidence envelopes. This lane-specific uncertainty growth validates our hierarchical safety framework, where outer-lane vehicles require larger safety margins due to their less constrained movement patterns.

We bound HDVs' reachable state space as:

$$\mathcal{S}_h^t = \left\{ \hat{s}_h \in \mathbb{R}^5 \left| \begin{array}{l} \|\mathbf{p}_h - \mathbf{p}_h(t)\| \leq (v_{\text{max}} + \sigma_v)t, \\ |v_h| \leq v_{\text{max}} + 2\sigma_v, \\ |\phi_h| \leq \pi, \\ l_h \in \{0, 1\} \end{array} \right. \right\}. \quad (20)$$

The collision probability for C2H interactions is:

$$C_{ih} = \int_{\hat{s}_h \in \mathcal{S}_h^t} \psi(\hat{s}_i, \hat{s}_h) \cdot \mathcal{N}(f_{\text{IDM}}(s_h), \Sigma_h^{\text{lane}}) d\hat{s}_h, \quad (21)$$

where $\psi(\hat{s}_i, \hat{s}_h) = \mathbb{I}(d(\hat{s}_i, \hat{s}_h) < d_{\text{safe}}(s_i, s_h))$ is the collision indicator function. The safety level for C2H interactions is:

$$Q_{\text{risk}}^{\text{ch}}(s_i, s_h) = w_1^{\text{ch}} r_{\text{inst}}(s_i, \hat{s}_h) + w_2^{\text{ch}} R_T(s_i, s_h) + w_3^{\text{ch}} C_{ih}, \quad (22)$$

where $i \in \mathcal{C}$, $h \in \mathcal{H}$, and $w_k^{\text{ch}}, k \in [1, 3]$ are given weights.

IV. MULTI-AGENT MCTS SOLUTION APPROACH

A. Multi-Agent MCTS for Roundabout Navigation

Building on the safety assessment framework, we propose a structured tree search approach where the risk assessment functions in (13) and (22) are used to evaluate safety at each node and prune unsafe nodes that exceed predefined safety thresholds ($Q_{\text{th}}^{\text{cc}}$ for C2C interactions and $Q_{\text{th}}^{\text{ch}}$ for C2H interactions). Let $\mathcal{T}$ be the search tree whose node $n \in \mathcal{T}$ is defined as: $n = (d_n, p_n, C_n, N_n, Q_n, u_n, \xi_n)$. where d_n is the node depth, p_n is the parent node, C_n is the set of child nodes, N_n is the visit count, Q_n is the cumulative reward value, and u_n is the action leading to this node. A node is considered

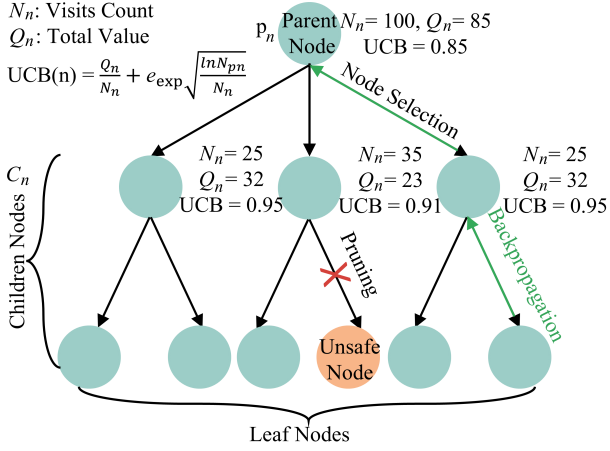


Fig. 5: Illustration of the safety-critical MCTS framework. Each node stores the visit count N_n , value estimate Q_n , and UCB score. Nodes identified as unsafe (shown in orange) are pruned during the safety validation stage. The green path indicates the backpropagation process.

safe ($\xi_n = \text{safe}$) if the safety metrics satisfy: $Q_{\text{risk}}^{\text{cc}} \leq Q_{\text{th}}^{\text{cc}}$ for C2C interactions, $Q_{\text{risk}}^{\text{ch}} \leq Q_{\text{th}}^{\text{ch}}$ for C2H interactions, and $d_{\text{c2b}} \geq d_{\text{min}}$. Otherwise, it is marked as unsafe ($\xi_n = \text{unsafe}$). The joint policy space $\mathcal{A}_{\text{joint}} = \bigotimes_{i=1}^{N+M} \mathcal{P}_i$ represents the Cartesian product of policy sets.

We set the safety thresholds to $Q_{\text{th}}^{\text{cc}} = 0.8$ (C2C) and $Q_{\text{th}}^{\text{ch}} = 0.6$ (C2H), using a more conservative margin for human-driven interactions due to higher prediction uncertainty. These values are calibrated in preliminary experiments to balance safety and traffic efficiency, and the sensitivity to d_{min} is further studied in Section V-C. Following standard surrogate safety analysis, we adopt a PET threshold of 1.0 s [44].

The safety-critical multi-agent MCTS algorithm uses a depth-first search strategy with safety validation at each node expansion. The search process evaluates nodes recursively:

$$\text{Search}(n_t) = \begin{cases} \text{Expand}(n_t) \cup \text{Rollout}(n_t), & \text{if new \& safe,} \\ \text{Search}(\text{UCB}(n_t)), & \text{if visited,} \\ \text{Terminate,} & \text{if unsafe.} \end{cases}$$

The selection of nodes for expansion is governed by the UCB formula, which balances exploration and exploitation:

$$\text{UCB}(n) = \frac{Q_n}{N_n} + e_{\text{exp}} \sqrt{\frac{\ln N_{p_n}}{N_n}}, \quad (23)$$

where e_{exp} is the exploration constant balancing exploitation and exploration, and N_{p_n} is the visit count of the parent node. Fig. 5 illustrates the structure of an MCTS guided by the UCB strategy. Leaf nodes represent terminal or intermediate states, with one marked as an “Unsafe Node” indicating trajectories that violate safety constraints.

B. Roundabout-Specific Reward Function Design

Roundabout navigation involves dense, asymmetric interactions that require both individual planning and group-level coordination. To accommodate these challenges, we propose

a structured reward function that integrates four key aspects: safety, efficiency, comfort, and lane-specific behavior.

1) *Group-aware reward formulation* We adopt a coalition-based design using the Shapley value ϕ_i to distribute the global reward among cooperative vehicles:

$$\phi_i(v) = \sum_{c \subset C \cup \mathcal{H}, i \in c} \frac{(|c|-1)!(n-|c|)!}{n!} [v(c) - v(c \setminus \{i\})],$$

where $v(c) = \sum_{j \in c} \mathcal{R}_j(\mathbf{s}_t, \mathbf{u}_t)$ denotes the coalition value at time t and $v(c \setminus \{i\})$ denotes the coalition value excluding CAV_{*i*}, such that the difference $v(c) - v(c \setminus \{i\})$ reflects how much value agent i brings to coalition c . The coefficient $\frac{(|c|-1)!(n-|c|)!}{n!}$ represents the probability that CAV_{*i*} is the last to join group c when CAVs join in a random order. This helps average i 's contribution across all possible team formations.

The global reward at each timestep is then defined as:

$$\mathcal{R}(\mathbf{s}_t, \mathbf{u}_t) = \sum_{i \in C} \phi_i \cdot \mathcal{R}_i(\mathbf{s}_t, \mathbf{u}_t). \quad (24)$$

Each CAV's individual reward $\mathcal{R}_i$ balances its own performance and the impact on others:

$$\mathcal{R}_i = \frac{Q_i^{\text{self}} + \lambda_i^t Q_i^{\text{other}}}{1 + \lambda_i^t (N - 1)}, \quad (25)$$

where $Q_i^{\text{self}} = Q_i^{\text{safety}} + Q_i^{\text{eff}} + Q_i^{\text{comfort}} + Q_i^{\text{lane}}$, and $Q_i^{\text{other}} = \sum_{j \neq i} (Q_j^{\text{safety}} + Q_j^{\text{eff}} + Q_j^{\text{lane}})$. The cooperation coefficient $\lambda_i^t \in [0, 1]$ is fixed for CAVs and empirical estimates based on typical driving patterns for HDVs (e.g., $\lambda_i^t = 0.4$).

2) *Reward components*: Each CAV's self-reward comprises four components that balance safety, efficiency, comfort, and lane-specific behaviors:

$$\begin{aligned} Q_i^{\text{safety}} &= -w_{\text{c2c}} Q_{\text{risk}}^{\text{cc}}(s_i) - w_{\text{c2h}} Q_{\text{risk}}^{\text{ch}}(s_i) - w_{\text{c2b}} \varphi_{\text{c2b}}, \\ Q_i^{\text{eff}} &= -w_v (v_i - v_{\text{des}})^2 - w_p \|p_i - p_{\text{ref}}\|^2, \\ Q_i^{\text{comfort}} &= -w_a |a_i|^2 - w_\phi |\dot{\phi}_i|^2, \\ Q_i^{\text{lane}} &= Q_i^{\text{position}}(l_i) - w_{\text{trans}} |\delta_i|^2 + Q_i^{\text{exit}}(\theta_i), \end{aligned} \quad (26)$$

where the safety terms directly incorporate the risk metrics from (9), (13) and (22), with φ_{c2b} penalizing boundary violations. The efficiency component encourages desired velocity v_{des} and path tracking to reference trajectory p_{ref} . The comfort term penalizes excessive acceleration a_i and yaw rate $\dot{\phi}_i$.

3) *Summary* This hierarchical reward design enables agents to balance local objectives with socially-aware behavior in the roundabout, while explicitly incorporating safety, comfort, and intent-awareness. The cooperation coefficient λ_i modulates each agent's autonomy versus altruism, and the Shapley-based allocation ensures fair credit assignment in joint outcomes.

C. Optimization Policy

For HDVs in the roundabout, we use the IDM model with lane-specific parameters as described in Section III-B. During MCTS rollouts, HDV actions are sampled from the predictive distribution in (14), with cooperation levels λ_h based on typical driving patterns. The optimality of our MCTS solution

for roundabout navigation can be formally characterized by the following optimization objective:

$$\mathbf{u}^* = \arg \max_{\mathbf{u} \in \mathcal{A}_{\text{joint}}} \mathbb{E} \left[\sum_{t=0}^{T-1} \gamma_r \sum_{i \in \mathcal{C}} \phi_i \mathcal{R}_i(\mathbf{s}_t, \mathbf{u}(\mathbf{s}_t)) \middle| \mathbf{u} \right], \quad (27)$$

subject to the following constraints at each time step t :

Safety Constraints:

$$\begin{aligned} \text{(C2C)} \quad & Q_{\text{risk}}^{\text{cc}}(s_i, s_j) \leq Q_{\text{th}}^{\text{cc}}, \quad \forall i, j \in \mathcal{C} \\ \text{(C2H)} \quad & Q_{\text{risk}}^{\text{ch}}(s_i, s_h) \leq Q_{\text{th}}^{\text{ch}}, \quad \forall i \in \mathcal{C}, h \in \mathcal{H} \\ \text{(C2B)} \quad & d_{\text{c2b}}(s_i) \geq d_{\text{min}}, \quad \forall i \in \mathcal{C} \end{aligned} \quad (28)$$

Dynamic Constraints:

$$\begin{aligned} \text{(Velocity)} \quad & v_i \in [0, v_{\text{max}}], \quad \forall i \in \mathcal{C} \\ \text{(Acceleration)} \quad & |a_i| \leq a_{\text{max}}, \quad \forall i \in \mathcal{C} \\ \text{(Steering)} \quad & |\dot{\phi}_i| \leq \dot{\phi}_{\text{max}}, \quad \forall i \in \mathcal{C} \end{aligned}$$

This formulation provides a comprehensive theoretical characterization of our roundabout navigation approach. The reward function $\mathcal{R}_i(\mathbf{s}_t, \mathbf{u}(\mathbf{s}_t))$ incorporates Shapley value-based contribution quantification. The lane-specific modeling through matrices $\mathbf{M}_{\text{lane}}(l_h)$ and $\mathbf{M}_{\text{exit}}(\mathbf{E}_h)$ enables precise characterization of uncertainty variations between inner and outer lanes, while the exit proximity effects provide realistic behavioral modeling for roundabout scenarios. Following the kinematic model in (7), our framework addresses roundabout conflicts by integrating safety assessment, reward design, and MCTS-based policy search, ensuring safe and efficient navigation through the balance of individual and cooperative behaviors.

D. Computational Complexity Analysis

The computational complexity of the proposed multi-agent MCTS-based planning algorithm arises from three main components: joint-action tree expansion, safety validation, and rollout simulation. Tree expansion considers joint actions over N CAVs, each with $|\mathcal{A}|$ possible actions, resulting in a branching factor of $|\mathcal{A}|^N$. At each expanded node, safety validation involves pairwise CAV interactions ($O(N^2)$), CAV-to-HDV checks ($O(NM)$), and lane-level uncertainty evaluations, typically contributing $O(NM \cdot d^2)$ complexity, where d is the state dimension.

Each new node further triggers a rollout simulation up to a depth d_{max} , with per-step computations dominated by state transitions, reward evaluations, and uncertainty updates, also scaling with $O(NM \cdot d^2)$.

As a result, the worst-case overall complexity across K MCTS iterations can be coarsely estimated as:

$$O(K \cdot |\mathcal{A}|^N \cdot d_{\text{max}} \cdot NM \cdot d^2), \quad (29)$$

which grows exponentially with the number of agents.

In practice, however, the actual computational cost is significantly lower due to three factors: 1) safety pruning that eliminates unsafe branches early; 2) UCB-guided selective search that limits unnecessary expansions; and 3) matrix-based parallel implementation of safety and reward computations. While the lane-specific uncertainty modeling adds overhead,

TABLE III: COMPUTATIONAL COMPLEXITY COMPARISON

Method	Complexity	Operations
Joint exhaustive search	$ \mathcal{A} ^{N \cdot H}$	$\approx 10^{85}$
Game-theoretic Nash	$O(\mathcal{A} ^{NH} \cdot I_{\text{Nash}})$	$> 10^{87}$
MCTS (no pruning)	$K \cdot \mathcal{A} \cdot N \cdot H$	$\approx 3.2 \times 10^5$
MCTS + safety pruning	$(1 - \rho) \cdot K \cdot \mathcal{A} \cdot N \cdot H$	$\approx 2.2 \times 10^5$

it substantially improves prediction robustness in congested or uncertain roundabout scenarios.

To provide practical runtime context, we report measured per-step computation times from our MATLAB implementation. In our most complex scenario (8-vehicle symmetric intersection), the framework plans for $N = 4$ decision-making CAVs while the remaining HDVs are modeled as reactive agents with negligible planning overhead. The average planning time is approximately 58 ms per decision step, well within a typical planning cycle of $\Delta t = 0.2$ s. Given the linear scaling with N in our MCTS framework, we estimate the current implementation can support approximately $N = 6$ – 8 planning agents in real time. As shown in Table III, safety pruning reduces the effective search space by approximately $1.4\times$ compared to unpruned MCTS, while UCB-guided selection further concentrates search effort on promising branches. A C++ implementation with parallelized rollouts would further extend this scalability envelope, left for future work.

V. EXPERIMENTAL EVALUATION

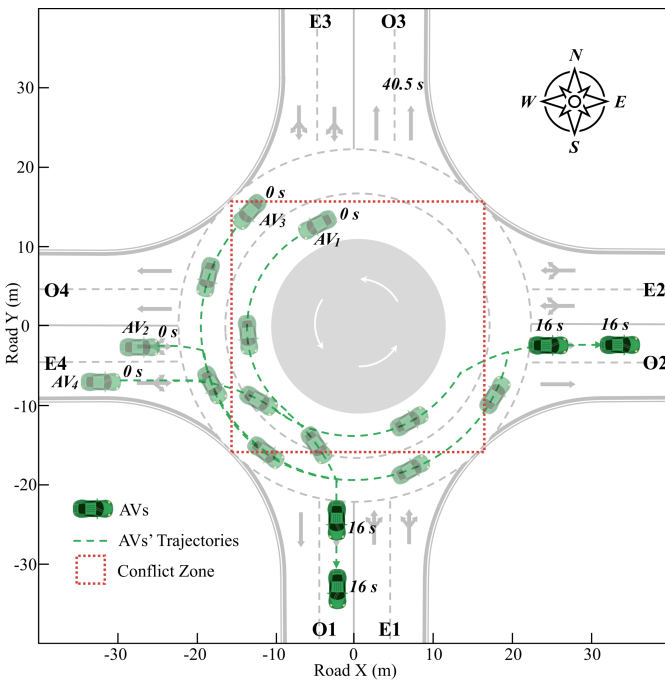
Simulations are conducted in MATLAB 2024a to evaluate the proposed approach for safe and efficient autonomous driving at a signal-free, dual-lane roundabout. We compare the proposed method with several advanced optimization algorithms, including the Stackelberg game approach [45] and the Nash equilibrium method [27].

A. Case 1: Dual-lane Roundabout (ROP = 100%)

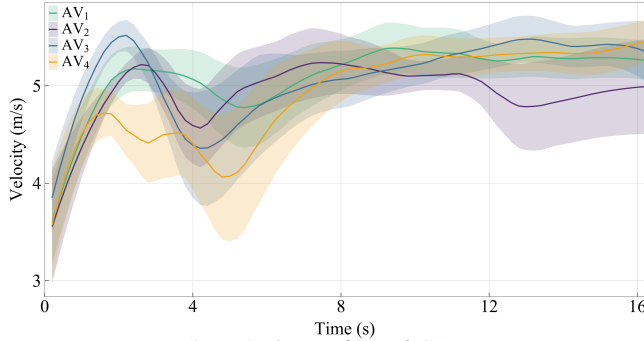
The experimental evaluation begins with a baseline scenario featuring a 100% rate of penetration (ROP) at a signal-free, dual-lane roundabout. As depicted in Fig. 6(a), four CAVs simultaneously approach the roundabout from different directions, creating a complex multi-agent coordination challenge. The conflict zone, highlighted by the red dashed box, marks the critical region where vehicle paths intersect and potential collisions may occur.

Fig. 6(b) illustrates the velocity profiles of the CAVs under our proposed method. Solid lines indicate the mean velocities, while the shaded bands represent the 95% confidence intervals. The results show that the vehicles maintain stable speeds ranging from 3 to 5 m/s, requiring only minimal speed modulation for safe interaction. This highlights the ability of our MCTS-based framework to ensure both safety and motion smoothness without unnecessary deceleration.

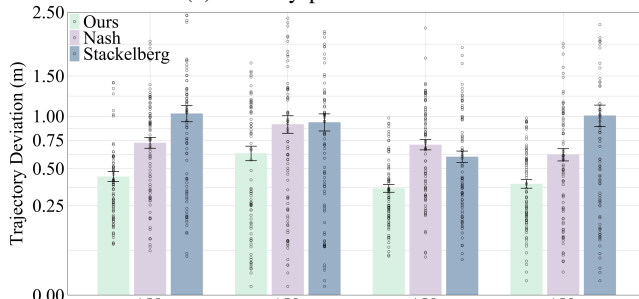
Further insights are provided in Fig. 6(c), which compares trajectory deviations among our method and two baselines: Nash and Stackelberg. Across all CAVs, our approach consistently yields smaller deviations from the reference trajectories. This significant improvement in tracking accuracy reflects the



(a) Illustration of MCTS simulation at a dual-lane roundabout.



(b) Velocity profiles of CAVs.

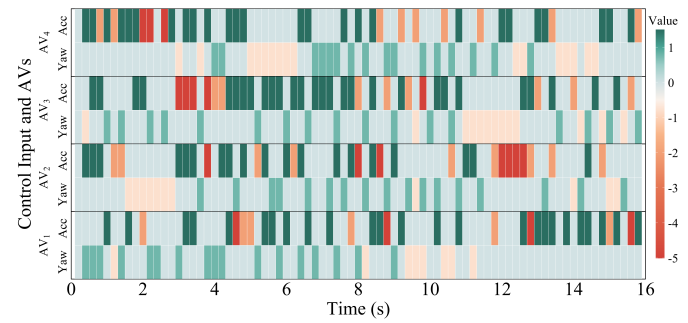


(c) Comparison of trajectory deviations with advanced approaches.

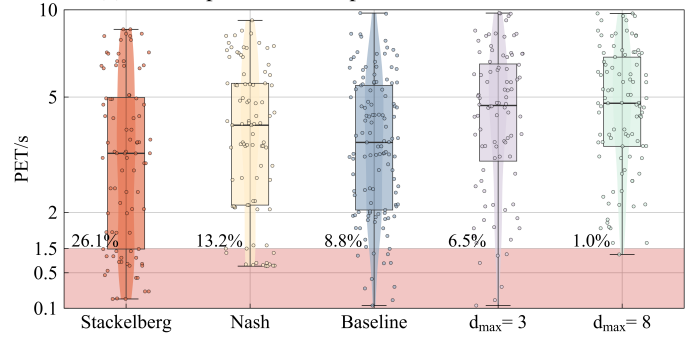
Fig. 6: Performance Analysis in Case 1 (ROP = 100%).

effectiveness of our planning strategy in preserving intended paths while adhering to safety constraints. Additionally, the reduced size of the error bars indicates greater behavioral consistency and robustness.

The decision-making process is further analyzed through the control input heatmap in Fig. 7(a), which reveals the temporal evolution of acceleration and yaw rate commands for each vehicle. The predominant green coloring indicates that most control actions are moderate, with occasional stronger interventions (darker colors) occurring primarily during critical in-roundabout driving phases. This pattern demonstrates the



(a) Heatmap of control inputs for controlled CAVs.



(b) PET comparison.

Fig. 7: Analysis of decision-making and safety performance. (a) Variations in control inputs. (b) PET distributions and violations: benchmarks, Baseline (without adaptive risk evaluation), and our methods with $d_{\max}$ of 3 and 8 (ROP = 100%).

TABLE IV:

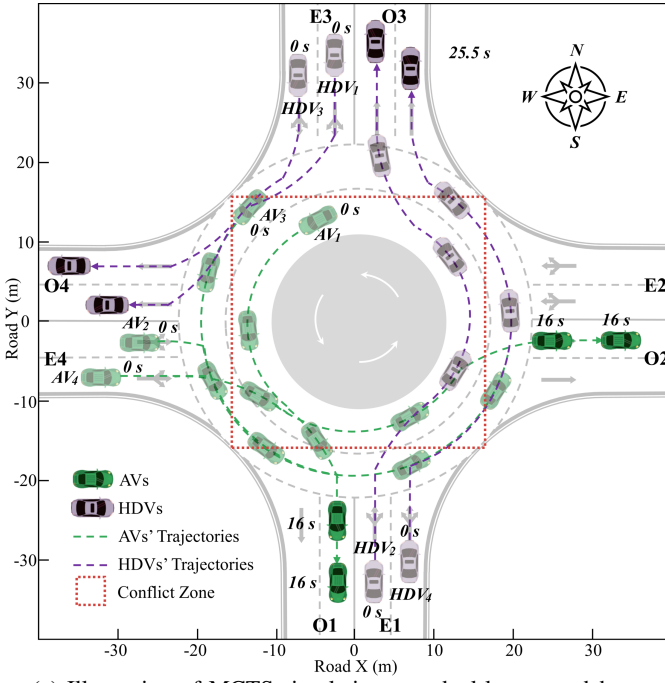
COMPARISON OF ALGORITHM PERFORMANCES IN CASE 1.

Methods	Average Speed (m/s)	Average Trajectory Deviation (m)	Collision Rate (%)
Stackelberg	5.22 ± 1.34	0.89 ± 0.84	16.0
Nash	5.16 ± 1.01	0.74 ± 0.59	11.0
Baseline	5.44 ± 1.10	0.65 ± 0.47	13.0
$d_{\max} = 3$	5.27 ± 1.03	0.57 ± 0.38	4.0
$d_{\max} = 8$	5.31 ± 0.89	0.45 ± 0.39	0.0

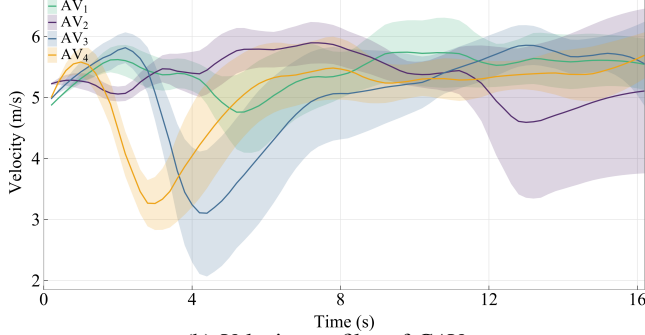
framework's ability to generate comfortable trajectories while responding appropriately to dynamic interaction scenarios.

Fig. 7(b) presents a quantitative assessment of safety performance based on Post-Encroachment Time (PET) distributions. The proposed method with a maximum tree depth of $d_{\max} = 8$ delivers the best outcomes, with no instances of PET falling below the critical threshold of 1.0 second. In comparison, the baseline approach exhibits 8.8% violations, while the Stackelberg and Nash methods perform worse, with violation rates of 26.1% and 13.2%, respectively. These results underscore the effectiveness of our framework, which leverages dynamic risk evaluation and adaptive safety constraints to proactively manage both current and anticipated interactions.

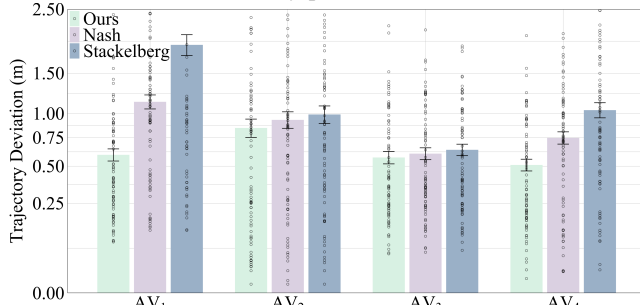
Table IV demonstrates the superior performance of our proposed method with $d_{\max} = 8$, which achieves the lowest collision rate (0.0%) and the smallest trajectory deviation (0.45 ± 0.39 m), while maintaining competitive driving speed (5.31 ± 0.89 m/s). In contrast, Stackelberg and Nash baselines exhibit substantially higher collision rates (16.0%



(a) Illustration of MCTS simulation at a dual-lane roundabout.



(b) Velocity profiles of CAVs.



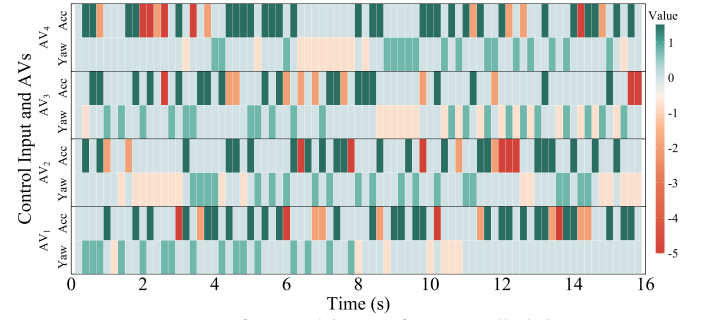
(c) Comparison of trajectory deviations with advanced approaches.

Fig. 8: Performance Analysis in Case 2 (ROP = 50%).

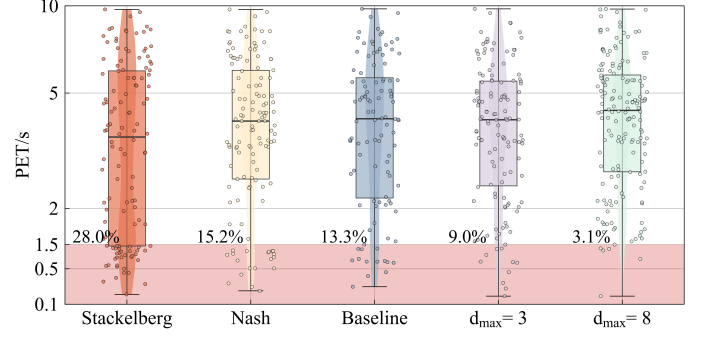
and 11.0%), indicating their limited robustness in complex interaction scenarios. Increasing $d_{\max}$ consistently improves safety, confirming the benefit of incorporating larger safety margins in roundabout coordination.

B. Case 2: Dual-lane Roundabout (ROP = 50%)

To assess the robustness of our proposed framework under mixed traffic conditions, we conduct experiments with a 50% CAV penetration rate, where CAVs and HDVs coexist and interact at the signal-free, dual-lane Roundabout. As illustrated in Fig. 8(a), the test setup features four CAVs and four



(a) Heatmap of control inputs for controlled CAVs.



(b) PET comparison.

Fig. 9: Analysis of decision-making and safety performance. (a) Variations in control inputs. (b) PET distributions and violations: benchmarks, Baseline (without adaptive risk evaluation), and our methods with $d_{\max}$ of 3 and 8 (ROP = 50%).

TABLE V:

COMPARISON OF ALGORITHM PERFORMANCES IN CASE 2.

Methods	Average Speed (m/s)	Average Trajectory Deviation (m)	Collision Rate (%)
Stackelberg	5.25 ± 1.31	1.14 ± 1.06	33.0
Nash	5.12 ± 1.24	0.85 ± 0.68	29
Baseline	4.83 ± 0.82	0.81 ± 0.65	17
$d_{\max} = 3$	5.16 ± 0.95	0.76 ± 0.62	5.0
$d_{\max} = 8$	5.23 ± 0.94	0.63 ± 0.59	2.0

HDVs entering the roundabout from different entrance ports, introducing additional coordination complexity due to the unpredictable behavior of human drivers.

The velocity profiles depicted in Fig. 8(b) highlight the effectiveness of our method in managing such complex roundabout. In contrast to the full-CAV scenario, the velocity trajectories exhibit greater variability—ranging between 3–5 m/s—and wider confidence intervals, reflecting the influence of HDV-induced uncertainty. Despite this, the velocity transitions remain relatively smooth, suggesting that our MCTS-based framework is capable of adapting to human behaviors while preserving both safety and traffic flow efficiency.

As shown in Fig. 8(c), the trajectory deviation analysis further confirms the advantages of our approach over baseline methods. Our framework consistently yields smaller deviations from reference paths, demonstrating improved path-tracking accuracy even in the presence of mixed vehicle types. These results underscore the effectiveness of our risk-aware decision-making mechanism and its adaptability in complex, uncertain

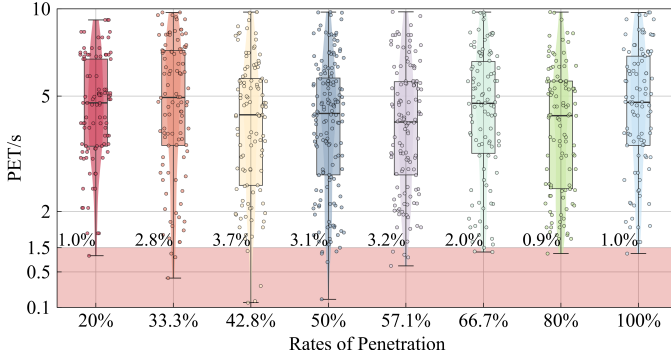


Fig. 10: PET violations under different CAV penetration rates.

traffic environments.

The control input heatmap in Fig. 9(a) reveals a greater diversity and frequency of CAV behavior adjustments compared to the fully autonomous scenario. The emergence of more pronounced light and dark regions indicates that CAVs engage in more dynamic control actions in response to the unpredictability of human-driven vehicles. This pattern reflects the adaptive nature of our framework, which effectively modulates between assertive and conservative behaviors to ensure safe and efficient interaction with HDVs.

The safety performance assessment based on PET distributions, as illustrated in Fig. 9(b), underscores the difficulties posed by mixed traffic scenarios. Our method with $d_{\max} = 8$ achieves the most favorable safety outcome, with only 3.1% of instances violating the PET threshold. In contrast, the baseline method results in a notably higher violation rate of 13.3%, while the Stackelberg and Nash strategies perform worse still, with violation rates of 28.0% and 15.2%, respectively. These findings emphasize the complexity of maintaining safe interactions when human-driven vehicles are involved.

As shown in Table V, our MCTS-based method with $d_{\max} = 8$ achieves the best overall performance in Case 2, yielding the lowest collision rate (2.0%) and the smallest trajectory deviation (0.63 ± 0.59 m), while maintaining a high average speed (5.23 ± 0.94 m/s). In comparison, Stackelberg and Nash exhibit substantially higher collision rates (33.0% and 29.0%), highlighting their limited robustness in mixed traffic with human-driven uncertainty. These results confirm that increasing $d_{\max}$ significantly improves safety and motion consistency in complex roundabout interactions.

To evaluate safety under varying CAV penetration rates, experiments were conducted with penetration rates ranging from 20% to 100%. Fig. 10 shows the PET distributions, revealing improved safety as CAV penetration increases. At low penetration rates (20%-33.3%), PET distributions are more variable, with low violation rates (1.0%-3.7%) due to the unpredictability of HDVs. In the medium range (42.8%-57.1%), violation rates initially rise to 3.7% at 42.8%, then stabilize around 3.2%, reflecting the complexity of mixed traffic. At high penetration rates (66.7%-100%), violation rates drop to 0.9%, and PET distributions narrow, demonstrating consistent safety margins due to the dominance of CAV behavior.

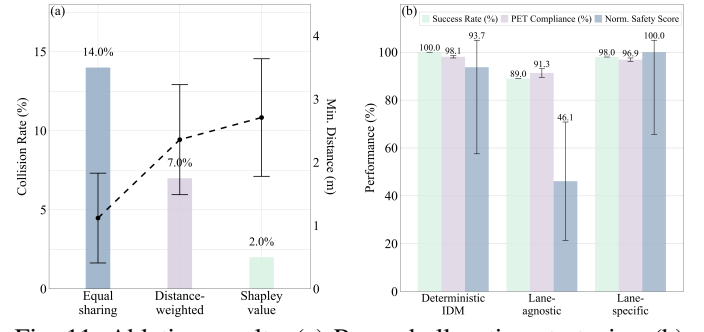


Fig. 11: Ablation results. (a) Reward allocation strategies. (b) HDV uncertainty models.

TABLE VI: SENSITIVITY AND ABLATION ANALYSIS

Setting	Collision Rate (%)	PET Viol. Rate (%)	Traj. Dev. (m)	Min. Distance (m)
<i>Min. Safe Distance $d_{\min}$ (Default $d_{\min} = 2.0$ m)</i>				
$d_{\min} = 1.0$ m	9.0	8.2 ± 0.9	0.56 ± 0.39	1.31 ± 0.74
$d_{\min} = 2.0$ m	2.0	3.1 ± 0.7	0.73 ± 0.61	2.72 ± 0.93
$d_{\min} = 3.0$ m	1.0	2.4 ± 0.5	1.42 ± 0.85	3.61 ± 1.12
<i>Reward Allocation Strategy (Default Shapley)</i>				
Equal sharing	14.0	10.2 ± 2.1	1.63 ± 0.85	1.12 ± 0.71
Distance-weighted	7.0	5.3 ± 1.6	1.12 ± 0.78	2.36 ± 0.87
Shapley value	2.0	3.1 ± 0.7	0.73 ± 0.61	2.71 ± 0.93
<i>HDV Uncertainty Model (Default Lane-Specific)</i>				
Deterministic IDM	0.0	1.9 ± 0.5	0.52 ± 0.43	2.54 ± 0.98
Lane-agnostic	11.0	8.7 ± 1.8	1.66 ± 0.82	1.25 ± 0.67
Lane-specific	2.0	3.1 ± 0.7	0.73 ± 0.61	2.71 ± 0.93

C. Sensitivity and Ablation Analysis

To evaluate the robustness of key design choices and validate the contributions of individual components, we conduct sensitivity and ablation studies. The quantitative results are summarized in Table VI, with visual comparisons for the two core algorithmic contributions presented in Fig. 11.

1) *Sensitivity to Minimum Safe Distance*: We vary the minimum safe distance threshold $d_{\min} \in \{1.0, 2.0, 3.0\}$ m. As shown in Table VI, a smaller threshold ($d_{\min} = 1.0$ m) permits closer interactions, resulting in elevated collision rates (9.0%) and reduced minimum distances (1.31 m). Conversely, a larger threshold ($d_{\min} = 3.0$ m) yields slightly lower collision rates (1.0%) but introduces excessive conservatism, increasing trajectory deviations to 1.42 m. The default setting $d_{\min} = 2.0$ m achieves an effective balance between safety and efficiency, maintaining a low collision rate (2.0%) with moderate trajectory deviations (0.73 m).

2) *Reward Allocation Strategy*: We compare three reward allocation strategies: equal sharing, distance-weighted sharing, and Shapley value allocation. As illustrated in Fig. 11(a), the Shapley value method significantly outperforms both alternatives. Equal sharing, which ignores individual contributions, leads to the highest collision rate (14.0%) and PET violation rate (10.2%), as agents lack incentive for cooperative behavior. Distance-weighted sharing improves upon equal sharing but still yields a 7.0% collision rate. The Shapley value method reduces collision rates to 2.0% and achieves the best minimum distance (2.71 m), demonstrating that fairly attributing

each agent's marginal contribution to the coalition reward is essential for effective multi-agent coordination.

3) *HDV Uncertainty Model*: We evaluate three uncertainty modeling approaches for human-driven vehicles: deterministic IDM, lane-agnostic uncertainty, and lane-specific uncertainty. As shown in Fig. 11(b), the deterministic IDM achieves zero collisions in simulation, which is expected since the prediction model matches the simulator. However, this zero-margin approach offers no robustness against real-world behavioral deviations. The lane-agnostic model, which applies uniform uncertainty across lanes, introduces excessive conservatism that paradoxically increases collisions to 11.0%, as overly cautious planning causes hesitation and deadlock. Our lane-specific model strikes an effective balance, reducing collisions to 2.0% while providing structured robustness margins calibrated to lane-dependent behavioral variability, as outer-lane vehicles exhibit inherently higher trajectory uncertainty than inner-lane vehicles.

VI. CONCLUSION

This paper introduces a safety-critical decision-making framework for autonomous vehicles navigating unsignalized, dual-lane roundabouts by integrating Monte Carlo Tree Search (MCTS) with a hierarchical risk assessment strategy. The framework offers three major innovations: a multi-agent MCTS structure for scalable and efficient action space exploration, a hierarchical safety assessment mechanism for robust spatiotemporal risk evaluation, and an adaptive reward function that effectively balances safety and efficiency. Experimental results confirm the effectiveness of the proposed method under varying autonomous vehicle penetration rates. In fully autonomous settings (100% CAVs), the framework achieves reduced trajectory deviations and eliminates PET violations when compared to baseline approaches. In mixed traffic scenarios (50% CAVs + 50% HDVs), the framework delivers even greater improvements by reliably handling the uncertainty introduced by human drivers, while maintaining low deviation and high safety margins. Sensitivity and ablation analyses further demonstrate robustness to key parameter choices and confirm that each proposed component—Shapley value allocation and lane-specific uncertainty modeling—contributes meaningfully to overall performance. We note that the current framework assumes accurate state observations and plans independently per CAV without inter-vehicle communication. While these assumptions are idealized, the probabilistic C2H uncertainty model provides inherent robustness margins that can partially absorb sensing inaccuracies. The current results, validated in simulation, demonstrate the framework's promise as an effective decision-support tool for complex roundabout coordination. Future work will focus on improving computational scalability through parallelized search strategies, extending the framework to diverse roundabout geometries, and integrating data-driven behavior prediction with validation under realistic sensing noise and communication delays.

REFERENCES

[1] S. Alkheder, F. Al-Rukaibi, and A. Al-Faresi, "Driver behavior at kuwait roundabouts and its performance evaluation," *IATSS research*, vol. 44, no. 4, pp. 272–284, 2020.

[2] L. Zhang, Y. Dong, H. Farah, and B. van Arem, "Social-aware planning and control for automated vehicles based on driving risk field and model predictive contouring control: Driving through roundabouts as a case study," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, 2023, pp. 3297–3304.

[3] E. Galceran, A. G. Cunningham, R. M. Eustice, and E. Olson, "Multi-policy decision-making for autonomous driving via changepoint-based behavior prediction," in *Proc. Robotics: Science and Systems (RSS)*, 2015, p. 6.

[4] E. Polders, S. Daniels, W. Casters, and T. Brijs, "Identifying crash patterns on roundabouts," *Traffic Inj. Prev.*, vol. 16, no. 2, pp. 202–207, 2015.

[5] C. Badue, R. Guidolini, R. V. Carneiro, P. Azevedo, *et al.*, "Self-driving cars: A survey," *Expert Syst. Appl.*, vol. 165, no. 113816, p. 113816, 2021.

[6] J. Pérez, V. Milanés, T. De Pedro, and L. Vlacic, "Autonomous driving manoeuvres in urban road traffic environment: A study on roundabouts," in *Proc. IFAC World Congress*, vol. 44, 2011, pp. 13795–13800.

[7] J. Choi and D.-K. Kim, "Calibration and validation of the rule-based human driver model for car-following behaviors at roundabout using naturalistic driving data," *Asian Transp. Stud.*, vol. 10, p. 100129, 2024.

[8] Y. Shi, J. Wu, B. Zhu, J. Zhao, R. He, and Z. Chen, "Safety-enhanced behavioral decision strategy for intelligent vehicles under roundabout scenarios," *Inf. Sci.*, p. 122367, 2025.

[9] R. Azimi, G. Bhatia, R. Rajkumar, and P. Mudalige, "V2v-intersection management at roundabouts," *SAE Int. J. Passeng. Cars-Mech. Syst.*, vol. 6, no. 2013-01-0722, pp. 681–690, 2013.

[10] X. Gong, P. Lyu, and B. Wang, "Cooperative motion planning and decision making for cavs at roundabouts: A data-efficient learning-based iterative optimization method," *IEEE Internet Things J.*, vol. 11, no. 19, pp. 32 205–32 220, 2024.

[11] Z. Lin *et al.*, "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.

[12] S. Lu, H. Wang, L. Chen, and Y. Cai, "A multi-agent federated reinforcement learning-based cooperative vehicle-infrastructure control approach framework for roundabouts," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–16, 2025.

[13] B. Peng *et al.*, "Communication scheduling by deep reinforcement learning for remote traffic state estimation with bayesian inference," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4287–4300, 2022.

[14] Z. Lin, J. Lan, and X. Zhao, "Kan-1stm enhanced multi-agent advantage actor-critic reinforcement learning for autonomous ramp merging," *IEEE Trans. Veh. Technol.*, pp. 1–12, 2025.

[15] P. Cai *et al.*, "Dq-gat: Towards safe and efficient autonomous driving with deep q-learning and graph attention networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 21 102–21 112, 2022.

[16] 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.

[17] L. Ferrarotti, M. Luca, G. Santin, G. Prevati, G. Mastinu, M. Gobbi, E. Campi, L. Uccello, A. Albanese, P. Zalaya, A. Roccasalva, and B. Lepri, "Autonomous and human-driven vehicles interacting in a roundabout: A quantitative and qualitative evaluation," *IEEE Access*, vol. 12, pp. 32 693–32 705, 2024.

[18] Y. Zhang, B. Gao, *et al.*, "Adaptive decision-making for automated vehicles under roundabout scenarios using optimization embedded reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5526–5538, 2021.

[19] F. Konstantinidis, M. Sackmann, O. D. Candido, U. Hofmann, J. Thielecke, and W. Utschick, "Parameter sharing reinforcement learning for modeling multi-agent driving behavior in roundabout scenarios," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, 2021, pp. 1974–1981.

[20] Z. Tian, D. Zhao, Z. Lin, D. Flynn, W. Zhao, and D. Tian, "Balanced reward-inspired reinforcement learning for autonomous vehicle racing," in *Proc. LADC*, 2024, pp. 628–640.

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

[22] 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.

[23] 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.

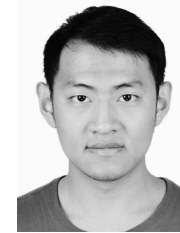
- [24] N. Ding, X. Meng, W. Xia, D. Wu, L. Xu, and B. Chen, "Multivehicle coordinated lane change strategy in the roundabout under internet of vehicles based on game theory and cognitive computing," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5435–5443, 2019.
- [25] M. Pourabdollah, E. Björkvik, F. Füller, B. Lindenberg, and K. Burgdorf, "Calibration and evaluation of car following models using real-world driving data," in *Proc. IEEE ITSC*, 2017, pp. 1–6.
- [26] Y. Bie, Y. Ji, and D. Ma, "Multi-agent deep reinforcement learning collaborative traffic signal control method considering intersection heterogeneity," *Transp. Res. Part C: Emerg. Technol.*, vol. 164, p. 104663, 2024.
- [27] P. Hang, C. Huang, Z. Hu, and C. Lv, "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.
- [28] 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. 12403–12413, 2022.
- [29] 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.
- [30] 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.
- [31] S. Bansal, M. Chen, S. Herbert, and C. J. Tomlin, "Hamilton-jacobi reachability: A brief overview and recent advances," in *Proc. IEEE Conf. Decis. Control (CDC)*, 2017, pp. 2242–2253.
- [32] 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 Trans. Comput. Intell. AI Games*, vol. 4, no. 1, pp. 1–43, 2012.
- [33] 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.
- [34] P. Zhou, X. Sun, and T. Chai, "Enhanced nmpe 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.
- [35] J. Wurts, J. L. Stein, and T. Ersal, "Design for real-time nonlinear model predictive control with application to collision imminent steering," *IEEE Trans. Control Syst. Technol.*, vol. 30, no. 6, pp. 2450–2465, 2022.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] M. Wang *et al.*, "Speed planning for autonomous driving in dynamic urban driving scenarios," in *Proc. ECCE*, 2020, pp. 1462–1468.
- [40] 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.
- [41] 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.
- [42] 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.
- [43] 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.
- [44] Y. Cui, S. Fang, Q. Chen, Y. Wang, P. Hang, and J. Sun, "A game-theoretic framework of interaction and cooperative driving for cavs at mixed unsignalized intersections," *IEEE Internet of Things Journal*, vol. 13, no. 1, pp. 1524–1538, 2026.
- [45] 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.



Zhihao Lin received the M.S. degree from the College of Electronic Science and Engineering, Jilin University, Changchun, China, in 2019. He is currently pursuing the Ph.D. degree with the James Watt School of Engineering, University of Glasgow, UK. His research interests focus on Structure-aware Reinforcement Learning and Multi-Agent Decision Making.



Jianglin Lan received the Ph.D. degree from the University of Hull in 2017. He has been a Leverhulme Early Career Fellow and Lecturer at the University of Glasgow since 2022. He was a Visiting Professor at the Robotics Institute, Carnegie Mellon University, in 2023. From 2017 to 2022, he held postdoc positions at Imperial College London, Loughborough University, and University of Sheffield. His research interests include AI, optimisation, control theory, and autonomy.



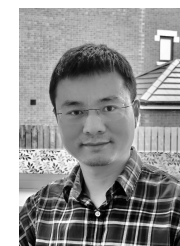
Shuo Liu (Student Member, IEEE) received his M.S. degree in Mechanical Engineering from Columbia University, New York, NY, USA, in 2020 and his B.Eng. degree in Mechanical Engineering from Chongqing University, Chongqing, China, in 2018. He is currently a Ph.D. candidate in Mechanical Engineering at Boston University, Boston, USA. His research interests include optimization, nonlinear control, deep learning, and robotics.



Zhen Tian received the B.Eng. degree in Electronic and Electrical Engineering from the University of Strathclyde, Glasgow, U.K., in 2020, and the Ph.D. degree from the College of Science and Engineering, University of Glasgow, Glasgow, U.K., in 2025. He is currently a postdoctoral researcher at the University of Glasgow. His research interests include interactive vehicle decision systems and autonomous racing strategies.



Dezhong Zhao received the B.Eng. and M.S. degrees from Shandong University, Jinan, China, in 2003 and 2006, respectively, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2010, all in Control Science and Engineering. He is a Reader in Autonomous Systems with the James Watt School of Engineering, University of Glasgow and a Turing Fellow with the Alan Turing Institute. He was awarded a Royal Society-Newton Advanced Fellow in 2020 and an EPSRC Innovation Fellow in 2018.



Chongfeng Wei received his Ph.D. degree in mechanical engineering from the University of Birmingham in 2015. He is now an Associate Professor (University Senior Lecturer) at University of Glasgow, UK. His current research interests include decision-making and control of intelligent vehicles, human-centric autonomous driving, cooperative automation, and dynamics and control of mechanical systems. He is also serving as an Associate Editor of IEEE TITS, IEEE TIV, IEEE TVT, and Frontier on Robotics and AI.