Mixed Platoon Control Strategy of Connected and Automated Vehicles based on Physics-informed Deep Reinforcement Learning

by

Haotian Shi

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Civil and Environmental Engineering)

at the

University OF WISCONSIN-MADISON

2023

Date of final defense: 04/24/2023

The dissertation is approved by the following members of the Final Oral Committee:

Prof. Bin Ran, Advisor, Civil and Environmental Engineering, UW-Madison
Prof. Soyoung (Sue) Ahn, Advisor, Civil and Environmental Engineering, UW-Madison
Prof. David A. Noyce, Civil and Environmental Engineering, UW-Madison
Prof. Xiaopeng Li, Civil & Environmental Engineering, UW-Madison
Prof. Xin Wang, Industrial and Systems Engineering, UW-Madison

ACKNOWLEDGEMENTS

Pursuing a Ph.D. degree has been a challenging yet unforgettable journey for me. I would like to express my deepest appreciation to the University of Wisconsin-Madison and the Department of Civil and Environmental Engineering for providing an exceptional environment for academic growth and research excellence. The guidance, encouragement, and resources provided by the esteemed faculty, staff, and colleagues have been instrumental in the successful completion of this dissertation thesis.

I would like to express my deepest gratitude to my advisor, Professor Bin Ran, for your unwavering guidance, support, and mentorship throughout the course of my Ph.D. journey at the Department of Civil and Environmental Engineering, University of Wisconsin-Madison. Your extensive knowledge, passion for research, and commitment to academic excellence have continually inspired and motivated me. You also provided support and guidance in improving my leadership, teamwork skills, and critical thinking. I deeply cherish your invaluable support to my Ph.D. journey.

I am grateful to Professor Yang Zhou. Your selfless and patient guidance on my research has significantly contributed to my academic and future career. It would have been impossible for me to accomplish this dissertation without you. As a bro, I have learned much from you and appreciate your support during my Ph.D. journey. I would also like to extend my gratitude to my other dissertation committee members, Professor David Noyce, Professor Xin Wang, Professor Sue Ahn, and Professor Xiaopeng Li, for their insightful advice that has promoted my growth. Additionally, I would like to thank Dr. Steven Parker and Glenn Vorhes for providing me with invaluable project experiences in the Tops Lab. I would also like to extend my appreciation to Dr. Yang Cheng for his support within Professor Ran's group.

I am also grateful to my lab mates, Keshu Wu and Sicheng Fu, who have been my closest friends and have supported me throughout my Ph.D. career. I appreciate the support from my lab friends, Dr. Shen Li, Dr. Zhen Zhang, Dr. Xiaotian Li, Dr. Tianyi Chen, Dr. Shuoxuan Dong, Dr. Kunsong Shi, Dr. Yangxin Lin, Dr. Yuan Zheng, Ran Yi, Yifan Yao, Shunyi Huang, Jingwen Zhu, Xiaodan Zhang, Han

Cao, Sixu Li, Kexin Tian, Junwei You, Rui Gan, Emily, and Rei Tamura. I also extend my gratitude to my friends in China and US. I truly appreciate the friendship and camaraderie.

I would like to thank my girlfriend, Yuhan Shao, for her unwavering support, patience, and understanding throughout this entire process. Your belief in me has often been the motivation I needed to persevere during the most challenging moments. You have helped me maintain balance and perspective during the highs and lows of this journey. Thank you so much.

Lastly, I would like to express my gratitude to my parents, my sister, my cat Sammy, and other family members. Thank you for your unwavering support and for believing in me even when I doubted myself. I am eternally grateful for the love and guidance you have provided me throughout my life.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This dissertation presents a distributed platoon control strategy of connected and automated vehicles (CAVs) based on physics-informed Deep Reinforcement Learning (DRL) for mixed traffic of CAVs and human-driven vehicles (HDVs). The dissertation will mainly consist of three parts: (i) generic DRL-based CAV control framework for the mixed traffic flow; (ii) DRL-based CAV distributed control under communication failure for the fully connected automated environment; (iii) distributed CAVs control for the mixed traffic flow, under real-time aggregated macroscopic car-following behavior estimation based on DRL.

For the first part, we first discussed the current challenges for CAV control in mixed traffic flow. For distributed CAV control, we categorize the local downstream environment into two broad traffic scenarios based on the composition of CAVs and HDVs to accommodate any possible CAV-HDV platoon configuration: (i) a fully connected automated environment, where all local downstream vehicles are CAVs, forming a CAV-CAVs topology; and (ii) a mixed local downstream environment, comprising the closest downstream CAV followed by one or more HDVs, creating a CAV-HDVs-CAV topology. This generic control framework effectively accommodates any CAV-HDV platoon topology that may emerge within the mixed traffic platoon. This part is discussed in Section 3.

For the second part, this study proposes a deep reinforcement learning (DRL) based distributed longitudinal control strategy for connected and automated vehicles (CAVs) under communication failure to stabilize traffic oscillations. Specifically, the Signal-Interference-plus-Noise Ratio (SINR) based vehicle-to-vehicle (V2V) communication is incorporated into the DRL training environment to reproduce the realistic communication and time-space varying information flow topologies (IFTs). A dynamic information fusion mechanism is designed to smooth the high-jerk control signal caused by the dynamic IFTs. Based on that, each CAV controlled by the DRL-based agent was developed to receive the real-time downstream CAVs' state information and take longitudinal actions to achieve the equilibrium consensus in the multi-agent system. Simulated experiments are conducted to tune the communication adjustment mechanism and further validate the control performance, oscillation dampening performance and generalization capability of our proposed algorithm. This part is discussed in Section 4.

The third part proposes an innovative distributed longitudinal control strategy for connected automated vehicles (CAVs) in the mixed traffic environment of CAV and human-driven vehicles (HDVs), incorporating high-dimensional platoon information. For mixed traffic, the traditional CAV control method focuses on microscopic trajectory information, which may not be efficient in handling the HDV stochasticity (e.g., long reaction time; various driving styles) and mixed traffic heterogeneities. Different from traditional methods, our method, for the first time, characterizes consecutive HDVs as a whole (i.e., AHDV) to reduce the HDV stochasticity and utilize its macroscopic features to control the following CAVs. The new control strategy takes advantage of platoon information to anticipate the disturbances and traffic features induced downstream under mixed traffic scenarios and greatly outperforms the traditional methods. In particular, the control algorithm is based on deep reinforcement learning (DRL) to fulfill car-following control efficiency and further address the stochasticity for the aggregated car following behavior by embedding it in the training environment. To better utilize the macroscopic traffic features, a general platoon of mixed traffic is categorized as a CAV-HDVs-CAV pattern and described by corresponding DRL states. The macroscopic traffic flow properties are built upon the Newell car-following model to capture the characteristics of aggregated HDVs' joint behaviors. Simulated experiments are conducted to validate our proposed strategy. The results demonstrate that the proposed control method has outstanding performances in terms of oscillation dampening, eco-driving, and generalization capability. This part is discussed in Section 5.

KEYWORDS

Mixed Traffic, Connected Automated Vehicles, Car Following, Deep Reinforcement Learning, Vehicle-to-Vehicle Communication.

1 INTRODUCTION

1.1 Background

Traffic oscillations, known as the stop-and-go phenomenon (Li et al., 2010), contribute to traffic flow instability, traffic unsafety, and energy inefficiency. With the fast development of vehicular automation and communication technology, connected and automated vehicles (CAV) gradually occupied some portion of the vehicle market. CAVs equipped with advanced communication and automation capability have great potential to alleviate traffic oscillations to enhance the traffic flow performance through adaptive cruise control (ACC) (Marsden et al., 2001) and cooperative adaptive cruise control (CACC) (Arem et al., 2006). It is envisioned that the CAVs and human-driven vehicles (HDVs) will co-exist in the near future (Zhou et al., 2019), which will change the pure traffic flow of HDVs to the mixed traffic flow of CAVs and HDVs (Lu & Liu, 2021; Zhang & Yang, 2021). Despite the changes in the traffic environment, traffic oscillations remain a demanding issue in the congested mixed traffic flow Therefore, efficiently controlling CAVs to drive safely and smartly in the mixed traffic environment is important both in academia and industry implementation.

The CAVs longitudinal control strategies have been comprehensively investigated in recent years, which can be divided into three categories: (i) analytical linear or non-linear CAVs control, (ii) model predictive control (MPC) based CAVs control, and (iii) deep reinforcement learning (DRL) based CAVs control. Each of these strategies has pros and cons. The analytical linear or non-linear CAVs controller (Morbidi et al., 2013; Stipanović et al., 2004; Treiber et al., 2000; Zhou & Ahn 2019; Guo et al., 2020; Zhu et al., 2020) is fast computing and easy to implement, and the corresponding closed-form control policy leads to the convenient stability analysis (Zheng et al., 2014, 2016; Li et al., 2019; Wang et al., 2020), where local stability and string stability can be mathematically guaranteed by properly choosing controller parameters. However, the analytical linear or non-linear controller is hard to explicitly incorporate multiple objectives and safety constraints with reasonable boundaries. As one of the most popular optimal control methods, MPC-based CAV controllers (Zhou et al., 2017; Gong et al., 2016; Wang et al., 2016; Zhou et al., 2019) can optimize multiple objectives in a flexible constrained framework. This method predicts the leading vehicle's future state (e.g., position, speed) and optimizes the driving behavior in a rolling/receding fashion. However, MPC usually requires the formulated problem to be convex, and it may need great demand for computation depending on the complexity of formulation, which makes it difficult for real-time implementation. The computation demand can be more intensive when stochasticity or uncertainties are explicitly considered (Chen et al., 2018).

Although there have been many studies about different approaches to optimizing car-following behavior and traffic flow, gaps in the studies remain in the following two aspects. Firstly, a fast computing multi-objective CAVs control strategy for the mixed platoon to improve mixed traffic string stability (e.g., Naus et al., 2010; Ge & Orosz, 2014), car following efficiency, and eco-driving performances, is still challenging. An exact optimization-based control such as MPC in a mixed connected automated traffic environment is hard to construct due to the uncertainty of HDVs (Zhou et al., 2019), and fast computation is required to satisfy the real-time implementation. Secondly, the unpredictable driving behavior of human-driven vehicles (HDVs) contributes to traffic disruptions and intensifies oscillation amplitude throughout the vehicular stream. This, in turn, negatively impacts the stability, efficiency, and energy consumption of mixed traffic flow (Zheng et al., 2020). Consequently, addressing the stochastic nature of HDV behavior is a crucial challenge in controlling connected and automated vehicles (CAVs) within mixed traffic environments. Thirdly, it is hard to optimize mixed traffic flow considering the different combinations of CAVs and HDVs due to the heterogeneity. Mixed platoons with different combinations have diverse characteristics, making it hard to develop a comprehensive model for optimal control, especially with multiple objectives.

In summary, despite extensive research on CAV control strategies, several research gaps persist:

Identified gaps

- Most studies assume flawless communication among CAVs, neglecting the potential for communication failures and dynamic information flow topology (IFT). This assumption leads to overly optimistic results.
- For deep reinforcement learning (DRL)-based CAV control, there is often a lack of equilibrium concepts derived from control theory, which could aid in analyzing string stability and local stability within a vehicular platoon.
- DRL-based CAV control typically lacks consensus concepts from multi-agent control theory, which can prevent disturbance accumulation and achieve system-level performance.
- In a mixed traffic environment consisting of CAVs and HDVs, addressing the adverse effects of HDV stochasticity presents a significant challenge.
- The heterogeneous nature of mixed traffic platoons complicates the development of a universal CAV control approach that ensures system-level control performance.
- A generic, distributed, and computationally efficient CAV control approach is needed, regardless of the CAV platoon size, dynamic IFT topology, and the CAV-HDV topology within mixed traffic environments.

1.2 Research Objectives and Scope of Work

Taking into account the advantages and disadvantages of various CAV longitudinal controllers and the identified gaps in existing research, this study aims to develop a physics-informed DRL-based CAV control strategy for mixed connected automated traffic environments by addressing the following aspects:

(i) Develop a generic and multi-objective CAV control strategy suitable for any CAV-HDV topology of a mixed vehicular platoon.

(ii) Create an integrated generic distributed control approach to control CAVs in a realistic communication environment, regardless of the CAV platoon size and dynamic IFT topology caused by the communication loss.

(iii) The generic CAV control strategy aims to efficiently stabilize traffic oscillations, enhance car-following efficiency, and improve eco-driving performances.

(iv) Integrate the equilibrium concept and the consensus concept into the DRL framework to achieve better control.

(v) Design a novel CAV control strategy that adeptly mitigates the adverse effects resulting from the stochastic nature of human-driven vehicles (HDVs).

1.3 Research Contributions

The research contributions of this dissertation work are summarized as follows.

(i) Propose an integrated generic distributed control approach to control CAVs in a realistic communication environment, regardless of the CAV platoon size and dynamic IFT topology caused by communication loss.

(ii) Propose a generic fast-computing CAV control strategy for the mixed vehicular platoon of any CAV-HDV topology.

(iii) Integrate the equilibrium concept and the consensus concept into the DRL framework to achieve better control by creatively fusing information of the downstream environment.

(iv) Embed the real ground-truth dataset and the dynamic communication mechanism in the distributed DRL training framework to better capture the stochasticity of driving behaviors and communication loss.

(v) Propose a novel and generic car-following structure, 'CAV-AHDV-CAV,' which can capture the aggregated HDVs' macroscopic traffic features (i.e., fundamental diagram) and embed them in the microscopic CAV longitudinal control, which efficiently alleviates the adverse impact brought by HDVs' stochasticity and optimizes the whole mixed traffic flow.

(vi) Better dampen traffic oscillations and meanwhile enhance car following efficiency, safety, and acceleration smoothness than other state-of-art CAV control methods.

(vii) Marry the merits of control theory, traffic flow theory, and Artificial Intelligence (AI), which better utilizes the information of sensed surrounding environment and leads to a promising control performance.

1.4 Dissertation Organization

This thesis is organized as follows. Chapter 2 introduces the relevant studies regarding CAV control strategies, DRL algorithms, V2V communications, and the equilibrium concept and consensus concept from control theories; Chapter 3 discusses the CAV control challenges in mixed traffic and introduces the proposed generic CAV control strategy for mixed traffic. Chapter 4 presents a DRL based distributed longitudinal control strategy for connected and automated vehicles (CAVs) under communication failure to stabilize traffic oscillations and enhances car-following control performances. Chapter 5 presents a generic CAV control strategy for any CAV-HDV topology of a mixed vehicular platoon, which integrates the macroscopic traffic features to handle HDVs' stochasticity. Conclusion and future studies are suggested in Chapter 6.

2 LITERATURE REVIEW

This chapter presents a literature review on research related to traditional CAV control strategy, deep reinforcement learning and DRL-based control, V2V communications, and equilibrium and consensus concepts from control theory.

2.1 Traditional Connected Automated Control Strategy

The study of CAV control strategies attracted many researchers, though the majority of them only focused on the pure connected automated environments (Gong & Du, 2018). The mainstream methods can be largely divided into three categories based on their modeling differences: (i) linear or non-linear CAV longitudinal controller, (ii) model predictive control (MPC) based CAV longitudinal controller with functions of objectives and constraints, (iii) deep reinforcement learning (DRL) based CAV longitudinal controller. First, linear (e.g., Stipanović et al., 2004; Naus et al., 2010; Morbidi et al., 2013) and non-linear (e.g., Bando et al., 1995; Treiber et al., 2000) CAV longitudinal controllers have closed-form formulations with parameters or gains. This type of model requires less calculation time due to its simplicity. In addition, its stability analysis is convenient due to the closed-form representation (e.g., Zheng et al., 2014; Shladover et al., 2015; Zheng et al., 2016; Petrillo et al., 2018; S. E. Li et al., 2019; Wang et al., 2020), and theoretically, the local stability and string stability of controllers can be guaranteed through appropriate parameter tuning. However, these linear and non-linear controllers have difficulties in designing an explicitly formulated framework to incorporate multiple control objectives (e.g., car following efficiency (local stability), string stability, energy efficiency) and collision-free constraints within reasonable vehicle acceleration/deceleration boundaries. Considering this limitation, MPC based CAV longitudinal controllers (e.g., Wang et al., 2016; Gong et al., 2016; Zhou et al., 2017; Gong & Du, 2018; Zhou et al., 2019;) have been favored in recent years. As an optimal control method, the MPC-based controller provides a flexible, constrained optimization framework that incorporates flexible optimizing objectives and constraints. Thus, CAV longitudinal control problem that considers string stability, efficiency, fuel consumption, and driving comfort can be solved at each timestep under safety constraints within a bounded acceleration range. Besides, this approach can provide driving decisions by predicting the future state of the leading vehicle trajectory, thus improving optimization and control performance in a rolling/receding horizon fashion. However, this approach is normally computational demanding and time consuming, which is not applicable for real-time implementation (Zhou et al., 2017). Since it solves a deterministic constrained optimization in a rolling horizon fashion, it usually requires the problem to be convex. Otherwise, it is generally forbidding to be solved efficiently. Furthermore, it is challenging to quantitatively guarantee its platoon string stability due to the formulation complexity (Zhou et al., 2019).

2.2 Deep Reinforcement Learning and DRL-based Control

Similar to the MPC controller, the recent breakthrough in the DRL community provides alternative algorithms to be utilized (Karnchanachari et al., 2020). The advantages of DRL based approaches are mainly reflected in the two aspects. First, DRL is a model-free and learning-based method that does not have any specific requirement for convexity of the problem and is suitable for capturing complex and stochastic system characteristics. Second, the computational burden of a DRL algorithm mainly lies in its offline training process, while the learned driving policy can be rapidly implemented in real-time (Görges, 2017). Specifically, the reinforcement learning basics are given as follows.

Modeling reinforcement learning is a Markov Decision Process (MDP) (Van Otterlo & Wiering, 2012), which contains a set of interactive objects, namely agent and environment. Except for two interactive objects, five model elements are included in MDP: state s_t, action a_t, policy $\pi(a|s)$, reward $r_t(s_t, a_t)$, and return R. RL aims to address problems with specific target through continuous exploration. At time step t, the input to the network is the observing state s_t of an agent, while an action a_t from action space is output according to the policy π. Reward r_t related to a_t and s_t is computed by the target-guided reward function and obtained by the agent. The agent then moves to the next states s_{t+1} due to a_t. Process terminates when the agent moves to the terminal state s_τ. At the same time, return R is obtained by summing cumulative rewards from timestep 0 to timestep τ. Thus, the whole training process from the first s_0 to the terminal s_τ is defined as an episode. Then, next episode restarts, back and forth. The target problem is solved through maximizing R.

Generally, model-free reinforcement learning can be divided into two categories: value-based reinforcement learning and policy-based reinforcement learning. Recently, deep reinforcement learning, combining reinforcement learning and deep learning, has made breakthrough that makes RL applied in different fields. A typical algorithm of value-based reinforcement learning is DQN (Deep Q Network), which was proposed by DeepMind, with its improved version published in (Mnih et al., 2015). Combined with deep learning, this algorithm uses a convolutional neural network to fit the value function, namely the Q function. Since DQN was proposed, it has shown a large number of improved algorithms which enhance the overall structure of the system and neuron networks. However, DQN is only applied to problems with discrete action space, which restricts the scope of application. To address more complex problems with continuous action space, Deep Deterministic Policy Gradient (DDPG), Asynchronous Advantage Actor-Critic (A3C), Trust Region Policy Optimization (TRPO), Proximal Policy Optimization (PPO) and other policy-based algorithms with actor-critic structure have been proposed and applied in diverse areas. DDPG is proposed by (Lillicrap et al., 2016), which has two separate and interdependent networks: actor network and critic network. DDPG combines DQN and deterministic policy gradient (DPG) by creating a memory buffer to separate training samples in a successive environment, which enhances stability of training process. Google DeepMind proposed A3C algorithm with multiple parallel agents (Mnih et al., 2016). These agents simultaneously update parameters in the primary structure on parallel environments, which reduces relevance and improves convergence of the algorithm. TRPO is proposed by (Schulman et al., 2015). The algorithm improves the convergence of policy updates by restricting KL divergence between the prediction distribution of the old strategy and new strategy on the same batch of data. PPO algorithm proposed by (Schulman et al., 2017) has become the default RL algorithm of Open AI. It is similar with TRPO but shows better sample efficiency due to multiple updates per batch sample. Combined with the advantage of A3C, Google DeepMind proposed a distributed PPO algorithm to update the policy of global agent in parallel through multiple working agents (Heess et al., 2017).

Recently, DRL algorithms have been gradually applied to design CAV controllers (Chong et al., 2013; Li et al., 2020; Zhou et al., 2020). Guan et al. (2019) applied a DRL algorithm to cooperatively control CAVs at intersections and address the computational burden by training offline. Duan et al. (2020) comprehensively consider both high-level and low-level motion control for CAVs based on DRL, which achieves a smooth and safe decision-making process. Wang et al. (2019) developed a Q-learning based bird-view approach for CAV control, which shows great control performance under complicated traffic environments. However, these works have not considered the mixed platoon's string stability and mainly focus on distance tracking (e.g., maintaining a car following headway with reasonable acceleration control). In addition, the equilibrium state concept for DRL-based CAV control is usually missing, which renders DRL a large search space for the optimal policy. As far as the authors know, only Qu et al. (2020) proposed a control strategy based on the DDPG algorithm to dampen traffic oscillations and improve energy efficiency. However, this study only considers non-cooperative vehicle control without exploiting information sharing. Besides, this study adopts a model-free gap policy, which cannot directly guarantee a stable traffic flow. In general, DRL algorithm applications are still rare from the perspective of stability analysis in a mixed traffic environment.

2.3 V2V Communications

Regarding the communication of CAVs, vehicle-to-vehicle (V2V) communication enables traffic information exchange through all surrounding CAVs, which improves CAVs’ situational awareness and performance in safety, mobility, and sustainability (Wang et al., 2019). However, the distance between sender and receiver vehicles and the mutual communication interference from equipped vehicles significantly affects the V2V communication connection, even causing a communication failure (Kim et al., 2017). The fading effect of the signal amplitudes over distance greatly influences the success rate of the V2V communication, especially when multiple CAVs frequently exchange and disseminate data simultaneously. Once communication failure occurs, the CAV loses the ability to receive the proceeding vehicles’ information, and the information flow topology (IFT) can change dynamically, which impairs CAV's performance in mobility, stability, and even safety. With these concerns, recent studies (Wang et al., 2020; Zhou et al., 2020b) proposed two-step algorithms based on a linear controller, whose first step is to optimize IFT for a time period that maximizes the expected string stability, and the second step is to find parameters for linear controllers to guarantee string stability under dynamic communication environment. However, the two-step algorithm is relatively separate and cannot update the IFT dynamically as an integrated system together with the control design.

2.4 Equilibrium and Consensus Concept

On the aspect of the equilibrium concept and consensus concept, they are critical for the car-following control approach to achieve stability-wise and system-level control performance. Without the equilibrium state concept, the stability, including string stability (Ploeg et al., 2014) and local stability (Willems et al., 2014) of DRL-based controller, is forbidden to be analyzed in the DRL-based control studies. Considering this issue, Shi et al., (2020) proposed a DRL-based centralized control strategy for connected and automated vehicles (CAVs). However, the approach divides mixed vehicular platoon into multiple sub-platoons, each of which is controlled by a centralized controller. Considering the potential communication loss, the sub-platoon size can also vary, which hinders the applications of sub-platoon centralized control. Further, with the increased size of the centralized DRL-based controller, the larger action and state dimensions make the training hard to converge. Hence, rather than developing a centralized controller, a distributed controller may be better fitted for the case with communication loss. To ensure the distributed controller can still achieve a great system-level control performance, a consensus concept from the multi-agent control theory (L. Zhang & Orosz, 2017), requiring all agents to maintain the desired relative states with respect to their neighbors, needs to be considered. The consensus property hinders the accumulation of disturbances and achieves system-level stability. Therefore, incorporating the consensus in the DRL-based CAV control framework can be expected to further stabilize traffic oscillations.

3 GENERIC CAV CONTROL STRATEGY FOR MIXED TRAFFIC FLOW

This chapter presents the CAV control challenges in mixed traffic flow and introduces the proposed generic CAV control strategy for mixed traffic flow.

3.1 CAV Control Challenges in Mixed Traffic Flow

Although the CAV longitudinal controllers described in Section 2 can provide strong tools to control CAVs efficiently, how to handle mixed traffic still remains a problem due to the heterogeneity and HDVs' stochastic and uncertain movements (Gong & Du, 2018). The stochastic HDV driving behavior triggers traffic disturbances and amplifies the oscillation amplitude through the vehicular stream, which impairs the mixed traffic flow stability, travel efficiency, and energy (Zheng et al., 2020). Moreover, the heterogeneous driving behaviors in mixed traffic may create voids and further reduce traffic throughput (D. Chen et al., 2020). To handle the HDV uncertainty and mixed traffic heterogeneities, approaches of recent studies can be largely divided into two categories: 1). predict the proceeding HDV's driving behaviors (Gong & Du, 2018; Bang & Ahn, 2019; Wang et al., 2020; Zhu et al., 2018) and incorporate the prediction in the control strategy (e.g., by MPC); 2). divide mixed traffic into sub-platoons (Shi et al., 2021; Wang, 2018b) for more efficient control and apply a cooperative control strategy for CAVs assuming that the disturbances triggered by HDVs are completely random. For the first type of method, Gong & Du (2018) utilized an online curve matching algorithm to predict the HDV trajectory and developed a cooperative platoon control for the mixed traffic environment. However, this study only predicts the last HDV of consecutive HDVs in the mixed traffic, which does not fully use the downstream traffic information that can be potentially conveyed by proceeding CAVs. The downstream traffic information can be very helpful in predicting the wave propagation and oncoming traffic scenario. On the other hand, Shi et al. (2021) proposed a DRL-based cooperative CAV longitudinal control strategy for the mixed traffic environment, which divides the mixed platoon into multiple subsystems for centralized cooperative control. However, the method cannot utilize the downstream information of each sub-platoon and model it as random noise in the environment.

Moreover, this approach lacks sufficient generalization with the increased centralized CAV sizes. With this concern, a distributed manner may be better for CAV control in mixed traffic. Besides the two primary approaches mentioned above, the connected cruise control (CCC) strategy was developed to handle the various connectivity structures in heterogeneous platoons by assuming an oversimplified nonlinear car following laws for HDVs. Based on that, they designed a control law for CAVs to improve the car following performance and traffic efficiency (e.g., Ge & Orosz, 2014; Orosz, 2016; Zhang & Orosz, 2016). However, these studies did not focus on addressing HDVs' inherent stochasticity and stabilizing the mixed platoon.

In general, although these approaches (e.g., HDV behavior prediction; sub-platoon) could improve CAV control performances in the mixed traffic environment, the limits still remain as follows. Firstly, it is challenging to effectively incorporate HDVs' behavior for control due to its inherent stochastic and personalized characteristics, especially for the aggregated (i.e., multiple consecutive) HDVs in the mixed traffic. The joint behaviors of multiple HDVs are hard to model. In the case of platooning, while only the individual behaviors are modeled, the predicting error will be accumulated and propagated over time and space (Lin et al., 2020). Even the microscopic behavior is stochastic and difficult to capture, the aggregated HDV driving behaviors exhibit macroscopic traffic flow properties (e.g., kinematic wave propagating time, density) with typical traffic phenomena (e.g., shock wave propagation), and they can be modeled by for example the fundamental diagram (Meng et al., 2021; Tian et al., 2021). Comparing to the microscopic behaviors, the aggregated driving behaviors show less stochasticity, as indicated by the central limit theorem (Kwak & Kim, 2017). Therefore, this study aims to attenuate the aggregated HDVs' stochasticity by incorporating their macroscopic traffic properties into the control framework. Secondly, the mixed traffic environment has various vehicular compositions due to the different combinations of HDVs and CAVs, as presented in Fig. 3-1. Approaches such as sub-platoon or centralized cooperative control (Du et al., 2020; Shi et al., 2021; Wang, 2018; Zheng et al., 2020) lack adequate flexibility and may suffer from computation burden. Such heterogeneity makes it challenging to develop a generic CAV control approach with a system-level control performance.

With this concern, the study aims to build a generic CAV control framework to generalize the varied CAV-HDV topologies and incorporate the macroscopic features.

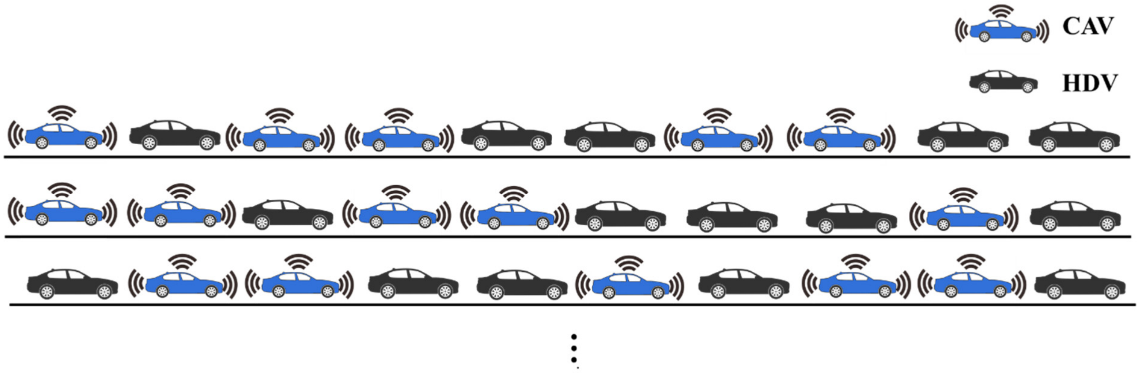


Fig. 3-1. Different compositions of mixed connected automated traffic environment

3.2 Generic DRL-based Control Framework

In a mixed traffic platoon, various combinations of connected and automated vehicles (CAVs) and human-driven vehicles (HDVs) can be observed. The controlling CAV typically processes information from the downstream environment to make control decisions. To accommodate any possible CAV-HDV platoon configuration, we categorize the local downstream environment into two broad traffic scenarios based on the composition of CAVs and HDVs, as illustrated in Fig. 3-2: (i) a fully connected automated environment, where all local downstream vehicles are CAVs, forming a CAV-CAVs topology; and (ii) a mixed local downstream environment, comprising the closest downstream CAV followed by one or more HDVs, creating a CAV-HDVs-CAV topology.

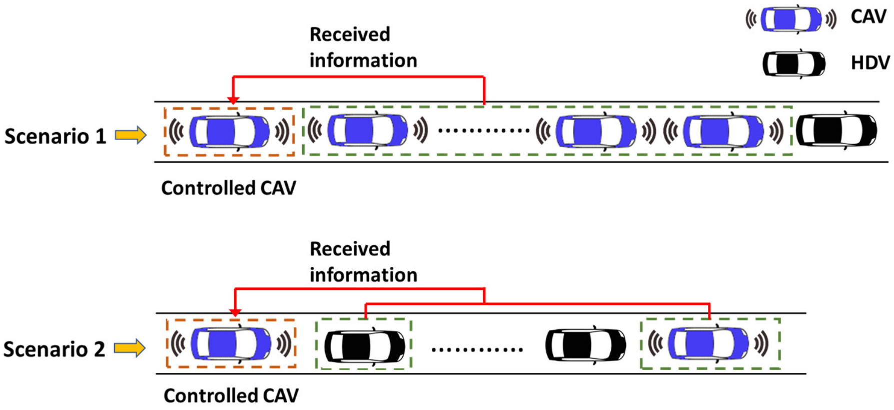


Fig. 3-2. The two scenarios of the local environment

In the fully connected automated environment, the controlling CAV gathers information from multiple downstream CAVs to make decisions. In the mixed local downstream environment, the controlling CAV receives information from the immediate preceding HDV as well as the nearest downstream CAV. This generic control framework effectively accommodates any CAV-HDV platoon topology that may emerge within the mixed traffic platoon.

3.2.1 Fully Connected Automated Environment

For the fully connected automated environment, we propose a generic DRL-based distributed framework for CAV control. The detailed methodology and results are discussed in Section 4 of the thesis.

Within the framework, we aim to develop an integrated generic distributed control approach with a dynamic IFT mechanism to control CAVs in a realistic communication environment. Specifically, our DRL framework is designed with the following novelties. First, we embed the real ground-truth dataset (i.e., Next Generation Simulation (NGSIM) datasets) and the dynamic communication mechanism (i.e., Signal-Interference-plus-Noise Ratio (SINR) (Du & Dao, 2015)) in the distributed training framework since DRL can etter capture stochastic behaviors of proceeding vehicles and stochastic communication loss. Second, we develop a generic DRL-based control framework, regardless of the CAV platoon size and dynamic information topology caused by communication loss. Specifically, the DRL state and reward function were specially designed to integrate the equilibrium concept and the consensus concept into the DRL framework by appropriately fusing multiple downstream CAVs' information in a weighted sum manner. Considering the potential fluctuations of the weighted sum of the downstream CAV's information and the resulting unstable control caused by communication loss, we developed a dynamic information fusion mechanism to smooth the high-jerk control signal and ensure the desired control performance. By this design, the fused state and predefined equilibrium state regulate CAVs to keep close to a predefined equilibrium point regardless of traffic scenario, which increases the generalizability and robustness of the control method. Moreover, the equilibrium state gives DRL an exploration direction in the training process to improve the convergence and ability to dampen traffic oscillations, compared with the decentralized control, which merely uses the very nearest proceeding vehicle information.

3.2.2 Mixed Local Traffic Environment

For the mixed traffic local downstream environment, we propose a novel vehicle following structure, "CAV-AHDV-CAV," as a generic unit for mixed traffic of any vehicle ordering and simultaneously embedded platoon-level features in the distributed CAV control framework. The detailed methodology and results are discussed in Section 5 of the thesis.

The defined 'AHDV' component in the "CAV-AHDV-CAC" structure means the aggregated HDVs between the two CAVs in the structure. This novel structure characterizes the aggregated consecutive HDVs in the mixed traffic as a whole, denoted as the 'AHDV,' whose aggregated HDV car-following behaviors and stochasticity can be further captured by the macroscopic traffic features. Specifically, we propose an estimated time-varying Newell car-following method (D. Chen et al., 2012), which links the fundamental diagram to the microscopic driving behavior parameters.

Furthermore, DRL is suitable for capturing stochastic characteristics and embedding them in the environment with great generalization capability. Thus, this structure is incorporated into the DRL framework to fulfill car-following control efficiency and further reduce stochasticity based on the following two aspects. First, the ground-truth HDV trajectory data are embedded into the DRL training process, by which we incorporate real HDV stochastic characteristics implicitly. Second, the macroscopic features captured by the 'CAV-AHDV-CAV' structure are weighted and fused into the DRL state and reward function based on the equilibrium concept. In this way, the HDVs' stochasticity is alleviated by regulating CAVs close to the pre-defined equilibrium state. With the proposed 'CAV-AHDV-CAV' structure and the designed DRL framework, the HDV stochasticity is efficiently alleviated for CAV control.

To summarize, the method utilizes the DRL framework to develop a generic distributed CAV longitudinal control approach for a mixed traffic environment. The contribution can be summarized in terms of methodology and application. From the methodology-wise perspective, the aggregated HDVs' macroscopic traffic flow features are real-time estimated based on the generic 'CAV-AHDV-CAV' structure. The structure is embedded into the DRL control framework by a specially designed DRL state and reward function, which efficiently alleviates the adverse impact of HDVs' stochasticity and optimizes the whole mixed traffic flow. From the application-wise side, a generic strategy for any CAV-HDV topology of a mixed vehicular platoon is developed to stabilize the traffic oscillations efficiently. Specifically, each controlled CAV receives the information from the local downstream vehicles for real-time control. The received information is fused as the DRL state based on the philosophy of equilibrium concept and the consensus concept, which helps develop a robust control policy and gives the base for analyzing car-following control efficiency and vehicular string stability. The DRL reward function is then designed based on the fused DRL state in a quadratic form to efficiently fulfill the car-following control efficiency and improve driving comfort performance.

4 DEEP REINFORCEMENT LEARNING BASED DISTRIBUTED CONNECTED AUTOMATED VEHICLE CONTROL UNDER COMMUNICATION FAILURE

This chapter proposes a deep reinforcement learning (DRL) based distributed longitudinal control strategy for connected and automated vehicles (CAVs) under communication failure to stabilize traffic oscillations. The control strategy is designed for the fully connected automated environment discussed in Section 3 of the thesis. Specifically, the Signal-Interference-plus-Noise Ratio (SINR) based vehicle-to-vehicle (V2V) communication is incorporated into the DRL training environment to reproduce the realistic communication and time-space varying information flow topologies (IFTs). A dynamic information fusion mechanism is designed to smooth the high-jerk control signal caused by the dynamic IFTs. Based on that, each CAV controlled by the DRL-based agent was developed to receive the real-time downstream CAVs' state information and take longitudinal actions to achieve the equilibrium consensus in the multi-agent system. Simulated experiments are conducted to tune the communication adjustment mechanism and further validate the control performance, oscillation dampening performance and generalization capability of our proposed algorithm.

The chapter is organized as follows. Section 4.1 presents the environment settings, including basic assumptions and the adopted V2V communication model. CAV longitudinal control framework and the proposed dynamic information fusion mechanism are described in Section 4.2. Section 4.3 proposed the details of DRL model development and training procedure. The proposed CAV longitudinal control strategy is validated by numerical experiments presented in Section 4.4. Section 4.5 gives the conclusion of this work.

4.1 Environment Setting

This research considers the car-following process without lateral movement in a straight highway segment with infinite length. The communication between CAVs applies a dedicated short-range communication (DSRC) radio with a 5.9-GHz frequency, which is adopted by the Federal Communications Commission for transportation safety and mobility (L. Du & Dao, 2015). The basic assumptions for the simulation environment are given as follows: (i) The CAV's state information (e.g., spacing, speed) can be broadcasted to the local upstream CAVs through V2V communication in real-time. (ii) SINR dynamically determines the successful transmission between two CAVs. (iii) The communication time is not considered in this study as it can be negligible for measuring the delay on a road segment (L. Du & Dao, 2015). (iv) The CAV can receive its immediate predecessor's state information through onboard sensors and its own state information through GPS.

For a given CAV platoon in a realistic environment, the V2V communications can be unreliable and constantly changing over time due to failures caused by communication interference or information congestion (Wang et al., 2019). The uncertain communication environment will impair CAV's driving behavior and thus the entire traffic flow. To optimize the CAV's driving behavior and stabilize traffic oscillations under the realistic environment, this study provides a control framework incorporating a distributed DRL-based CAV control strategy and a dynamic adjusted V2V IFT. The V2V communication topology with broadcast mechanism, widely utilized for the CACC framework (Noor-A-Rahim et al., 2019; Wang et al., 2018, 2020), is adopted in this study. For this communication topology, each CAV broadcasts its information to multiple upstream CAVs and simultaneously receives its downstream CAVs' information for real-time control.

Specifically, the IFT, which demonstrates the information links of all vehicles in the platoon, changes dynamically based on the SINR condition. To describe the IFT from the receiver side, we introduced a vector $\boldsymbol{\xi}_{\mathbf{i}}^{\mathbf{t}} = [\eta_{i,i-1}^{t}, \eta_{i,i-2}^{t}, \dots, \eta_{i,i-N}^{t}]$, whose each $\eta_{i,i-m}^{t} \in \{0, 1\}$ indicates the information transmission status between the receiver $CAV\ i$ and the transmitter $CAV\ i - m$: $\eta_{i,i-m}^{t}$= 1 denotes a successful transmission; otherwise, $\eta_{i,i-m}^{t}$= 0. Notably, we assume that $\eta_{i,i-1}^{t} \equiv 1$ due to the robust onboard sensors, representing that the CAV can always receive its immediate predecessor's state information. For instance, $CAV\ i$ with dynamic IFTs receives real-time information of the three downstream CAVs for control, as presented in Fig. 4-1. The three possible real-time IFTs in Fig. 4-1(a), Fig. 4-1(b), Fig. 4-1(c) have $\boldsymbol{\xi}_{\mathbf{i}}^{\mathbf{t}} = [1, 1, 1], [1, 0, 1], [1, 0, 0]$, respectively.

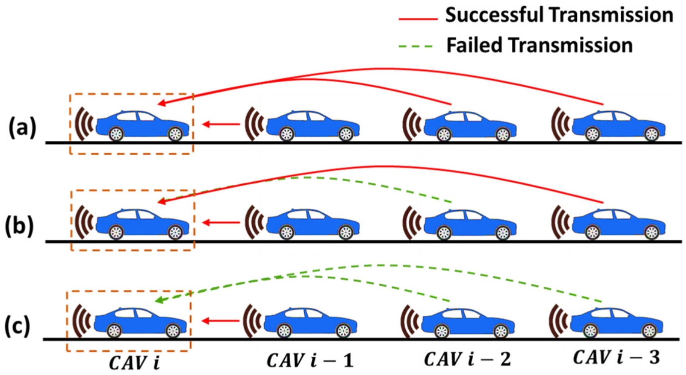


Fig. 4-1. A schematic diagram showing different IFTs of $\boldsymbol{CAV\ i}$ in a four-vehicle platoon.

To reproduce the realistic communication environment, this study, inspired by (L. Du & Dao, 2015), uses the SINR communication model to determine the successful wireless communication condition. The SINR, presented in Equation (4-1), is a commonly used standard that considers multiple realistic factors to measure the wireless connection quality. The SINR quality $y_{j,i}^t$ between the transmitter CAV i and the receiver CAV j at timestep t is specified as:

$$y_{j,i}^t = \frac{P_i\left(x_{ij}^t\right)^{-\alpha}}{\sum_{k=1,k\neq i}^{n} e_k P_k (X_{kj}^t)^{-\alpha} + O} \tag{4-1}$$

where P_i represents the transmission power of CAV i; α is the signal power decay; X_{ij}^t denotes the distance between the two vehicles. $\sum_{k=1,k\neq i}^{n}(e_k P_k X_{kj}^t)^{-\alpha}$ is the sum of the interference signal power from all proceeding vehicles in the communication range, where e_k is a Boolean parameter that determines whether $CAV\ k$ can share its information. In this study, all CAVs are allowed to share information (i.e., $e_k \equiv 1$). O denotes the noise term. A normal distribution ($O \sim N(\mu, \sigma^2)$) is adopted to illustrate the noise effect. Based on the SINR quality $y_{j,i}^t$, the information transmission status $\eta_{i,i-m}^t$ is defined as:

$$\eta_{i,i-m}^t = \begin{cases} 1, \ if\ y_{i,i-m}^t > \beta \\ 0, \ if\ y_{i,i-m}^t \leq \beta \end{cases}, \tag{4-2}$$

where β is a threshold value determined by the communication modulation and code rate.

Considering multiple V2V environmental factors, the proposed SINR model captures critical communication features in the actual condition, which reliably determines a successful connection between CAVs. The default parameter settings are shown in **Table 4-1**.

4.2 Distributed Control Scheme

Based on the environment setting in Section 4.1, this section describes the control scheme of the proposed strategy, including a distributed control framework for regulating CAVs' longitudinal movements (in Section 4.2.1) and a dynamic information fusion mechanism for reducing control signal's high jerks caused by the time-varying IFT (in Section 4.2.2). The related notations are defined in **Table 4-1.**

Table 4-1. Notations of the control scheme

Symbol	Definition
$y_{j,i}^t$	The SINR quality between the transmitter CAV i and the receiver CAV j
β	The SINR threshold value for a successful transmission
$\eta_{i,i-m}^t$	The information transmission status between the receiver $CAV\ i$ and the transmitter $CAV\ i-m$ at timestep t
Δt	The timestep size (update interval)
$\mathbf{s_i^t}$	The fused DRL state for $CAV\ i$ at timestep t
u_i^t	The desired acceleration signal of $CAV\ i$ at timestep t
a_i^t	The realized acceleration of $CAV\ i$ at timestep t
v_i^t	The velocity of $CAV\ i$ at timestep t
$\Delta v_{i,i-m}^t$	The relative speed between $CAV\ i$ and $CAV\ i-m$

$d_{i,i-m}^{t}$	The spacing between $CAV\ i$ and $CAV\ i-m$
$d_{i,i-m}^{*t}$	The equilibrium spacing between $CAV\ i$ and $CAV\ i-m$
$\varDelta d_{i,i-m}^{t}$	The deviation from the equilibrium spacing between $CAV\ i$ and $CAV\ i-m$
w_{i-m}	The state coefficient of the transmitter $CAV\ i-m$
$\Delta\tilde{d}_{i}^{t}$	The weighted deviation of spacing for $CAV\ i$ at timestep t after IFT adjustment
$\Delta\tilde{v}_{i}^{t}$	The weighted deviation of speed for $CAV\ i$ at timestep t after IFT adjustment
$\mathrm{c}_{i,i-m}^{t}$	The permission parameter for determining whether to fuse the information of $CAV\ i-m$
$\tilde{\eta}_{i,i-\mathrm{m}}^{t}$	The adjusted information transmission status between $CAV\ i$ and $CAV\ i-m$ at timestep t
e_{i}^{t}	The first-order difference of $\Delta\tilde{d}_{i}^{t}$; $e_{i}^{t} = \lvert\Delta\tilde{d}_{i}^{t} - \varDelta\tilde{d}_{i}^{t-1}\rvert$

4.2.1 Distributed Control Scheme

In this section, we proposed a distributed CAV longitudinal control strategy under the unreliable communication environment, whose framework is as presented in Fig. 4-2. In this general scenario, the controlled $CAV\ i$ communicates with local downstream CAVs and receives their fused state information for control at each timestep. The fused state information is generated based on $CAV\ i$'s adjusted IFT within a certain communication range (i.e., k local downstream vehicles).

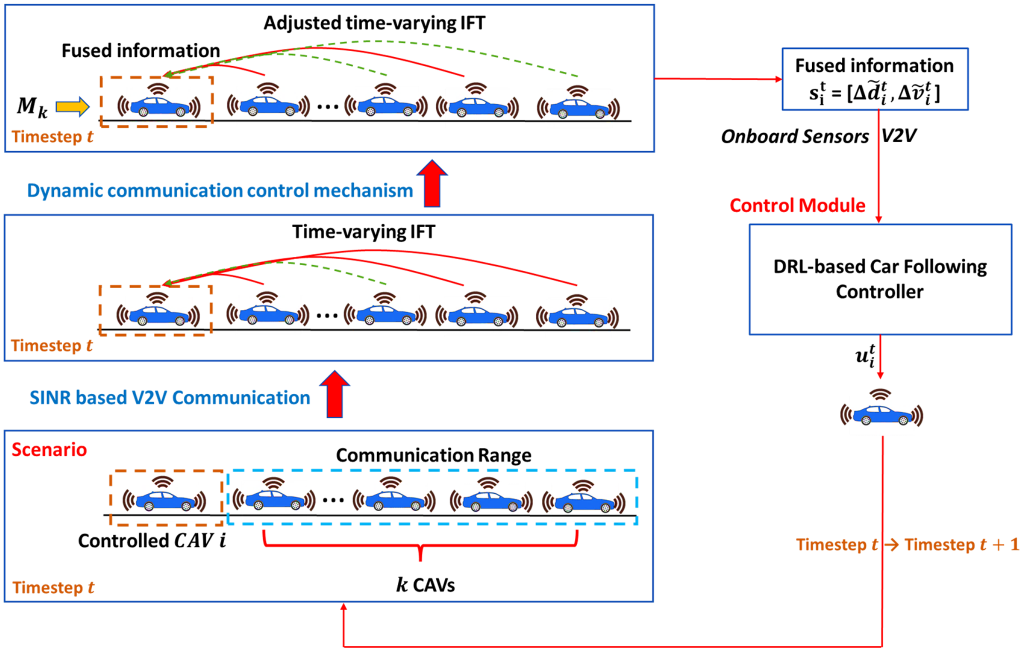

Adjusted time-varying IFT
Fused information
M_k
Timestep t
Dynamic communication control mechanism
Time-varying IFT
Timestep t
SINR based V2V Communication
Scenario
Communication Range
Controlled $CAV\ i$
Timestep t
k CAVs
Fused information
$\mathbf{s}_\mathbf{i}^\mathbf{t} = [\Delta\tilde{d}_i^t, \Delta\tilde{v}_i^t]$
Onboard Sensors
V2V
Control Module
DRL-based Car Following Controller
u_i^t
Timestep t → Timestep $t+1$

Fig. 4-2. A distributed control framework for $\boldsymbol{CAV\ i}$ in the multi-agent vehicle platoon.

Specifically, the $CAV\ i$'s time-varying IFT, determined by the SINR-based communication model, is adjusted by our proposed dynamic information fusion mechanism, which will be explained in Section 4.2.2. Then, the fused state information $\mathbf{s}_\mathbf{i}^\mathbf{t}$ is generated based on the adjusted IFT and sent to the DRL-based controller. $\mathbf{s}_\mathbf{i}^\mathbf{t}$ is calculated from the local downstream vehicles' information (i.e., speed difference, gap, position) to the required DRL state $\mathbf{s}_\mathbf{i}^\mathbf{t} = [\Delta\tilde{d}_i^t\ ,\ \Delta\tilde{v}_i^t]$, representing the weighted deviations from the target equilibrium, which will be explained with details later. Based on the fused state information, the DRL-based controller (denoted as M_k if controlled CAV receives k downstream CAV's fused information) outputs the desired acceleration signal u_i^t to regulate the CAV's longitudinal movement at each timestep. Given the above framework, the detailed design is given as below.

In our longitudinal control, we consider a vehicle's linearized dynamics which captures the air drag force, gear position and road gradient. It is modeled with the first-order approximation using the generalized vehicle dynamics (GLVD) equation (Li et al., 2011; Wang, 2018):

$$\dot{a}_i^t = -\frac{1}{T_{i,L}}a_i^t + \frac{K_{i,L}}{T_{i,L}}u_i^t, \tag{4-3a}$$

$$a_i^{t+1} = e^{-\frac{\Delta t}{T_{i,L}}} \times a_i^t + \left(1 - e^{-\frac{\Delta t}{T_{i,L}}}\right) \times K_{i,L} u_i^t \tag{4-3b}$$

where $T_{i,L}$ is the actuation time lag and $K_{i,L}$ is the ratio of the demanded acceleration that can be realized for vehicle i. Readers can find more details for the value of the above two parameters under different conditions in (Wang, 2018). $a_i^t$is the realized acceleration, usually within a boundary $[a_{min}, a_{max}]$; $\dot{a}_i^t$ is the jerk. Based on the realized acceleration a_i^t, the vehicle state is updated using the kinematic point-mass model (Zhu et al., 2018):

$$v_i^{t+1} = v_i^t + a_i^t \Delta t, \tag{4-4a}$$

$$\Delta v_{i,i-1}^{t+1} = v_{i-1}^{t+1} - v_i^{t+1}, \tag{4-4b}$$

$$d_{i,i-1}^{t+1} = d_{i,i-1}^t + \frac{\Delta v_{i,i-1}^t + \Delta v_{i,i-1}^{t+1}}{2} \times \Delta t, \tag{4-4c}$$

where Δt is the control interval; v_i^t denotes $CAV\ i$'s velocity at timestep t; v_{i-1}^{t+1} indicates the velocity of $CAV\ i$'s preceding vehicle; $d_{i,i-1}^t$ is the vehicle spacing, representing the distance between front bumpers of $CAV\ i$ and $CAV\ i-1$.

With above vehicle longitudinal dynamics, the fused DRL state $\mathbf{s_i^t}$ is specially designed in our DRL-based controller. The control design follows the concept of a distributed control framework for cooperation in the multi-agent vehicular platoon (L. Zhang & Orosz, 2017), aiming to achieve a consensus of CAV platoon and meanwhile regulate CAVs to keep close to a predefined equilibrium point for each car following pair. The merit of utilizing the equilibrium concept is to avoid the arbitrary change of the car following spacing that may render unstable traffic flow. Specifically, the equilibrium concept is based on the constant time gap (CTG) policy from the Society of Automotive Engineer Standard. It regulates each car following pair formed by CAV i and $i-1$ to reach the same speed and maintain the preset equilibrium spacing as below:

$$d_{i,i-1}^{*t} = v_i^t \tau_i^* + l_i, \tag{4-5a}$$

$$v_i^{*t} = v_{i-1}^t, \tag{4-5b}$$

where v_i^t is the speed of $CAV\ i$; l_i is the standstill spacing; τ_i^* is the constant time gap.

Considering the downstream vehicles' impact and platoon-level consensus, we can further expand the local equilibrium for each car following pair to a distributed system-level equilibrium $d_{i,i-m}^{*t}$ between CAV i and $i - m$, $\forall m = [1,2, \dots k]$, whose equilibrium spacing follows:

$$d_{i,i-m}^{*t} = m\left(v_i^t \tau_i^* + l_i\right), \tag{4-6a}$$

$$v_i^{*t} = v_{i-m}^t, \tag{4-6b}$$

Based on that, the deviation from the equilibrium spacing $\Delta d_{i,i-m}^t$ and the relative speed $\Delta v_{i,i-m}^t$ are defined as:

$$\Delta d_{i,i-m}^t = d_{i,i-m}^t - d_{i,i-m}^{*t}, \tag{4-7}$$

$$\Delta v_{i,i-m}^t = v_{i-m}^t - v_i^t. \tag{4-8}$$

To reduce the dimension of state in DRL for better convergence, and meanwhile better achieve the multi-vehicle consensus by utilizing downstream traffic information, similar to CAV multi-agent linear control (Bian et al., 2019; Chen et al., 2021), the equilibrium deviations between $CAV\ i$ and its k downstream vehicles ($\Delta d_{i,i-m}^t$, $\Delta v_{i,i-m}^t$, $1 \leq m \leq k$) are weighted averaged to the fused state information $\boldsymbol{s}_{\mathbf{i}}^{\mathbf{t}}$ = $[\Delta \tilde{d}_i^t\, , \Delta \tilde{v}_i^t]$ for $CAV\ i$. The weighted deviations of spacing $\Delta \tilde{d}_i^t$ and speed $\Delta \tilde{v}_i^t$ are given as Equation (4-9) and Equation (4-10), respectively:

$$\Delta \tilde{d}_i^t = \frac{w_{i-1}\tilde{\eta}_{i,i-1}^t \Delta d_{i,i-1}^t + w_{i-2}\tilde{\eta}_{i,i-2}^t \Delta d_{i,i-2}^t + \cdots + w_{i-k}\tilde{\eta}_{i,i-k}^t \Delta d_{i,i-k}^t}{w_{i-1}\tilde{\eta}_{i,i-1}^t + w_{i-2}\eta_{i,i-2}^t + \cdots w_{i-k}\tilde{\eta}_{i,i-k}^t}, \tag{4-9}$$

$$\Delta \tilde{v}_i^t = \frac{w_{i-1}\tilde{\eta}_{i,i-1}^t \Delta v_{i,i-1}^t + w_{i-2}\tilde{\eta}_{i,i-2}^t \Delta v_{i,i-2}^t + \cdots + w_{i-k}\tilde{\eta}_{i,i-k}^t \Delta v_{i,i-k}^t}{w_{i-1}\tilde{\eta}_{i,i-1}^t + w_{i-2}\tilde{\eta}_{i,i-2}^t + \cdots w_{i-k}\tilde{\eta}_{i,i-k}^t}, \tag{4-10}$$

where the coefficient w_{i-m} is defined in Equation (4-11), reflecting that the closer downstream CAVs are paid more attention.

$$w_{i-m}=\begin{cases}\frac{1}{2^m}, & 1\le m\le k-1\\ \frac{1}{2^{m-1}}, & m=k\end{cases}. \tag{4-11}$$

$\tilde{\eta}_{i,i-m}^t$ denotes the adjusted transmission status between $CAV\ i$ and $CAV\ i-m$, which will be given by an information fusion mechanism in Section 4.2.2.

4.2.2 Dynamic Information Fusion Mechanism

The continuity and smoothness of the DRL state ($\Delta\tilde{d}_i^t$ and $\Delta\tilde{v}_i^t$) are essential for DRL-based control methods since the learned policy directly maps the DRL state to the control action (i.e., $u_j^i=\pi_\theta(\mathbf{s_j^i})$). However, if the IFT of $CAV\ i$ is not adjusted (i.e., $\tilde{\eta}_{i,i-m}^t$= $\eta_{i,i-m}^t$), the communication loss will cause the transmission status $\tilde{\eta}_{i,i-m}^t$ to switch frequently and correspondingly make the DRL state $\mathbf{s_i^t}$ fluctuate, which can lead to undesirable high-jerk accelerations, as presented in Fig. 4-3.

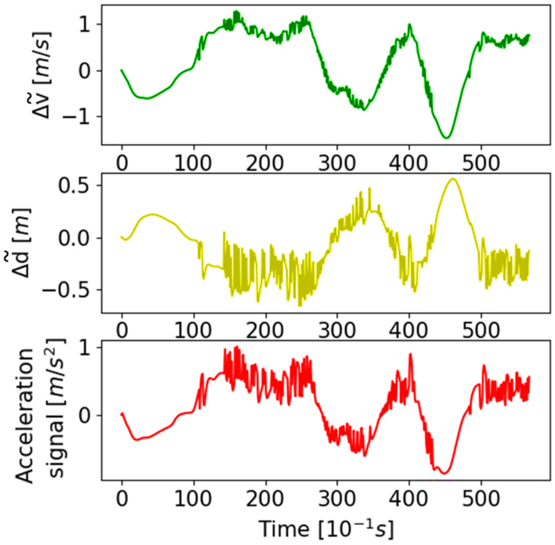


Fig. 4-3. The high jerk phenomenon due to the unstable transmission.

Rather than directly fusing the information received by letting$\tilde{\eta}_{i,i-m}^t$= $\eta_{i,i-m}^t$, we add one more mechanism, ‘dynamic information fusion mechanism’, by determining whether CAV i is allowed to fuse the information of CAV $i-m$ at each timestep. The proceeding CAVs’ information will be fused if and only if the controlled CAV is allowed to fuse the information and meanwhile the information is received. Mathematically, our dynamic control mechanism introduces a ‘fusion permission’ parameter $\mathrm{c}_{i,i-m}^{t} \in \{0, 1\}$ for $CAV\ i$ to determine whether it is allowed to fuse the information from $CAV\ m$ at each timestep t, where 1 represents permission and vice versa. Based on this mechanism, the adjusted IFT of $CAV\ i\ \tilde{\xi}_i^t$ is defined as $\tilde{\boldsymbol{\xi}}_{\mathbf{i}}^{\mathbf{t}} = [\tilde{\eta}_{i,i-1}^{t},\ \tilde{\eta}_{i,i-2}^{t}, \dots, \tilde{\eta}_{i,i-k}^{t}]$, where $\tilde{\eta}_{i,i-k}^{t}$ represents whether CAV $i-k$’s information will be utilized for the state fusion by CAV i. The detailed definition $\tilde{\eta}_{i,i-k}^{t}$ is given as:

$$\tilde{\eta}_{i,i-k}^{t} = \begin{cases} 1, if\ \eta_{i,i-m}^{t} = 1 \text{ and } \mathrm{c}_{i,i-m}^{t} = 1 \\ 0,\ otherwise \end{cases} . \tag{4-12}$$

Equation (4-12) represents that the fusion only happens when fusion permission and information receipt both hold. It should be noted that $\mathrm{c}_{i,i-1}^{t} \equiv 1$ since the information of $CAV\ i-1$ is necessary due to safety concerns and $\eta_{i,i-1}^{t} \equiv 1$ since information can be directly measured by vehicle on-board sensors.

As presented in Fig. 4-4, a rule-based method for determining the permission parameter $\mathrm{c}_{i,i-m}^{t}$ is designed. Specifically, we firstly set the default value of the permission parameter $\mathrm{c}_{i,i-m}^{t}$ =1, aiming to utilize the information of proceeding vehicles as much as possible. $\mathrm{c}_{i,i-m}^{t} = 0$ only happens when the following two conditions hold simultaneously: (i) the transmission status between $CAV\ i$ and $CAV\ i-m$ changes from “fail” at timestep $t-1$ to “success” at timestep t (i.e., $\tilde{\eta}_{i,i-m}^{t-1} = 0$ and $\tilde{\eta}_{i,i-m}^{t} = 1$); (ii) the first-order difference of the weighted deviation e_i^t (i.e., $e_i^t = |\Delta\tilde{d}_i^t - \Delta\tilde{d}_i^{t-1}|$), triggered by condition (i), is larger than a threshold q (i.e., $e_i^t > q$). The above rule helps to reduce the sudden state change caused by communication status change, which may result in control non-smoothness. The threshold q determines and adjusts the smoothness of the control signal, which is flexible to meet the requirements of varied control actuators. The sensitivity analysis of the threshold q regarding signal smoothness and control performances is conducted in Section 4.4.1 to seek its optimal range.

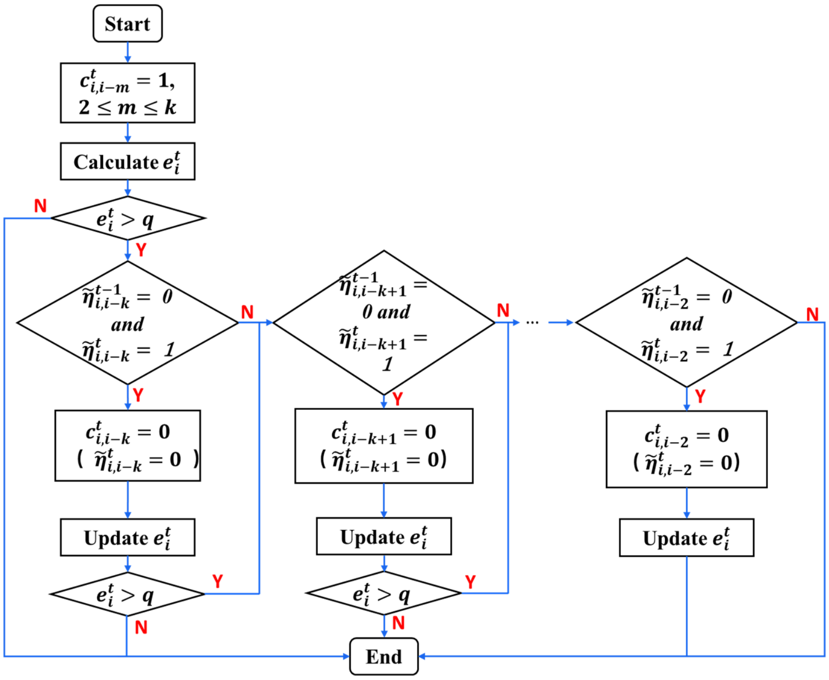

Start
$c^t_{i,i-m} = 1,$
$2 \le m \le k$
Calculate e^t_i
N
$e^t_i > q$
Y
$\tilde{\eta}^{t-1}_{i,i-k} = 0$ and $\tilde{\eta}^{t}_{i,i-k} = 1$
N
$\tilde{\eta}^{t-1}_{i,i-k+1} =$ 0 and $\tilde{\eta}^{t}_{i,i-k+1} =$ 1
N
...
$\tilde{\eta}^{t-1}_{i,i-2} = 0$ and $\tilde{\eta}^{t}_{i,i-2} = 1$
N
Y
$c^t_{i,i-k} = 0$
$(\ \tilde{\eta}^t_{i,i-k} = 0\ )$
Y
$c^t_{i,i-k+1} = 0$
$(\ \tilde{\eta}^t_{i,i-k+1} = 0)$
Y
$c^t_{i,i-2} = 0$
$(\ \tilde{\eta}^t_{i,i-2} = 0)$
Update e^t_i
Update e^t_i
Update e^t_i
Y
$e^t_i > q$
Y
$e^t_i > q$
N
N
End

Fig. 4-4. The flow chart of the dynamic communication control mechanism.

4.3 DRL MODEL DEVELOPMENT

Based on the above control scheme, this section develops the DRL-based models. We first describe the DRL framework design (given in Section 4.3.1), including the representations of the four basic DRL elements (state, action, policy, and reward). Then, the DRL algorithm (DPPO) details for policy updating are given in Section 4.3.2. The training procedure is described in Section 4.3.3.

4.3.1 DRL Design

DRL can be modeled as a Markov decision process, consisting of two interactive objects: DRL agent (CAV control algorithm) and environment (given in Sections 4.1 and 4.2). The DRL framework has four basic elements: state, action, policy, and reward: state, action, policy, and reward (**s,** A, π, r).

The state information **s** contains two components: the weighted deviations of spacing $\Delta\tilde{d}^t_i$ and speed $\Delta\tilde{v}^t_i$, as discussed in the last section. When the DRL agent receives the state information $\mathbf{s^t_i}$, it outputs the action A, namely u^t_i, to control $CAV\ i$ according to a control policy π. The policy π is an implicit function updated through the training process to achieve optimal performance described by the reward r.

The reward r determines control targets. In our design, cooperative control efficiency and driving comfort are considered to achieve the consensus equilibrium and simultaneously smooth driving behavior. The cooperative control efficiency f_i^tmeasures the deviation from the consensus equilibrium in a quadratic form:

$$f_i^t = (\mathbf{s_i^t})^T \mathbf{Q_i} \mathbf{s_i^t}, \tag{4-13}$$

where Q_i is a positive definite diagonal coefficient matrix with tuning weights $\alpha_{1,i}$, $\alpha_{2,i}$, defined as:

$$\mathbf{Q_i} = \begin{bmatrix} \alpha_{1,i} & \\ & \alpha_{2,i} \end{bmatrix}, \alpha_{1,i}, \alpha_{2,i} > 0. \tag{4-14}$$

Especially, the cooperative car following control efficiency f_i^t regulates the equilibrium spacing deviation $\Delta d_{i,i-1}^t \rightarrow 0$ and relative speed $\Delta v_{i,i-1}^t \rightarrow 0$, which greatly reduces the driving risks as manifested by the safety surrogate measure such as time-to-collision (TTC), where $TTC_i^t = \begin{cases} \frac{d_{i,i-1}^t - l_v}{\Delta v_{i,i-1}^t}, \text{if } v_i^t > v_{i,i-1}^t \\ \infty, if\ v_i^t \leq v_{i,i-1}^t \end{cases}$ (Jiménez et al., 2013). When $\Delta d_{i,i-1}^t \rightarrow 0$ and $\Delta v_{i,i-1}^t \rightarrow 0$, TTC→ ∞. During the non-steady state, small $\Delta v_{i,i-1}^t$ also suggests large TTC. Moreover, it is worth noting that our newly designed approach is different from (Zhou, et al., 2019; Shi, et al., 2020), whose objective only aims to minimize the quadratic term of local control efficiency cost $\hat{f}_i^t = (\mathbf{\hat{s}_i^t})^T \mathbf{Q_i} \mathbf{\hat{s}_i^t}$, where $\mathbf{\hat{s}_i^t}$ = [$\Delta d_{i,i-1}^t$, $\Delta v_{i,i-1}^t$]. $\hat{f}_i^t$ merely measures the local stability of the CAV longitudinal control, indicating a vehicle's capability to remain in a car following pair of the local equilibrium state. However, this term does not incorporate the consensus of the whole CAV platoon, which makes the CAV react very myopically and may lead to large $\Delta v_{i,i-m}^t$ and $\Delta d_{i,i-m}^t$.

Further, a trade-off cost term g_i^t=$Z_i(a_i^t)^2$ that evaluates driving comfort is incorporated in the car-following control, alleviating the control force to increase the driving comfort and string stability (i.e., acceleration magnitude decreases through vehicular string). Z_i is the acceleration weighting coefficient. Thus, the running cost l_i^t is formulated below:

$$l_i^t = f_i^t + g_i^t. \tag{4-15}$$

Since Equation (4-15) is quadratic, which makes the DRL property similar to constrained quadratic control, we adopted the settings of coefficient Z_i and matrix $\mathbf{Q_i}$ (Zhou et al., 2019) to enhance the empirical string stability.

The cost function (15) is converted to the immediate reward r_i^t for CAV i at timestep t using the exponential function, specified as:

$$r_i^t = \exp\left(-l_i^t\right). \tag{4-16}$$

Therefore, we formulate an infinite-horizon optimal control problem with the DRL policy π* to maximize the discounted cumulative rewards:

$$\pi^* = \arg\max_{\pi} \sum_{m=0}^{\infty} \Upsilon^m r_i^{t+m}\left(\mathbf{s_i^{t+m}}, a_i^{t+m}\right), \tag{4-17}$$

where $r\left(s_i^t, a_i^t\right)$ represents the reward function (16).

4.3.2 Distributed Proximal Policy Optimization (DPPO) Algorithm

The DRL solves the optimal control problem in Equation (4-17) by updating policy π in the training procedure. We used the DPPO algorithm (Heess et al., 2017) that supports continuous action space to update policy due to its great performance in sampling efficiency and convergence.

The DPPO algorithm consists of an actor network and a critic network, whose parameters need to be updated by training to find the optimal π^*. Specifically, the actor network's parameter θ, which directly determines the latest policy π, is updated by maximizing the objective function $L^{CLIP}(\theta)$:

$$L^{CLIP}(\theta) \; = \; \hat{E}_t[\min\left(p_t(\theta)\hat{A}_t, clip(p_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t\right], \tag{4-18}$$

where $p_t(\theta)$ represents the probability ratio of the new policy $\pi_\theta$and the old policy π_{old}, denoted as $p_t(\theta) = \frac{\pi_\theta(a_t|\mathbf{s_t})}{\pi_{\theta_{old}}(a_t|\mathbf{s_t})}$. The function $clip(p_t(\theta), 1-\varepsilon, 1+\varepsilon)$ limits $p_t(\theta)$ between $1-\varepsilon$ and $1+\varepsilon$, which prevents a large difference between the updated new policy and the old policy, thus improving the converging performance. ε is a parameter of the clipping function. $\hat{A}_t$ is the estimated advantage at timestep t:

$$\hat{A}_t \; = \; R_t - V_\Phi(\mathbf{s_i^t}), \tag{4-19}$$

where R_t represents the T-step discounted sum of rewards:

$$R_t = \sum_{m=0}^{T-1} \gamma^m r_i^{t+m} + \gamma^T V_\Phi(\mathbf{s_i^{t+T}}), \tag{4-20}$$

where r_i^{t+m} denotes the reward value given in Equation (4-16); γ is a discount factor.

On the other hand, the critic network evaluates the action u_i^t output by the actor network. A critic loss function $L_c(\Phi)$ is defined to be minimized to update the critic network:

$$L_c(\Phi) = \hat{E}_t\left(R_t - V_\Phi(\mathbf{s_i^t})\right)^2. \tag{4-21}$$

The detailed hyperparameter setting of the DPPO algorithm is given in **Table 4-2.**

Table 4-2. DPPO algorithm's hyperparameters.

Hyperparameters	Value
Clipping value ε	0.2
Discount factor Υ	0.99
Minibatch T	256
Actor learning rate	0.00001
Critic learning rate	0.00001
Number of the parallel agents	4

4.3.3 Training Procedure

This section provides the detailed training procedure, which aims to update the policy of the DRL agent. DPPO algorithm with the actor-critic network structure is adopted for updating policy π. The DPPO algorithm includes one global agent for updating the actor-critic networks' parameters and multiple parallel agents collecting data to improve the sampling efficiency. To be noted that, if there are fewer than k downstream vehicles, the controlled CAV integrates all the downstream information and takes proper action from DRL-based models (M_1~M_{k-1}), which are developed based on the same framework with M_k but with a smaller number of proceeding vehicles (1~$k-1$ downstream CAVs).

The training process of M_5 is demonstrated in Fig. 4-5. We set the number of communicated downstream CAVs $k=5$ in the control framework since the impact of far downstream CAV can be neglected ($w_{i-k} < \frac{1}{16}$ when $k > 5$). The numerical environment is built via Python. Specifically, the trajectory of the leading CAV ($CAV\ i-5$) is ground-truth data from the NGSIM datasets. The other downstream CAVs are controlled by the corresponding DRL-based models (M_1~M_4). Without losing generality, all following CAVs start with the equilibrium states defined in section 4.2. During the training process, each parallel agent receives state $\mathbf{s_i^t}$ at time step t and outputs u_i^t to control $CAV\ i$ based on the latest policy π. Simultaneously, the reward r_i^t is computed by the reward function (16) and stored in the memory buffer with state $\mathbf{s_i^t}$ and action u_i^t. After a specific batch of data is collected, the global agent will update policy π by optimizing the actor network and critic network parameters.

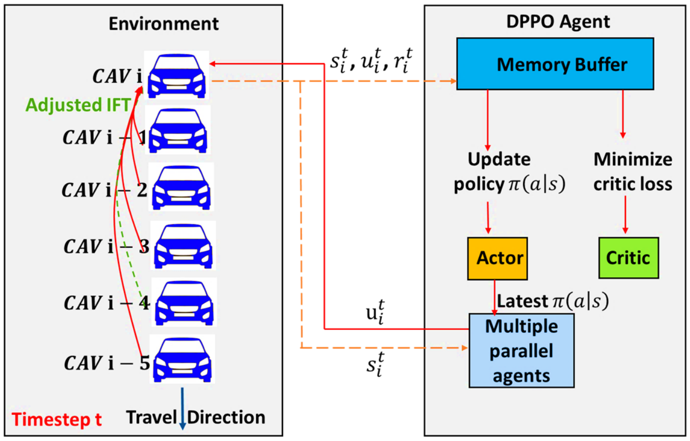


Fig. 4-5. The schematic diagram of training procedure.

The distributed controller is trained on 200 episodes, consisting of 218 timesteps with a 0.1s time interval. The moving reward trajectories (Qu et al., 2020) of developed models (M_1~M_5), presented in Fig. 4-6, show an almost monotonous increase with episodes until stable convergence, suggesting the good converging performance of our designed DRL. The main reason is that the predefined equilibrium state regulates CAVs to keep close to equilibrium, which gives DRL an exploration direction to improve the convergence and reduce the computation burden.

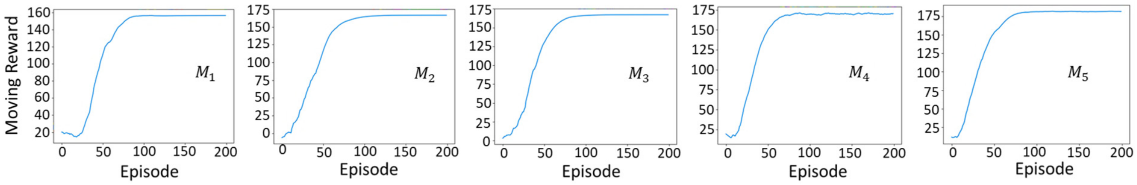


Fig. 4-6. Reward trajectories of the proposed models.

4.4 Numerical Experiments

Several experiments embedded with NGSIM datasets are conducted to evaluate the DRL-based distributed control strategy in this section. The raw field data was processed using a low-pass filter to efficiently clean noises (Montanino & Punzo, 2015). The trajectories of vehicles in Lane 2 of I-80 from 4:00 pm to 4:15 pm are adopted for experiments due to the frequent traffic oscillations period. Without losing generality, for each experiment, the followers start with initial equilibrium states. The default experimental settings are shown in Table 4-3.

Table 4-3. Experiment parameter settings.

Parameters	Value
Number of downstream CAVs k	5
Update interval Δt	0.1 s
Vehicle length l_v	4.5 m
Standstill spacing l_i	6.4 m
Constant time gap τ_i^*	1 s
Acceleration weighting coefficient Z_i	0.5
Coefficient matrix $\boldsymbol{Q_i}$	$\begin{bmatrix} 1 & \\ & 0.5 \end{bmatrix}$
SINR threshold value β	0.055
Ratio of the demanded acceleration $K_{i,L}$	1
Actuation time lag $T_{i,L}$	0.1
Threshold for the dynamic information infusion mechanism q	0.1
SINR noise parameter μ	0
SINR noise parameter σ^2	0.1
Acceleration boundary $[a_{min}, a_{max}]$	[-4 m/s^2, 4 m/s^2]

The experiments consist of the following three aspects: (i) parameter tuning for the dynamic information fusion mechanism (in Section 4.4.1); (ii) control performance evaluation and comparison with the decentralized strategy (in Section 4.4.2); (iii) generalization capability analysis (in Section 4.4.3).

Regarding part (i), the sensitivity analysis was conducted to seek the optimal range of the information fusion mechanism threshold q, aiming to achieve a great balance between the acceleration jerk and communication utilization rate. We average the absolute value of acceleration signal jerk j_i^t for all time steps to measure the control signal smoothness, specified as:

$$\tilde{j}_i = \frac{\sum_{t=0}^{N} j_i^t}{N}, \tag{4-22}$$

where N is the number of timesteps; $j_i^t = |u_i^t - u_i^{t-1}|/\Delta t$. The communication utilization rate is to indicate the percentage of fused information for a communication link, given by:

$$\gamma_{i,i-m} = \frac{\sum_{t=0}^{N} \tilde{\eta}_{i,i-m}^t}{\sum_{t=0}^{N} \eta_{i,i-m}^t}. \tag{4-23}$$

Regarding part (ii), the distributed control performance is analyzed and compared with a decentralized control strategy and a linear-based CACC strategy. The CAV controlled by the decentralized strategy can only receive the preceding vehicle's information through onboard sensors, which means the CAV is downgraded to the automated vehicle (AV). M_1 is applied for decentralized control. The compared CACC algorithm (Zhou et al., 2020c) is a linear CAV longitudinal controller also based on CTG policy, which greatly dampens traffic oscillations with guaranteed string stability performance. To quantitatively evaluate the performance of the control strategy, four performance indexes: driving comfort cost g_i^t, cooperative control efficiency cost f_i^t, local control efficiency cost $\hat{f}_i^t$, and the cumulative dampening ratio $d_{p,i}$, are incorporated in the analysis. The cumulative dampening ratio $d_{p,i}$ is defined to evaluate the string stability that measures the performance in dampening traffic oscillations through a platoon (Zhou et al., 2019), defined as:

$$d_{p,i} = \frac{\|a_i^t\|_2}{\|a_0^t\|_2} = \frac{\left(\sum_{t=0}^{N} |a_i^t - a_{i,mean}|^2\right)^{\frac{1}{2}}}{\left(\sum_{t=0}^{N} |a_0^t - a_{0,mean}|^2\right)^{\frac{1}{2}}}, \tag{4-24}$$

where i is the vehicle index; index 0 represents the leader of a platoon; $a_{i,mean}$ is the average acceleration of $CAV\ i$ over all timesteps. The smaller dampening ratio indicates the more string stable driving behavior. Particularly, the platoon is strict string stable when all vehicles satisfy $d_{p,i} \leq d_{p,i-1}$.

After performance evaluation, the control strategy's generalization ability is validated in part (iii), using multiple ground-truth datasets. Finally, the proposed strategy is implemented in different traffic conditions to demonstrate the oscillation-dampening performance compared with the (intelligent driver model) IDM-based HDV platoon (Treiber et al., 2000).

4.4.1 Sensitivity Analysis of the Communication Control Mechanism

The experiments in this section aim to optimize the dynamic information fusion mechanism through tuning the threshold q to achieve smooth control subject to sufficient communication utilization and control performances.

As an example, we analyzed an unstable communication link between the controlled $CAV\ i$ and the transmitter $CAV\ i-4$. Fig. 4-7 presents the acceleration jerk trajectories under different threshold q. The “Information fully adopted” case indicates the communication utilization rate $\gamma_{i,i-m} = 1$, which means the CAV utilizes (fuses) all the received information from $CAV\ i-4$ and thus can better achieve the downstream consensus with richer information. However, the acceleration signal jerk ranges between 1.1 m/s^3 to 5.5 m/s^3 in this case when the communication status is unstable. The “fully dropped” case represents $\gamma_{i,i-m} = 0$, in which the jerk trajectory is merely below 0.5 m/s^3. However, the CAV ignores all information from $CAV\ i-4$ in this case, which is not desired for achieving the equilibrium consensus. The two extreme cases demonstrate the trade-off relationship between the smoothness of the acceleration signal and the communication utilization, which needs to be balanced by finding an ‘optimal’ threshold q^*.

Thus, the experiment with threshold q (meters) ranging from 0.01 to 0.25 was enumerated, presented in Fig. 4-7. Compared with the case that fully utilizes the communication information, cases with smaller q help alleviate the jerk, with only a few time points with high jerk values. To quantify the results, we use 30 datasets with each over 50 seconds to average the acceleration jerk $\tilde{j}_i$ and communication utilization rate $\gamma_{i,i-4}$, presented in Fig. 4-8 (a). The quantified results further illustrate the trade-off relevance. Particularly, the average jerk $\tilde{j}_i$ increases monotonically from 0.173 m/s^3 to 0.309 m/s^3, with the communication utilization rate $\gamma_{i,i-m}$ growing from 28.4% to 87.8%. When $q \geq 0.03$, $\gamma_{i,i-m}$ is less sensitive while the trend of $\tilde{j}_i$ is more sensitive due to the nearly exponential growth.

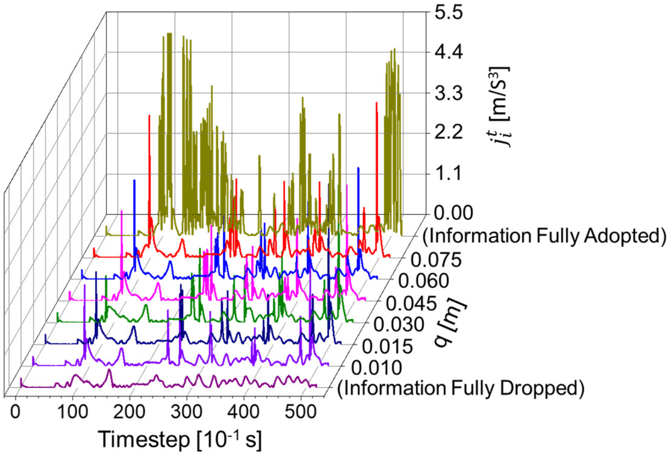
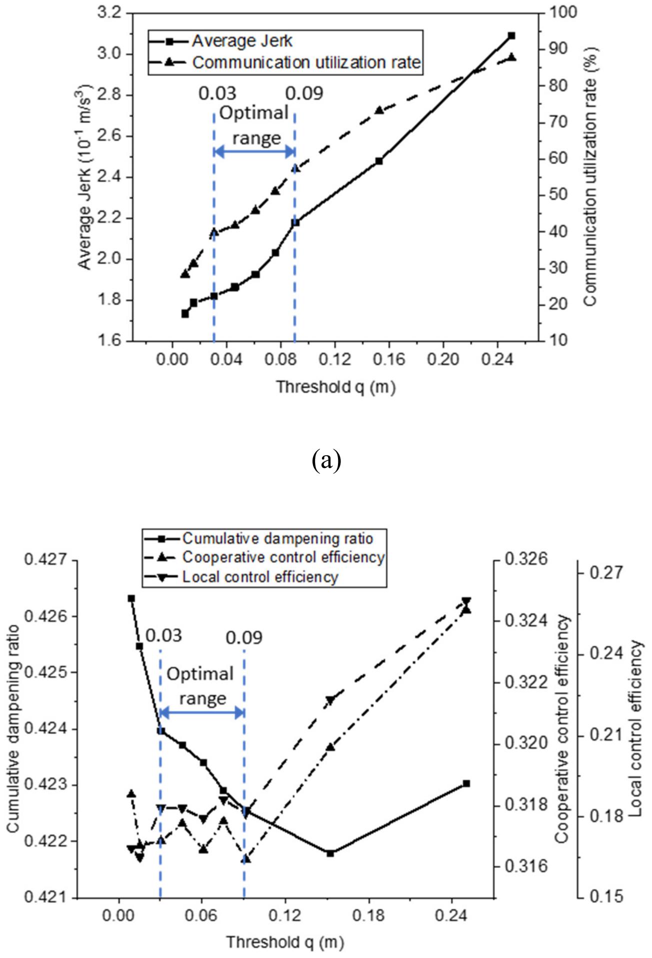


Fig. 4-7. The acceleration jerk trajectory under different thresholds.

Furthermore, the sensitivity analysis of threshold q was conducted regarding its impact on control performances (i.e., dampening ratio $d_{p,i}$, cooperative control efficiency f_i^t, and local control efficiency $\hat{f}_i^t$), as presented in Fig. 4-8 (b). As threshold q increases, the dampening ratio gradually decreases before since more information is utilized to facilitate the string stability performance. Nevertheless, the dampening ratio re-rises after q reaches some point (0.15) due to the high-jerk acceleration. On the other hand, the two types of control efficiency costs remain low when q is within a certain range (0.03 ~ 0.09) and then rise monotonically as q gradually increases, indicating that the high-jerk equilibrium deviations (i.e., $\Delta\tilde{d}_i^t, \Delta\tilde{v}_i^t$) negatively affect the control efficiency.



(a)

(b)

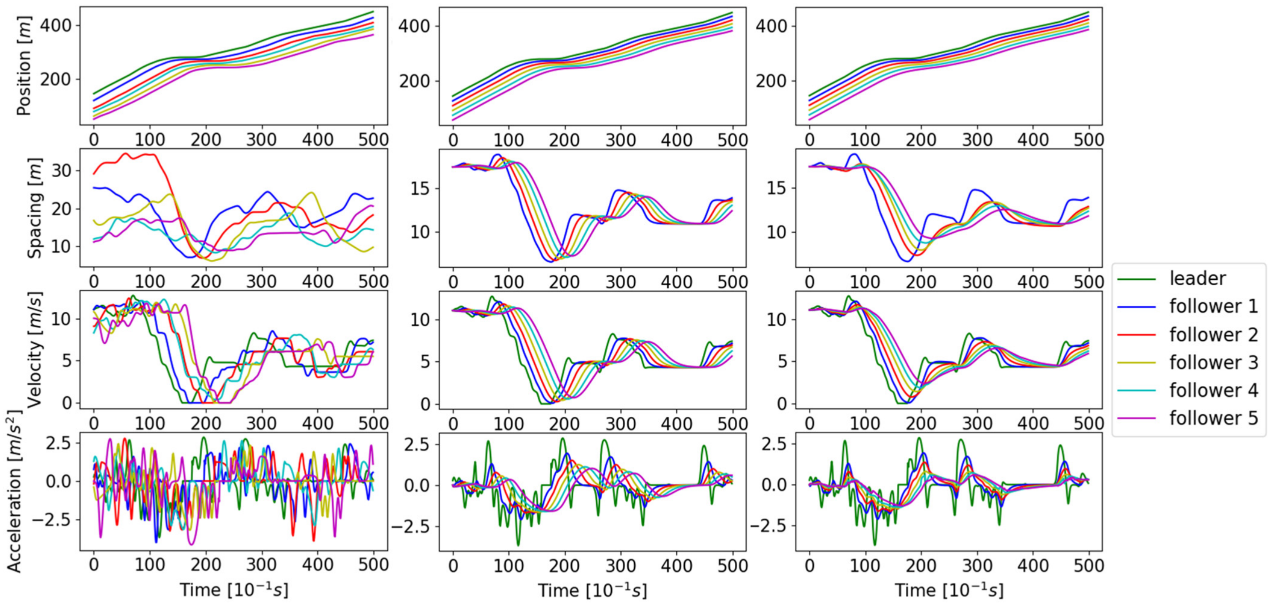
Fig. 4-8. Sensitivity analysis of threshold q regarding (a): average jerk and communication utilization rate; (b): dampening ratio, cooperative control efficiency, and local control efficiency.

Taking different aspects regarding signal smoothness, oscillation dampening performance, and control efficiency into consideration, the optimal range of the threshold q can be found as 0.03 to 0.09. In this study, $q^* = 0.03$ is adopted for the follow-up experiments due to the relatively small jerk and sufficient control performance.

4.4.2 Control Performance Evaluation

Based on the optimal threshold q^*, we systematically compared the proposed DRL-based distributed control strategy's performance with a DRL-based decentralized control strategy (i.e., single predecessor-follower topology) and a linear-based CACC strategy in this section. A six-vehicle platoon' trajectory from the NGSIM dataset was used in the experiment, where the first vehicle trajectory is selected as the leading vehicle trajectory of these strategies.

As an illustrative example, Fig. 4-9 shows the six-vehicle platoon's trajectories of the field data and simulated results based on the decentralized strategy and distributed control strategy. The acceleration of HDVs in Fig. 4-9 (a) fluctuates significantly due to the traffic oscillations, leading to traffic congestion during 25 seconds to 35 seconds. In contrast, the CAV follower in Fig. 4-9 (b-c) is more responsive to the preceding CAV with smaller spacing and smoother realized acceleration, indicating great car following efficiency and driving comfort. Compared with the decentralized control, which merely utilizes the very nearest following vehicle's information, the distributed control strategy in Fig. 4-9 (c) significantly improves the dampening performance with the magnitude of spacing, velocity, and acceleration attenuated through the platoon, indicating more excellent string stability and cooperative control efficiency. Moreover, we assessed the decision-making time of the proposed DRL-based method. The average decision-making time (the average result of 500 timesteps) of follower 5' s DRL controller takes 0.236 milliseconds (0.023 standard deviations), which meets the decision-making requirements for CAVs.



(a) (b) (c)

Fig. 4-9. The position, velocity, and realized acceleration results of a vehicle platoon: (a) NGSIM ground-truth data; (b) Simulated results using decentralized control strategy; (c) Simulated results using distributed control strategy.

For quantitative evaluation, Fig. 4-10 demonstrates evaluation indexes of the three vehicular platoons, including dampening ratio $d_{p,i}$ (in Equation (4-24)), comfort cost g_i^t (in Equation (4-15)), cooperative control efficiency cost f_i^t (in Equation (4-13)), and local control efficiency cost $\hat{f}_i^t$, among which last three indicators are based on the average value per time step. The dampening ratio of each HDV (NGSIM data) remains around 1.0. On the other hand, the dampening ratio and driving comfort cost of the linear-based CACC approach, decentralized control approach, and proposed distributed control approach show a similar downward trend through the platoon, satisfying the strict string stability criteria and indicating the improved driving comfort. Despite the similar tendency, the distributed control strategy demonstrates a smaller dampening ratio and driving comfort cost, and the advantage to the decentralized control and linear-based CACC becomes more obvious towards the traffic upstream. By the above results, we can conclude that the distributed CAV controller receives downstream multi-vehicle information and achieves equilibrium consensus, which outperforms the decentralized control and linear CACC in smoothing the driving behavior and alleviating traffic oscillations.

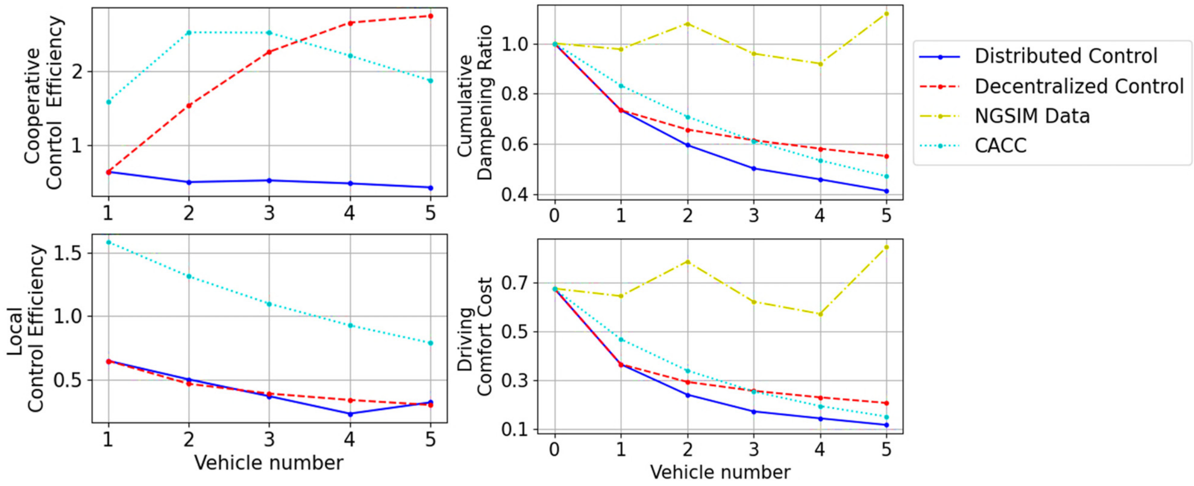


Fig. 4-10 Comparison of distributed control, decentralized control, and NGSIM data in terms of four evaluation indexes.

On the other hand, both decentralized and distributed control shows a small cost of local control efficiency (below 1.0), indicating that local stability is well guaranteed. However, the decentralized control shows a growing trend in the cooperative control efficiency cost, while the distributed control demonstrates much smaller and attenuated costs through the platoon. This proves that the decentralized control cannot achieve the consensus due to the large deviation for the CAVs far apart. The distributed control strategy achieves the consensus, which contributes to the stable platoon dynamics with little deviations from the consensus equilibrium. Furthermore, the proposed DRL-based control strategy outperforms the linear-based CACC strategy in every aspect, which can be attributed to two reasons. First, the DRL can better capture stochastic vehicle behaviors since the real ground-truth dataset (NGSIM) is embedded in the training framework, whereas (Zhou et al., 2019) may neglect some nuanced vehicle driving characteristics. In addition, the specially designed DRL state and reward function incorporate the merits of the multi-agent consensus by fusing the weighted information of the downstream CAVs, which further improves the control performance and the dampening performance from a system level. Besides, the system-level control performances further enhance driving safety based on the equilibrium concept. To demonstrate the safety performance more intuitively, we defined the ‘safety cost’ as the reciprocal of TTC (i.e., $1/TTC_i^t$) and plotted the safety cost trajectory for the vehicle platoon, as presented in Fig. 4-11**.** According to the previous studies (Sultan et al., 2002), the safety cost threshold of 0.5 (i.e., TTC threshold as two seconds) is adopted to identify whether the car following behavior is safe or not. In Fig. 4-11, the safety cost trajectories are below 0.5 under any condition, which further proves that the proposed controller enhances safety performance.

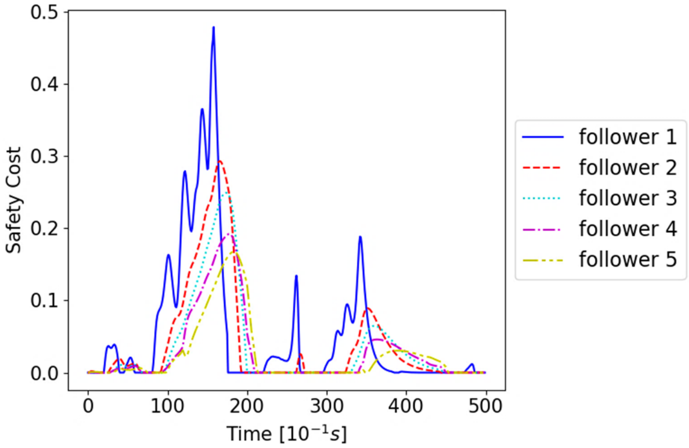
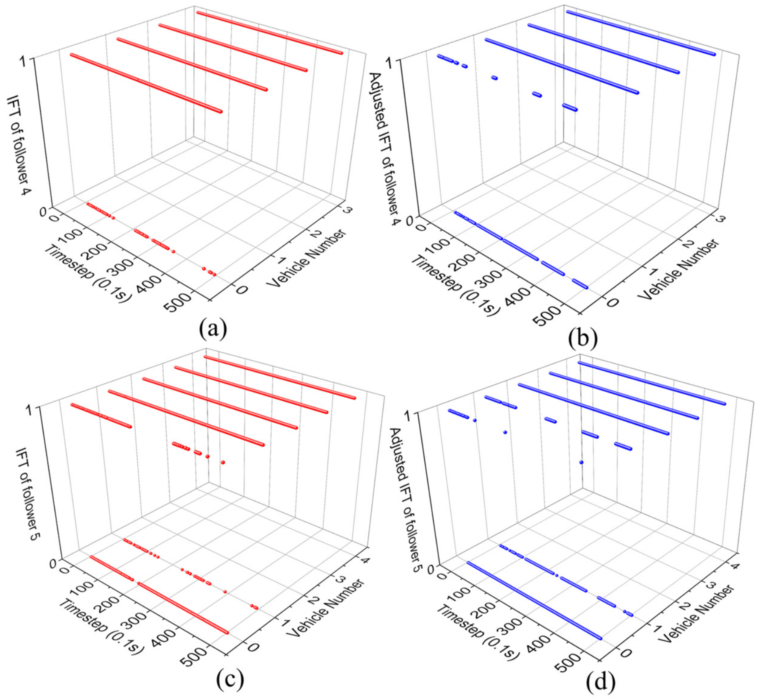


Fig. 4-11. The safety cost trajectory of the vehicle platoon

Moreover, to visualize our dynamic information fusion mechanism and the unstable communication status of the vehicular platoon, the IFT and the adjusted IFT of follower 4 and follower 5 are given in Fig. 4**-**12. It demonstrates the communication statuses between the receiver (follower 4 and follower 5) and downstream transmitters (vehicle 3 – vehicle 0 for follower 4; vehicle 4 – vehicle 0 for follower 5) during the whole period. It shows that the communication becomes unstable as the distance of the transmitter and receiver increases. In current settings, the successful information delivery of the transmitter $CAV\ i-m$ cannot be guaranteed when $m>3$ (e.g., vehicle 0 to follower 4; vehicle 0 and vehicle 1 to follower 5), which negatively affects the control implementation and performance. By the information fusion mechanism, the acceleration of follower 4 and follower 5, presented in Fig. 4-9 (c), ensures smooth control and simultaneously achieves desired control performance.



(a)

(b)

(c)

(d)

Fig. 4-12. IFT and adjusted IFT of the follower 4 and follower 5: (a) IFT of follower 4; (b) Adjusted IFT of follower 4; (c) IFT of follower 5; (d) Adjusted IFT of follower 5

4.4.3 Generalization Analysis

After evaluating the great control performance of the proposed approach, this section validates its generalization capability through validating the statistical robustness and demonstrating extended experimental cases.

Statistical Robustness Validation

To validate the statistical robustness, multiple field trajectories from NGSIM datasets (180 ground-truth datasets with each over 30 seconds) are selected to quantitatively evaluate the generalized performance of our proposed strategy for different leading vehicles trajectories. The detailed performance indexes are as presented in Fig. 4-13. As can be found that, the dampening ratio (give in Equation (4-24)) starts at 1 and gradually decreases to 0.335 through five distributed CAVs, proving that the oscillations are weakened during traffic propagation. From the control perspective, the monotonically decreased running cost (given in Equation 4-15) demonstrates that the proposed strategy can make CAVs maintain close to the predefined equilibrium point through the vehicular string. Moreover, upstream CAVs can better achieve equilibrium consensus due to the weakened oscillations and more comprehensive information received.

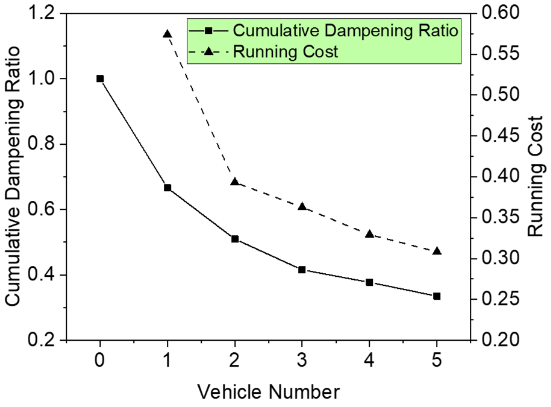


Fig. 4-13. Average cumulative dampening ratio and running cost of each vehicle in a six-vehicle platoon.

Further, we statistically show the advantage of the proposed strategy over ground-truth data and other approaches mentioned above (i.e., decentralized control; linear-based CACC). The superiority percentage P is introduced as follows:

$$P = \frac{PI_o - PI_d}{PI_o} * 100\%, \tag{4-25}$$

where PI_o and PI_d are the generalized performance index of the compared strategies and proposed strategy, respectively. For generalizing these indexes, we first averaged five followers' performance indicators in the platoon as representation for the whole platoon and then took the average result of the 180 datasets. Thus, the superiority percentage of each indicator is calculated and demonstrated in Fig. 4-14. The distributed control outperforms the field data with 55.92% in dampening ratio and 76.05% in driving comfort, demonstrating significant improvement of string stability. Compared with the decentralized control and linear-based CACC, the distributed control has advantages in every aspect, especially in the cooperative control efficiency (74.96% and 78.3% in cost reduction, respectively).

These results further validate the generalization capability of the proposed method for different leading vehicle trajectories.

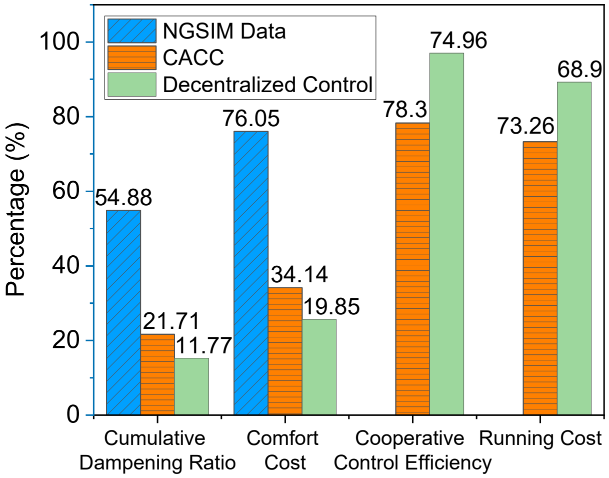


Fig. 4-14. The superiority percentage of proposed strategy compared with NGSIM data and decentralized control strategy.

Extended Cases

Finally, extended experiments are conducted to demonstrate the robustness of our controllers under different traffic conditions. Fig. 4-15 shows acceleration profiles of ten-follower vehicular platoon in six different scenarios, whose leading vehicle trajectories, picked from NGISM, are different. The results show that the proposed control method quickly and significantly dampens traffic oscillations, efficiently stabilizing the driving behaviors of the upstream CAVs.

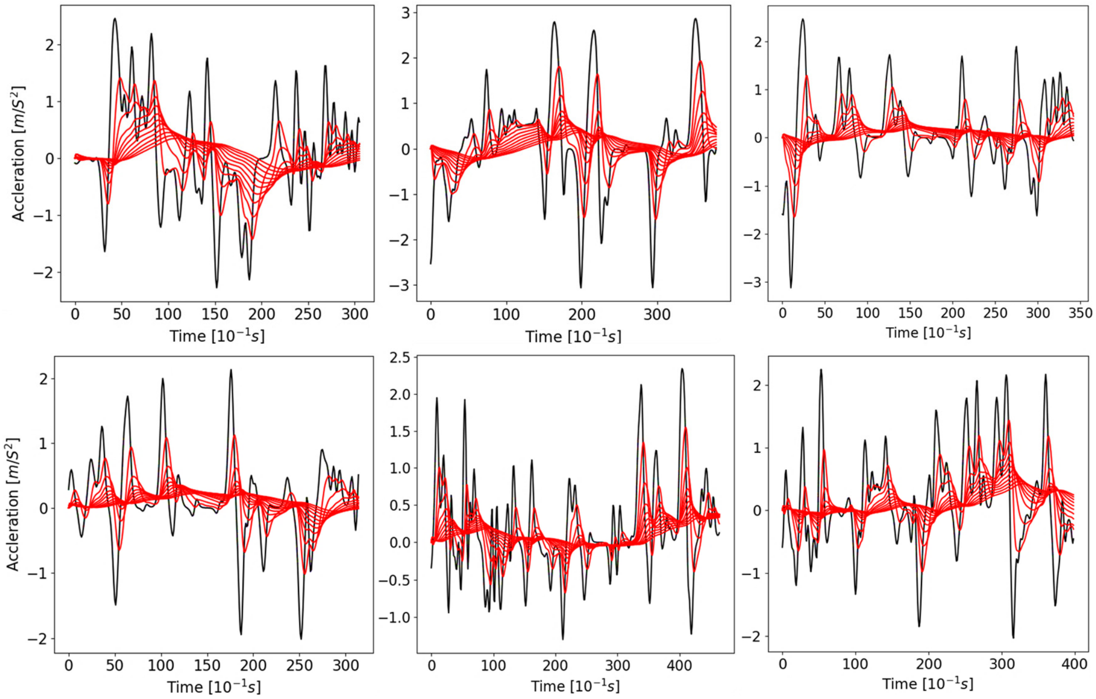


Fig. 4-15. Acceleration profiles of ten-follower vehicular platoon in six different scenarios.

Further, the proposed control strategy is applied to control 50 CAV followers with a customized leading vehicle trajectory incorporating one rapid deceleration-acceleration cycle (-2.4 m/s^2 – 1.5 m/s^2) disturbance. The 50-HDV platoon generated by the IDM is set as a comparison, where the IDM model is calibrated by (Kesting & Treiber, 2008) based on the datasets with complex situations. Fig. 4-16 illustrates the velocity heat map of the CAV platoon and HDV platoon, and Fig. 4-17 intuitively demonstrates the velocity and acceleration portfolios of the CAV platoon. For the HDV platoon, the traffic oscillation is amplified upstream with increasing traffic jams. However, for the CAV platoon, the disturbance is dampened significantly through the platoon so the followers can quickly recover from the disturbance, showing outstanding robustness and resilience. The velocity and acceleration portfolios also demonstrate the excellent dampening performance of the CAV platoon.

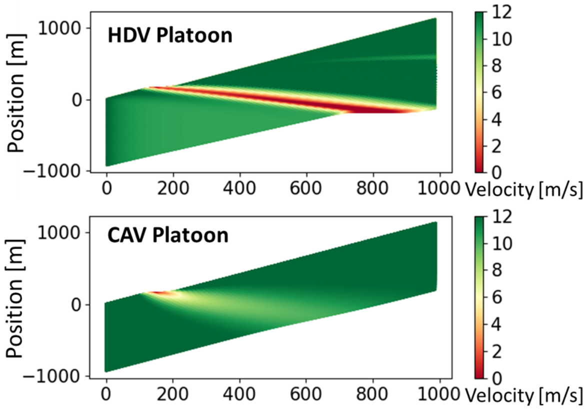


Fig. 4-16. Velocity heat map based on the position trajectories for the HDV platoon and CAV platoon.

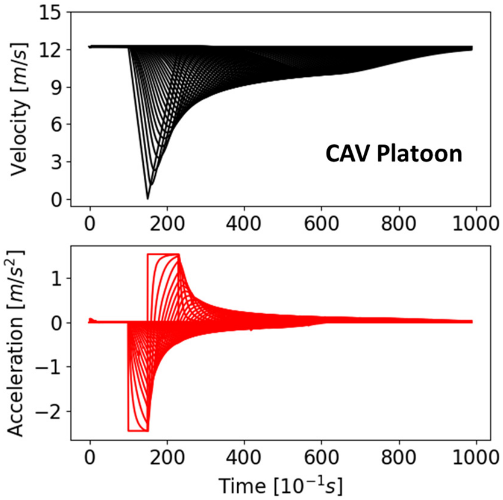


Fig. 4-17. Velocity and acceleration trajectories of the CAV platoon under rapid deceleration-acceleration scenario.

Since the proposed controller is trained by the NGSIM dataset, which is normally low-speed congested, we conduct experiments to simulate a 20-vehicle platoon under the medium-speed speed scenario (steady state 20 m/s) and high-speed scenario (steady state 30 m/s), with results presented in Fig. 4-18. Similarly, one typical rapid deceleration-acceleration cycle (i.e., -2.5 m/s^2 – 2.5 m/s^2 for medium-speed scenario; -3 m/s^2 – 3 m/s^2 for high-speed scenario) is customized to represent a large-amplitude traffic oscillation for the leading vehicle. The results demonstrate that the controller greatly dampens the traffic oscillation in both medium-speed and high-speed scenarios, which further validates the generalizable capability.

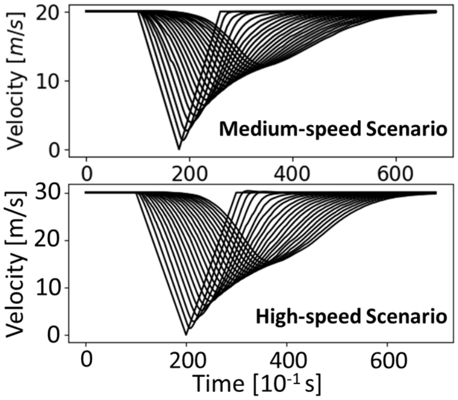


Fig. 4-18. Velocity trajectories of the CAV platoon in the medium-speed and high-speed scenarios.

4.5 Conclusion

This study presents a DRL-based generic distributed CAV longitudinal control approach in a relatively realistic communication environment. To better capture stochastic characteristics of the preceding vehicles and communication loss, we embed the NGSIM datasets and the SINR based dynamic communication mechanism into the training framework. Each CAV in the framework receives its downstream CAVs' fused information as the DRL state for real-time control. The fused DRL state and reward function are specially designed to incorporate the merits of the equilibrium concept and consensus concept, which maintains CAVs around the predefined equilibrium point and achieves the system-level consensus to better dampen traffic oscillations. A dynamic information fusion mechanism is proposed to smooth the fluctuated DRL state and the high-jerk control signal caused by the dynamic communication loss.

For evaluating the proposed strategy, we conducted several numerical experiments using NGSIM datasets. The sensitivity analysis was conducted first to optimize the parameter of the dynamic information fusion mechanism. Then the control performance of the distributed control approach is evaluated by comparing with the decentralized control, linear control, and field data in terms of dampening ratio, driving comfort, and cooperative car following efficiency. The results suggest that the distributed control strategy significantly outperforms other strategies and field data in every aspect and can greatly stabilize the traffic oscillations based on the platoon's equilibrium consensus, demonstrating its robustness and resilience against disturbances. Finally, the generalization capability of the proposed strategy is validated using large amounts of the NGSIM datasets and customized traffic scenarios.

Some future studies can be investigated based on current results. The CAV lateral control can be incorporated in the control framework for merging, diverging or lane-changing maneuvers. In addition, other dynamic or validated communication models (Kim et al., 2017; Wang et al., 2019) or topologies (e.g., relay communication topology, V2I, V2C) can be embedded in the framework to conduct extended experiments. The dynamic communication delay can be considered to make the control framework more realistic. Moreover, the complex mixed traffic flow properties can be further studies and optimized based on this study by extending the control framework.

Furthermore, we can incorporate the prediction process (i.e., predicting the behavior of the surrounding vehicles) into the control framework to achieve more efficient control performance. The prediction process using advanced supervised machine learning algorithms (Ahmadlou & Adeli, 2010; Alam et al., 2020; Pereira et al., 2020; Rafiei & Adeli, 2017) are considered as the extension based on the current control framework.

5 Distributed Connected Automated Vehicles Control in mixed traffic: Real-time Aggregated Macroscopic Car-following Behavior Estimation based on Deep Reinforcement Learning

This chapter proposes an innovative distributed longitudinal control strategy for connected automated vehicles (CAVs) in the mixed traffic environment of CAV and human-driven vehicles (HDVs), incorporating high-dimensional platoon information. The control strategy is designed for the mixed traffic environment discussed in Section 4 of the thesis. For mixed traffic, the traditional CAV control method focuses on microscopic trajectory information, which may not be efficient in handling the HDV stochasticity (e.g., long reaction time; various driving styles) and mixed traffic heterogeneities. Different from traditional methods, our method, for the first time, characterizes consecutive HDVs as a whole (i.e., AHDV) to reduce the HDV stochasticity and utilize its macroscopic features to control the following CAVs. The new control strategy takes advantage of platoon information to anticipate the disturbances and traffic features induced downstream under mixed traffic scenarios and greatly outperforms the traditional methods. In particular, the control algorithm is based on deep reinforcement learning (DRL) to fulfill car-following control efficiency and further address the stochasticity for the aggregated car following behavior by embedding it in the training environment. To better utilize the macroscopic traffic features, a general platoon of mixed traffic is categorized as a CAV-HDVs-CAV pattern and described by corresponding DRL states. The macroscopic traffic flow properties are built upon the Newell car-following model to capture the characteristics of aggregated HDVs' joint behaviors. Simulated experiments are conducted to validate our proposed strategy. The results demonstrate that the proposed control method has outstanding performances in terms of oscillation dampening, eco-driving, and generalization capability. Finally, we further analyze the vehicle sequencing’s impact on the mixed traffic flow, which has rarely been discussed in previous researches. This will provide guidance and reference for future research that considers lane-changing maneuvers.

The chapter is organized as follows. Section 5.1 provides the CAV longitudinal control scheme, including the environment setting, the distributed control scheme, and the state fusion strategy. Section

5.2 gives the details of DRL-based control model development, in which the basics of the DRL algorithm are discussed in Section 5.2.1; the policy updating algorithm is given in Section 5.2.2, and the training procedure is described in Section 5.2.3. Section 5.3 analyzes the results of simulated experiments in terms of control performance, driving comfort, and generalization capability. Section 5.4 analyzes the impact of CAV-HDV topologies on mixed traffic flow. Section 5.5 concludes the work.

5.1 CAV Control Scheme

5.1.1 Assumptions and Environment Setting

Assumptions

This study focuses on the CAV longitudinal control in mixed traffic of CAVs and HDVs. We consider the car following process without lateral movement in an infinite highway segment. The communication between CAVs follows the Federal Communications Commission, allocating a dedicated short-range communication (DSRC) radio with a 5.9-GHz frequency (Du & Dao, 2015). The environment assumptions are given as follows: (i) The CAV can obtain the real-time state information (e.g., speed, position) of its immediate preceding vehicle using onboard sensors. (ii) The CAV can receive its own real-time state information. (iii) The CAV's real-time state information is broadcasted to the upstream CAVs by vehicle-to-vehicle (V2V) communication. (iv) The Signal-Interference-plus-Noise Ratio (SINR) condition dynamically determines the transmission status (fail/success) between any CAV pairs. (v) The communication delay is negligible due to the short communication distance in a road segment. (vi) HDVs have no communication capability. (vii) The lane-changing maneuvers are not considered in the vehicular platoon.

Communications

For the DSRC-based V2V communication environment given in the above assumptions, the information transmission status between CAVs can change dynamically under communication failure due to communication interference or information congestion (Wang et al., 2019). The communication failure will undermine driving performance. We embed this realistic communication property into the control framework to enhance the robustness and practicality of the CAV controller. The information flow topology (IFT), which indicates the dynamic transmission status of links in the vehicular platoon, is described from the receiver side (i.e., controlled CAV) to illustrate the communication environment. Specifically, the IFT of CAV i at timestep t is defined as $\boldsymbol{\xi}_{\boldsymbol{i}}^{\boldsymbol{t}} = [\eta_{i,i-1}^{t}, \eta_{i,i-2}^{t}, \dots, \eta_{i,i-N}^{t}]$, where $\eta_{i,i-m}^{t} \in \{0, 1\}$ denotes the information transmission status between the receiver CAV i and the downstream vehicle $i-m$. $\eta_{i,i-m}^{t} = 1$ indicates a successful transmission, while $\eta_{i,i-m}^{t} = 0$ can happen either when a communication loss or vehicle $i-m$ is an HDV. In addition, we assume a permanently successful transmission status for the immediate preceding vehicle (i.e., $\eta_{i,i-1}^{t} \equiv 1$) due to CAV's robust onboard sensors. To better replicate the DSRC-based V2V communication, the SINR communication model (Du & Dao, 2015), which demonstrates great estimation of communication loss on a one-way road segment, is adopted to identify CAVs' IFTs. The SINR model determines the real-time transmission quality $y_{i,j}^{t}$ between the transmitter CAV j and the receiver CAV i at timestep t, defined as Equation (5-1):

$$y_{i,j}^{t} = \frac{P_j\left(X_{ji}^{t}\right)^{-\alpha}}{\sum_{k=j+1}^{i-1} P_k (X_{ki}^{t})^{-\alpha} + O}, \tag{5-1}$$

where P_j denotes the transmission power of vehicle i; X_{ji}^{t} is the distance between two CAVs; α is the parameter adjusting the signal power decay. $\sum_{k=j+1}^{i-1} P_k (X_{ki}^{t})^{-\alpha}$ represents the sum of the interference signal power of vehicles between the receiver CAV i and transmitter CAV j. The noise term $O \sim N(\mu, \sigma^2)$ is used to simulate the random noise affecting the communication environment. Based on $y_{i,j}^{t}$, a threshold value β related to the communication capability (e.g., modulation, code rate) is introduced to determine the real-time transmission status $\eta_{i,i-m}^{t}$:

$$\eta_{i,i-m}^{t} = \begin{cases} 1, \ y_{i,i-m}^{t} > \beta \\ 0, \ y_{i,i-m}^{t} \leq \beta \end{cases}. \tag{5-2}$$

Based on the SINR model, we embed critical communication features of the practical condition in the simulated V2V communication environment, making the simulation more realistic.

Vehicle Dynamics

Given the assumptions and communication environment, the vehicle dynamics are modeled by a first-order approximation to capture multiple factors (e.g., gear position, road gradient, air drag force) of the vehicle linearized dynamics (Li et al., 2011; Wang, 2018b):

$$a_i^{t+1} = e^{-\frac{\Delta t}{I_{i,L}}} \times a_i^t + \left(1 - e^{-\frac{\Delta t}{I_{i,L}}}\right) \times K_{i,L} u_i^t, \tag{5-3}$$

where $K_{i,L}$ and $I_{i,L}$ denote the system gain (ratio of the control demand that can be realized) and actuation time lag of CAV i, respectively; u_i^t and a_i^t are the demanded acceleration and realized acceleration. With acceleration a_i^t, the real-time vehicle state is updated according to the kinematic point-mass equations (M. Zhu et al., 2018):

$$v_i^{t+1} = v_i^t + a_i^t \Delta t \tag{5-4a}$$

$$\Delta v_{i,i-1}^{t+1} = v_{i-1}^{t+1} - v_i^{t+1} \tag{5-4b}$$

$$d_{i,i-1}^{t+1} = d_{i,i-1}^t + \frac{\Delta v_{i,i-1}^t + \Delta v_{i,i-1}^{t+1}}{2} \Delta t, \tag{5-4c}$$

where Δt is the update interval; v_i^t is the velocity of CAV i at timestep t; $d_{i,i-1}^t$ denotes the front-bumper distance between CAV i and CAV $i-1$.

5.1.2 Distributed Control Scheme

Based on the above environment setting, this section provides a generic distributed control framework to regulate CAVs' longitudinal movements in a mixed traffic environment, as presented in Fig. 5-1. The communication topology setting assumed by the SINR model in the framework is a common V2V communication topology, which is widely utilized in the CACC control (Wang et al., 2020). For this topology, each controlled CAV (i.e., CAV i) broadcasts its state information (e.g., velocity, position) to the upstream vehicles within a certain communication range and simultaneously communicates with the multiple downstream vehicles within the communication range (i.e., at most K downstream vehicles) at each timestep for real-time longitudinal control. For each timestep, the received information from the downstream vehicles is fused as a weighted DRL state $\boldsymbol{s}_{\boldsymbol{i}}^{\boldsymbol{t}}$, which will be explained with details later. After the fusion process, the DRL-based controller generates the real-time demanded acceleration u_i^t, and u_i^t is then implemented based on the above vehicle dynamics, regulating CAV i's longitudinal movements.

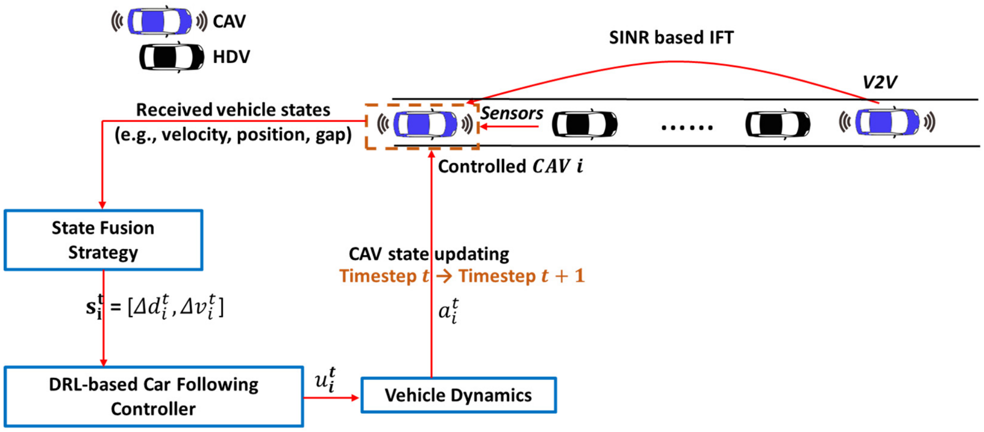
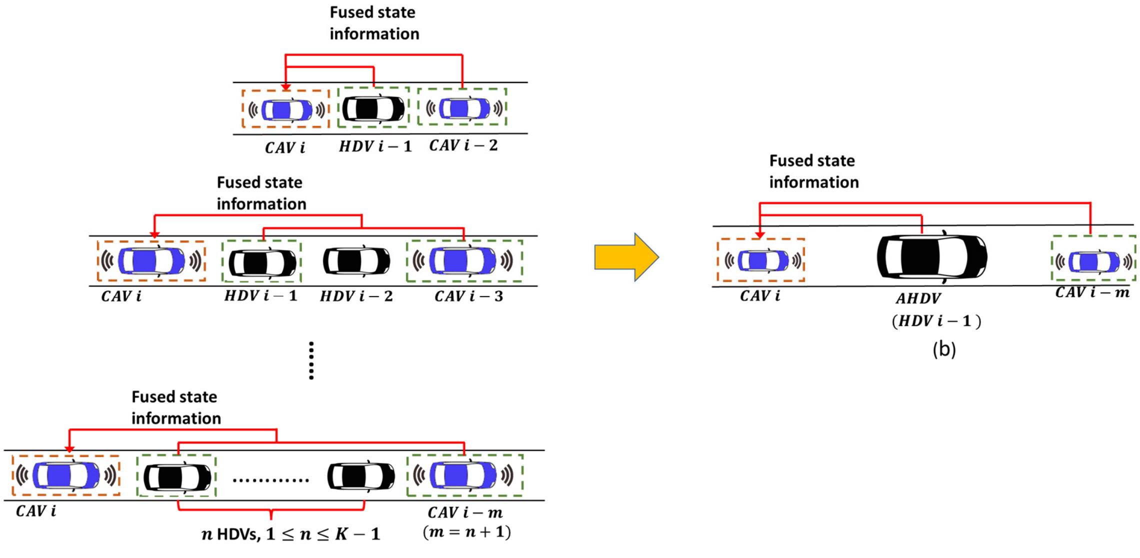


Fig. 5-1. Velocity and acceleration trajectories of the CAV platoon under rapid deceleration-acceleration scenario.

Within the above framework, the fused DRL state $\boldsymbol{s}_{\boldsymbol{i}}^{\boldsymbol{t}}$ is notably designed to better utilize the downstream vehicles' information. The communication range for state fusion, defined as the 'local downstream environment,' is restricted to cover at most K downstream vehicles. The communication quality within the range should be basically stable (i.e., rarely fails), and the kinematic traffic waves (Whitham, 1955) can be quickly propagated to the controlled CAV i. Despite the limited range, the diversified downstream CAV-HDV topologies make developing a generic distributed controller challenging. To this end, we describe any local mixed downstream environment as the generic CAV-HDVs-CAV pattern, which consists of a nearest downstream CAV followed by a single or multiple HDVs, as presented in Fig. 5-2(a). In this heterogeneous local environment, the traffic oscillation amplitude usually grows upstream through the consecutive HDVs between CAV i and CAV $i-m$ (Zhou et al., 2019), which hinders CAV i from driving smoothly. To alleviate this issue, we firstly fuse the nearest downstream CAV (i.e., CAV $i-m$)' s state information to 'actively' anticipate its relatively smooth and stable driving behavior for more efficient control. Furthermore, directly modeling or predicting each HDV's stochastic behavior between CAV i and CAV $i-m$ is very challenging. To this end, we characterize the consecutive HDVs as a whole 'large' HDV (i.e., AHDV) and utilize its macroscopic traffic features to attenuate stochasticity, thus enhancing CAV i' s driving behavior. As presented in Fig. 5-2(b), we neglect each HDV's microscopic driving behavior between the preceding HDV (i.e., HDV $i-1$) and CAV $i-m$ and define this 'CAV-HDVs-CAV' pattern as a novel car-following structure 'CAV i → AHDV → CAV $i-m$.' In this way, CAV i receives the real-time state information of its preceding HDV $i-1$ and the nearest downstream CAV $i-m$ to generate the fused DRL state $\boldsymbol{s}_i^t$ for the DRL-based control. It should be noted that if CAV $i-m$ is out of the communication range (i.e., $m > K$), CAV i only receives the information of HDV $i-1$ for control. The proposed distributed control scheme is downgraded to the 'decentralized control', which will be explained in Section 5.3.2 with details.



(a)

(b)

Fig. 5-2. The mixed local downstream environment: (a) characterized as 'CAV-HDVs-CAV' pattern; (b) "three-vehicle" car-following structure

5.1.3 State Fusion Formulation

A generic state fusion strategy is designed based on the equilibrium concept to regulate each CAV close to the pre-defined equilibrium state and meanwhile effectively stabilize traffic oscillations. The equilibrium concept from the modern control theory defines the equilibrium state (equilibrium point) for a dynamical system, which represents a state where the system can stabilize after being affected by external disturbances or forces (Absil & Kurdyka, 2006). A system will remain at the (stable) equilibrium state once reached, given the perturbation and inputs are small enough.

In longitudinal car-following control, the equilibrium state represents the desired ideal vehicle state (i.e., equilibrium spacing and speed) during driving for each car following pair, which avoids arbitrary variation of the inter-vehicle spacing for control. Incorporating the equilibrium concept in DRL provides the exploration direction in DRL training to help develop a robust control policy and gives the base for analyzing vehicle string stability and car following control efficiency. Based on the concept, this subsection derives the DRL state $\boldsymbol{s}_i^t$ as the weighted deviation from the equilibrium spacing Δd_i^t and the weighted deviation from the equilibrium speed Δv_i^t regarding its downstream vehicles HDV $i - 1$ and CAV $i - m$. Four parameters are predefined to fuse the DRL state, including the equilibrium spacing $d_{i,i-1}^{*t}$ and equilibrium speed $v_{i,i-1}^{*t}$ regarding HDV $i - 1$; the equilibrium spacing $d_{i,i-m}^{*t}$ and equilibrium speed $v_{i,i-m}^{*t}$ regarding CAV $i - m$. The following content describes the derivations of the DRL state.

Local Equilibrium

The derivation of the DRL state starts with the local equilibrium state. The local equilibrium state for each CAV car following pair follows the constant time gap (CTG) policy from the Society of Automotive Engineer Standard (SAE), which regulates the CAV to set the same speed as its preceding vehicle and maintain the preset equilibrium spacing, defined as below:

$$d_{i,i-1}^{*t} = v_i^t \tau_i^* + l_i, \tag{5-5a}$$

$$v_{i,i-1}^{*t} = v_{i-1}^t, \tag{5-5b}$$

where v_i^t denotes CAV i's real-time velocity at timestep t; τ_i^* and l_i are the constant time gap and the standstill spacing between CAV i and vehicle $i-1$, respectively. The two equations above define the local equilibrium for a car following pair.

Multi-agent Equilibrium

Furthermore, to consider the impact of multiple vehicles in the local downstream environment, Equation (5-5a) and Equation (5-5b) are expanded to a distributed multi-agent version, whose equilibrium spacing $d_{i,i-m}^{*t}$ and speed $v_{i,i-m}^{*t}$ between CAV i and any downstream vehicle $i-m$ is specified as Equation (5-6a) and Equation (5-6b):

$$d_{i,i-m}^{*t} = v_i^t T_{i,i-m}^* + L_{i,i-m}, \tag{5-6a}$$

$$v_{i,i-m}^{*t} = v_{i-m}^t. \tag{5-6b}$$

$T_{i,i-m}^*$ and $L_{i,i-m}$ denote the equilibrium time gap and standstill spacing between CAV i and CAV $i-m$. The two terms regulate CAV i's desired microscopic driving behavior considering its downstream vehicles, which will be explained with mathematical details later. To facilitate the system-optimal consensus of the vehicular platoon, we measure and embed the actual spacing and speed deviations from the equilibrium between CAV i and CAV $i-m$ into the DRL framework, which is specified as:

$$\Delta d_{i,i-m}^t = d_{i,i-m}^t - d_{i,i-m}^{*t}, \tag{5-7a}$$

$$\Delta v_{i,i-m}^t = v_{i-m}^t - v_i^t, \tag{5-7b}$$

Based on Equation (5-7), the equilibrium deviations for multiple downstream vehicles can be determined for the state fusion.

Estimation from Newell's Car-following Model

The remaining problem lies in calculating the equilibrium spacing $d_{i,i-m}^{*t}$ in Equation (5-6a), which needs to specify the corresponding time gap $T_{i,i-m}^{*}$ and the standstill spacing $L_{i,i-m}$. Specifically, for the defined 'CAV-AHDV-CAV' structure, the equilibrium spacing regarding HDV $i-1$ (i.e., $d_{i,i-1}^{*t}$) and CAV $i-m$ (i.e., $d_{i,i-m}^{*t}$) needs to be configured. Regarding HDV $i-1$, $d_{i,i-1}^{*t}$ is directly given in Equation (5-5). Based on that, the deviation from equilibrium spacing is specified as $\Delta d_{i,i-1}^{t} = d_{i,i-1}^{t} - d_{i,i-1}^{*t}$. Regarding CAV $i-m$, the equilibrium time gap $T_{i,i-m}^{*}$, standstill spacing $L_{i,i-m}$, and equilibrium spacing $d_{i,i-m}^{*t}$ are defined as follows based on the three-vehicle following scheme 'CAV i → HDV $i-1$ → CAV $i-m$':

$$T_{i,i-m}^{*} = \tau_{i}^{*} + T_{i-1,i-m}^{*t}, \tag{5-8a}$$

$$L_{i,i-m} = l_{i} + L_{i-1,i-m}^{t}, \tag{5-8b}$$

$$d_{i,i-m}^{*t} = v_{i}(\tau_{i}^{*} + T_{i-1,i-m}^{*t}) + \left(l_{i} + L_{i-1,i-m}^{t}\right), \tag{5-8c}$$

where $T_{i-1,i-m}^{*t}$ and $L_{i-1,i-m}^{t}$ represent the time-varying time gap and spacing between HDV $i-1$ and CAV $i-m$, respectively. Considering the aggregated HDVs in-between, $T_{i-1,i-m}^{*}$ can be denoted as:

$$T_{i-1,i-m}^{*t} = \sum_{j=1}^{m-1} \tau_{i-j}^{*t}, \tag{5-9}$$

where τ_{i-j}^{*t} denotes the time gap between HDV $i-j$ and its preceding vehicle. Since HDV has inherent stochastic nature with great diversities, τ_{i-j}^{*t} is time-varying and follows varied distributions for different HDVs. Moreover, τ_{i-j}^{*t} is unmeasurable due to the lack of communication capability of HDVs, making it challenging to determine $T_{i-1,i-m}^{*t}$. Compared with a single HDV's microscopic behavior, the aggregated HDV driving behaviors exhibit macroscopic traffic flow properties, which show less stochasticity. Thus, rather than measuring τ_{i-j}^{*t} individually, the aggregated HDV driving characteristics can be better captured by the macroscopic traffic features to address stochasticity, as indicated by the philosophy of central limit theorem (CLT) (Kwak & Kim, 2017). Though m may not be sufficiently large to apply CLT, the aggregation treatment of multiple HDVs brought promises to capture HDVs features. Moreover, we leave the remaining uncertainties by embedding the field-measured HDV trajectories in the DRL training process.

Precisely, the two time-varying terms $T_{i-1,i-m}^{*t}$ and $L_{i-1,i-m}^{t}$ need to be real-time estimated in the state fusion process to capture the macroscopic features. The schematic diagram for the real-time estimation is presented in Fig. 5-3(c). Newell's car following model (Newell, 2002), which bridges the fundamental diagram and microscopic driving behavior features, and meanwhile efficiently models the kinematic oscillation waves (Richards, 2013), is adopted after modification, by allowing the two time-varying terms $T_{i-1,i-m}^{*t}$ and $L_{i-1,i-m}^{t}$ to be time-variant and real-time estimated. Rather than directly modeling the microscopic driving behaviors of any CAV or HDV, the time-varying version of Newell's car-following model describes the aggregated HDVs' (AHDV's) driving behavior to capture its exhibited macroscopic traffic features in real-time.

Fig. 5-3(a) and Fig. 5-3(b) demonstrates the principle of Newell's car-following model. From the microscopic perspective, Newell's car-following model gives a linear speed-spacing relationship in congested traffic flow for the following vehicle i, which assumes the follower reproduces the preceding leader's trajectory with a time-space displacement (τ, l):

$$d_i = v\tau + l, \tag{5-10}$$

where v is the vehicle speed; d_i is the spacing; τ represents the time shift for vehicle i to match its leader's speed; l denotes the displacement of the speed change point. Moreover, from a macro perspective, the Newell's car following model models the kinematic wave with a triangular fundamental diagram, in which the parameters τ and l represent macroscopic traffic features to describe traffic wave speed w and jam density k:

$$w = \frac{1}{\tau k}, \tag{5-11a}$$

$$k = \frac{1}{l} = \frac{1}{w\tau}, \tag{5-11b}$$

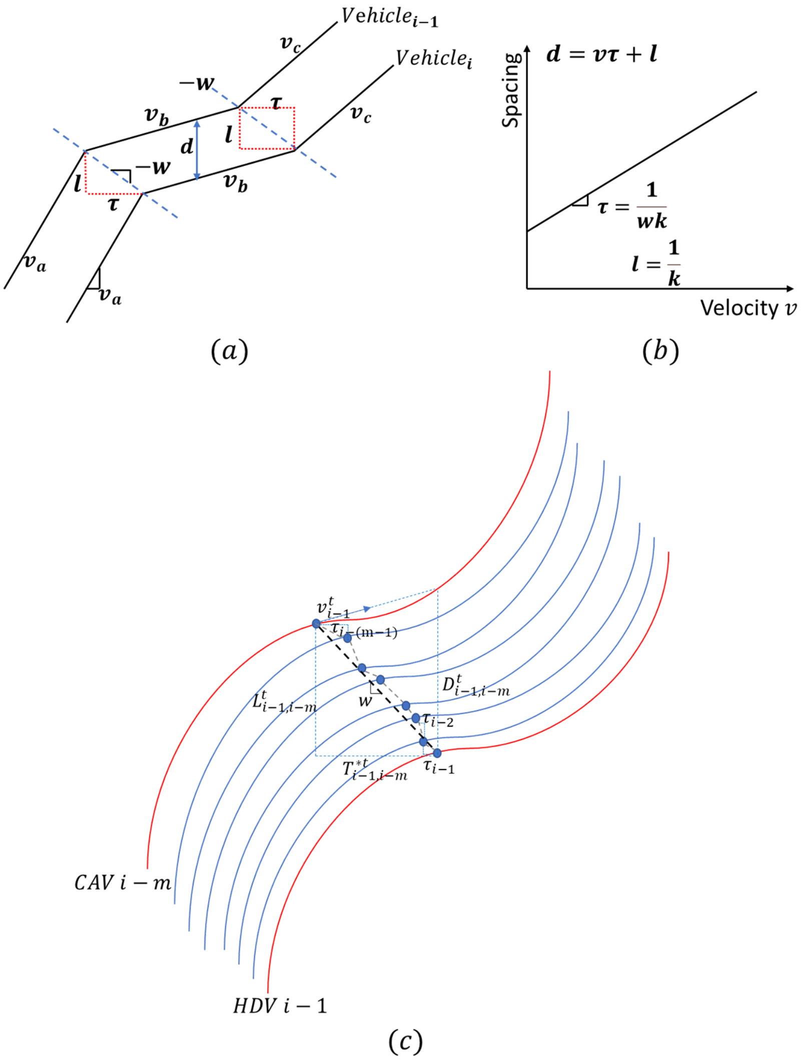


Fig. 5-3. Model schematic diagram: (a) Newell's car following model; (b) Speed-spacing relationship; (c) Real-time estimation diagram of the time-varying time-gap $\boldsymbol{T^{*t}_{i-1,i-m}}$ and spacing $\boldsymbol{L^{t}_{i-1,i-m}}$

where τ denotes the wave propagating time between two consecutive vehicles; l indicates the jam spacing. Based on Equation (5-10) and Equation (5-11), τ and l are two key terms for modeling the car-following behavior and simultaneously capturing the macroscopic features. Furthermore, for the vehicle following structure 'CAV-AHDV-CAV' described above, the time gap $T^{*t}_{i-1,i-m}$ and spacing $L^{t}_{i-1,i-m}$ can be interpreted as Newell's parameters τ and l in the car following pair 'HDV $i-1$ → CAV $i-m$', which anticipates the relatively smooth behavior of CAV $i-m$ and incorporates macroscopic features of the aggregated HDVs. Specifically, Equation (12) is proposed to real-time estimate $T_{i-1,i-m}^{*t}$ and $L_{i-1,i-m}^{t}$ based on the integration form of Equation (10) and Equation (11):

$$T_{i-1,i-m}^{*t} = \frac{D_{i-1,i-m}^{t}}{w+v_{i-1}^{t}}, \tag{5-12a}$$

$$L_{i-1,i-m}^{t} = wT_{i-1,i-m}^{*t}, \tag{5-12b}$$

where v_{i-1}^{t} is the speed of HDV $i-1$ at timestep t; $D_{i-1,i-m}^{t}$ is the actual spacing between HDV $i-1$ and CAV $i-m$; w denotes the average kinematic wave speed, which is a pre-calibrated value determined by the road infrastructure's features and configuration. Since w plays an important role in the method, applications of our methods should regularly measure and update the w value. There are methods available for w measurement, including direct measurement w using wavelet transform (Zheng et al., 2011; Zheng & Washington, 2012), or indirect estimation by first estimating the fundamental diagram to derive w per Li et al., (2022). We set w to 16 km/h due to the generalized settings in studies using Next Generation Simulation (NGSIM) data collected on eastbound I-80 with an on-ramp at Powell Street (e.g., Laval & Leclercq, 2010; Duret et al., 2011; Chen et al., 2012).

In addition, it should be noted that $T_{i-1,i-m}^{*t}$ and $L_{i-1,i-m}^{t}$ are not derived from the steady-state spacing defined in Chen et al. (2012). We use the actual spacing $D_{i-1,i-m}^{t}$ to approximate the steady-state spacing for real-time estimation, which better anticipates the actual disturbances from downstream (e.g., sudden change in spacing) and thus achieves adaptive control performances. Moreover, the estimation method for $T_{i-1,i-m}^{*t}$ and $L_{i-1,i-m}^{t}$ are only suitable for heavily congested traffic conditions, where the traffic oscillations are continuously propagated upstream.

DRL State Fusion

Based on the estimated time-varying gap $T_{i-1,i-m}^{*t}$ and spacing $L_{i-1,i-m}^{t}$ in Equation (5-12), the equilibrium spacing $d_{i,i-m}^{*t}$ between CAV i and CAV $i-m$ in Equation (5-8c) can be real-time determined. Thus, the deviation between actual spacing and equilibrium spacing is defined as $\Delta d_{i,i-m}^{t} = d_{i,i-m}^{t} - d_{i,i-m}^{*t}$. To better regulate the CAVs close to the equilibrium and reduce the DRL state dimension for greater training performance, the DRL state $\boldsymbol{s}_{\boldsymbol{i}}^{\boldsymbol{t}}$ = $[\Delta d_i^t, \Delta v_i^t]$ is generated by fusing the weighted equilibrium deviations of $HDV\ i-1$ and $CAV\ i-m$:

$$\Delta d_i^t = \frac{q_{i-1}\eta_{i,i-1}^t \Delta d_{i,i-1}^t + q_{i-m}\eta_{i,i-m}^t \Delta d_{i,i-m}^t}{q_{i-1}\eta_{i,i-1}^t + q_{i-m}\eta_{i,i-m}^t}, \tag{5-13a}$$

$$\Delta v_i^t = \frac{q_{i-1}\eta_{i,i-1}^t \Delta v_{i,i-1}^t + q_{i-m}\eta_{i,i-m}^t \Delta v_{i,i-m}^t}{q_{i-1}\eta_{i,i-1}^t + q_{i-m}\eta_{i,i-m}^t}, \tag{5-13b}$$

where weights q_{i-1} and q_{i-m} represent the information importance for HDV $i-1$ and CAV $i-m$, respectively. The coefficient for the two components is computed based on the function $q_{i-j} = \begin{cases} \frac{1}{2},\ j=1 \\ \frac{1}{2},\ j=m \end{cases}$. The sum of all the weights should be equal to 1 without loss of generality. Since the information of both the two components is critical, we make it decay with 1/2 order to give equal weight to the two components based on the above function. Despite the equal weight settings in this study, the weights can be adjusted to balance the impact of the two components. The impact of both components can be summarized as follows. The preceding HDV $i-1$'s information is necessary due to safety concerns. The information $\Delta d_{i,i-1}^t$ and $\Delta v_{i,i-1}^t$ from HDV $i-1$ should always be incorporated into the fused DRL state to enhance safety. In particular, it regulates the local equilibrium spacing deviation $\Delta d_{i,i-1}^t \to 0$ and relative speed $\Delta v_{i,i-1}^t \to 0$, which significantly lowers the driving risks as manifested by safety surrogate measures such as time-to-collision (TTC) (Jiménez et al., 2013). On the other hand, the information from CAV $i-m$ is the key part of the state fusion process, which anticipates the relatively smooth driving behaviors from CAV$i-m$ and alleviates the HDVs' stochasticity to facilitate control performances.

For a better understanding, the above-proposed estimation method and state fusion process can be interpreted in this way. The controlled CAV i adapts its car-following strategy according to the state of its actual immediate preceding vehicle HDV $i-1$ and the state of the fictive vehicle (i.e., AHDV) in front of the controlled CAV. To refine the effective position of the fictive vehicle, the Newell-based methodology is introduced to estimate the traffic state and propagate the current state of the leader CAV $i-m$ through a set of HDV vehicles. Specifically, the time-varying terms $T_{i-1,i-m}^{*t}$ and $L_{i-1,i-m}^{t}$, which are real-time estimated from Newell's model, determine the position of the fictive vehicle. Adopting the interpretation makes the introduced method similar to Multi-Anticipative ACC car-following rules (Lin et al., 2012; Wang et al., 2014a, 2014b). Rather than defining a fictive vehicle like in MA-ACC rules only based on perception and communication sensors, which provide information regarding the immediate leader HDV $i-1$ and the CAV leader CAV $i-m$, the introduced method refines the approach by making use of the Newell's car-following model to capture the macroscopic traffic features between the two leaders.

From the equilibrium concept perspective, the control design maintains the CAV in the pre-defined equilibrium state, considering its preceding HDV and the nearest downstream CAV. The equilibrium state with the preceding HDV (i.e., $d_{i,i-1}^{*t}$, v_{i-1}^{t}) ensures local stability (Willems and Polderman, 2013), representing the capability to remain in a car-following pair of equilibrium under disturbances. The equilibrium state with the nearest downstream CAV (i.e., $d_{i,i-m}^{*t}, v_{i-m}^{t}$) incorporates the relatively stable driving motion of the downstream CAV and the macroscopic traffic features of aggregated HDVs, which further enhances the car-following performances. Moreover, this control design is generic since it is suitable for diversified compositions of the mixed local downstream environment.

Extension to the full CAV environment

It should be noted that the local communication range can be a pure connected automated environment, in which consecutive downstream vehicles are CAVs (i.e., CAV-CAVs patten), as presented in Fig. 5-4. The above generic state fusion approach can also be applied to this full CAV condition, whose control design is to achieve a platoon-level consensus by fusing received information from all the aggregated CAVs (i.e., m CAVs, $1 \leq m \leq K$) as the DRL state $\boldsymbol{s_t^i}$ for control (Shi et al., 2022). The methodology and detailed results are discussed in Section 4. Specifically, the fusion process follows the same formulations from Equation (5-5) to Equation (5-7) to calculate the equilibrium speed deviation $\Delta v_{i,i-m}^{t}$ and spacing deviation $\Delta d_{i,i-m}^{t}$ between $CAV\ i$ and $CAV\ i-m$. Due to the full CAV environment, the equilibrium time gap $T_{i,i-m}^{*}$ and equilibrium standstill spacing $L_{i,i-m}$ in Equation (5-6) can be directly defined in a multi-agent version, which means $T_{i,i-m}^{*} = m\tau_{i}^{*}$; $L_{i,i-m} = ml_{i}$. Similar to Equation (5-13), the fused DRL state $\boldsymbol{s_i^t}$ = $[\Delta d_i^t\ , \Delta v_i^t\ ]$ incorporates the weighted equilibrium deviations for the m aggregated CAVs to anticipate the disturbances induced downstream and thus achieve great system-level consensus:

$$\Delta d_i^t = \frac{q_{i-1}\eta_{i,i-1}^{t}\Delta d_{i,i-1}^{t}+q_{i-2}\eta_{i,i-2}^{t}\Delta d_{i,i-2}^{t}+\cdots+q_{i-k}\eta_{i,i-m}^{t}\Delta d_{i,i-m}^{t}}{q_{i-1}\eta_{i,i-1}^{t}+q_{i-2}\eta_{i,i-2}^{t}+\cdots q_{i-m}\eta_{i,i-m}^{t}}, \tag{5-14a}$$

$$\Delta v_i^t = \frac{q_{i-1}\eta_{i,i-1}^{t}\Delta v_{i,i-1}^{t}+q_{i-2}\eta_{i,i-2}^{t}\Delta v_{i,i-2}^{t}+\cdots+q_{i-k}\eta_{i,i-m}^{t}\Delta v_{i,i-m}^{t}}{q_{i-1}\eta_{i,i-1}^{t}+w_{i-2}\eta_{i,i-2}^{t}+\cdots q_{i-k}\eta_{i,i-m}^{t}}, \tag{5-14b}$$

where the coefficient $q_{i-j} = \begin{cases} \frac{1}{2^j},\ 1 \le j \le m-1 \\ \frac{1}{2^{j-1}},\ j = m \end{cases}$ represents that the closer CAV is assigned with greater power on the control decision.

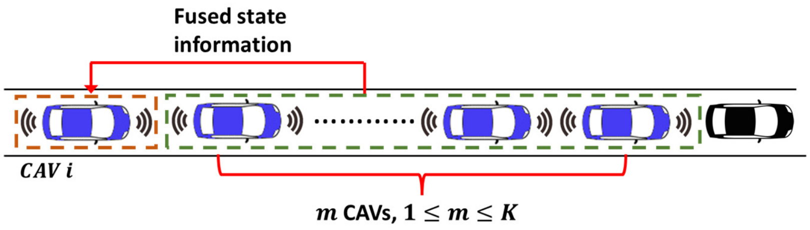


Fig. 5-4. Scenario extended to the full CAV downstream environment

5.2 Development of DRL-based Controller

Based on the defined control scheme in Section 5.1, this section develops the DRL-based controller. We discuss the detailed DRL scheme (Section 5.2.1), the adopted DRL algorithm (Section 5.2.2), and the training process (Section 5.2.3). The simulation experiments, including training and evaluation, are performed via Python. TensorFlow package is used to build the DRL algorithm. Pyomo package is applied to develop the MPC-based controller for performance comparison in the experimental part (Section 5.3).

5.2.1 DRL Scheme and Formulation

The basis of DRL is Markov Decision Process (MDP), in which the DRL agent (i.e., CAV controller) and environment (given in Section 5.1) interact with each other based on four basic elements: state, action, policy, and reward (***s, a, π, r***). As discussed in the previous section, state ***s*** represents the fused state information, which contains the weighted deviations of spacing Δd_i^t and speed Δv_i^t, denoted as $\boldsymbol{s_i^t}$ = $[\Delta d_i^t, \Delta v_i^t]$. The DRL agent receives $\boldsymbol{s_i^t}$ at each timestep and outputs the action ***a*** (i.e., the control signal u_i^t) based on the policy $\boldsymbol{\pi}$ to regulate CAV i's longitudinal movements. As an implicit function that assigns the action probability for each state, policy $\boldsymbol{\pi}(\boldsymbol{a}|\boldsymbol{s})$ needs to be updated to achieve optimal control performance through the training process.

The reward ***r*** determines the control objectives. In our design, the objectives of the car following control efficiency, which aims to maintain CAV in the pre-defined equilibrium state, and driving comfort, which pursues a smoother driving behavior with greater eco-driving performance, are incorporated in the DRL framework. In particular, the cost of the car following control efficiency c_i^t is defined as the quadratic form of deviation from the equilibrium state, which is a common control design in modern control theories such as Linear Quadratic Regulator (LQR) and MPC. This design facilitates stability analysis, as manifested by numerous control papers (e.g., Fisher & Bhattacharya, 2009; Zhou et al., 2019b). Specifically, the quadratic cost c_i^t is defined as:

$$c_i^t = (\boldsymbol{s_i^t})^T Q_i \boldsymbol{s_i^t}, \tag{5-15}$$

where $Q_i = \begin{bmatrix} \alpha_{1,i} & \\ & \alpha_{2,i} \end{bmatrix}$ is a positive definite diagonal coefficient matrix with weights $\alpha_{1,i} > 0$ and $\alpha_{2,i} > 0$.

Further, the driving comfort cost g_i^t suggested by (Wang et al., 2014) is defined in Equation (5-16), which alleviates the acceleration to improve the eco-driving performance and empirical string stability (i.e., acceleration energy) (Shi et al., 2021; Feng et al., 2019):

$$g_i^t = \alpha_{3,i}\left(a_i^t\right)^2, \tag{5-16}$$

where $\alpha_{3,i}$ denotes the weight for driving comfort. It should be noted that both the two costs regarding the car following control efficiency cost c_i^t and driving comfort cost g_i^t are unitless. The weight's unit is the reciprocal of its valuable unit. Thus, each weight of $\alpha_{1,i}$, $\alpha_{2,i}$, and $\alpha_{3,i}$ offsets the units of its variables, making the whole cost unitless.

Combining the two control objectives above, the running cost e_i^t of CAV i at timestep t is defined as the sum of the car following efficiency cost c_i^t and driving comfort cost g_i^t:

$$e_i^t = c_i^t + g_i^t. \tag{5-17}$$

Since the quadratic running cost e_i^t is similar to the cost function in the constrained optimization framework, and the training environment is similar to the state space as Zhou et al. (2019b), the coefficients setting ($\alpha_{1,i}$, $\alpha_{2,i}$, $\alpha_{3,i}$) is set to be same as Zhou et al. (2019b) to improve the string stability performance further. Though it is prohibitive to conduct mathematical string stability as Zhou et al. (2019b), due to the intrinsic complexity of DRL, we envision the setting coefficients in the same fashion as is helpful to enhance the string stability by the similarities mentioned above.

Based on the above systematic cost design, we convert the running cost e_i^t as the immediate reward r_i^t using the exponential function, as shown in Equation 5-18, which calculates the reward value as feedback for the control action at each timestep. The exponential equation serves the following two purposes. First, the reward value needs to be maximized in the DRL framework, whose optimization direction is opposite to the cost function of optimal control. Using the exponential function changes the optimization direction from minimization to maximization. Second, the above exponential function plays a normalization function, normalizing the immediate reward r_i^t within the boundary [0, 1] and further enhancing the training performance.

$$r_i^t = \exp\left(-e_i^t\right). \tag{5-18}$$

The above exponential function also normalizes the immediate reward r_i^t within the boundary [0, 1] to enhance the training performance. With the reward value, an infinite-horizon optimal control problem is formulated for maximizing the discounted cumulative rewards to find the optimal control policy π^*:

$$\pi^* = \arg\max_{\pi} \sum_{m=0}^{\infty} \Upsilon^m \, r_i^{t+m}, \tag{5-19}$$

where Υ is the discount factor.

5.2.2 Policy Update Algorithm

The DRL solves the optimization problem in Equation (5-19) by continuously updating policy π in training. The choice of DRL algorithm for the CAV control is based on the following aspects: (i) the action space (discrete/continuous); and (ii) the algorithm performance. The algorithm should support continuous action space for the instance of the microscopic CAV control with great sampling efficiency and converging performances. The Distributed Proximal Policy Optimization (DPPO) algorithm (Heess et al., 2017), one of the Actor-Critic DRL algorithms, is adopted for policy updating in training. The Actor-Critic DRL algorithm combines the merits of Policy-Based RL and Value-Based RL algorithms, which performs faster than traditional RL algorithms and supports continuous action space in the training process. Based on its merits, the Actor-Critic framework is widely used in the most popular reinforcement learning algorithms, such as the A3C algorithm (Mnih et al., 2016), DDPG algorithm (Lillicrap et al., 2016), and PPO (DPPO) algorithm (Heess et al., 2017). In this study, we adopted the DPPO algorithm to update policy due to its great balance between sampling efficiency, implementation simplicity, and converging performance (Schulman et al., 2017). Compared with traditional policy gradient RL algorithms, the DPPO algorithm makes policy gradient less sensitive to a large step and improves the convergence of policy updates by clipping the divergence of the strategy update. Besides, the DPPO algorithm updates the policy of the global agent in parallel through multiple parallel agents, which further improves training efficiency.

The Distributed Proximal Policy Optimization (DPPO) algorithm (Heess et al., 2017) is adopted for policy updating due to its great balance between sampling efficiency, implementation simplicity, and converging performance. The DPPO algorithm is a typical Actor-Critic DRL algorithm, with objective $L^{CLIP}(\theta)$ updating in the actor network and critic loss $L_c(\Phi)$ updating in the critic network. The overall actor-network framework with network structures is presented in Fig. 5-5. The detailed hyperparameters settings are demonstrated in Table 5-1. The number of neurons for the actor network (200) and critic network (100) is tuned by experiences to achieve the desired performances without causing underfitting and overfitting issues. Since the actor network learns a more complex policy function that maps the DRL state to a probability distribution over all actions, thus setting with more neurons in this study (Grondman et al., 2012).

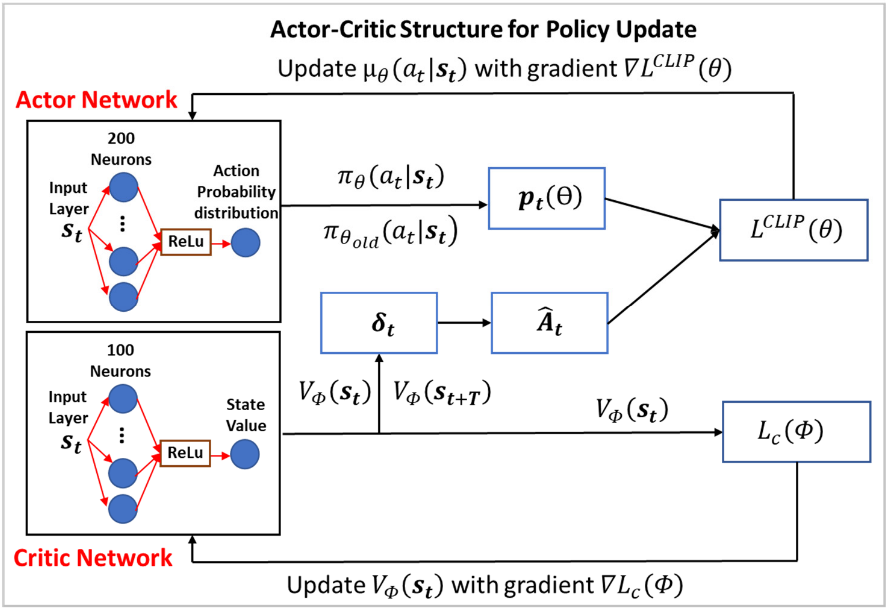

Actor-Critic Structure for Policy Update
Update $\mu_\theta(a_t|\boldsymbol{s_t})$ with gradient $\nabla L^{CLIP}(\theta)$
Actor Network
200 Neurons
Input Layer
$\boldsymbol{s_t}$
Action Probability distribution
ReLu
$\pi_\theta(a_t|\boldsymbol{s_t})$
$\pi_{\theta_{old}}(a_t|\boldsymbol{s_t})$
$\boldsymbol{p_t}(\Theta)$
$L^{CLIP}(\theta)$
$\boldsymbol{\delta_t}$
$\boldsymbol{\hat{A}_t}$
100 Neurons
Input Layer
$\boldsymbol{s_t}$
State Value
ReLu
$V_\Phi(\boldsymbol{s_t})$
$V_\Phi(\boldsymbol{s_{t+T}})$
$V_\Phi(\boldsymbol{s_t})$
$L_c(\Phi)$
Critic Network
Update $V_\Phi(\boldsymbol{s_t})$ with gradient $\nabla L_c(\Phi)$

Fig. 5-5. The actor-critic structure of the policy iteration algorithm

Table 5-1. Hyper parameters of the DPPO algorithm

Hyperparameter	**Value**
clipping value ε	0.2
minibatch T	256
discount factor Υ	0.99
Hidden layer of actor	1
Hidden layer of critic	1
actor network neurons	200
critic network neurons	100
parallel worker numbers	4
actor learning rate	0.00001
critic learning rate	0.00001

Actor Network

The actor network determines the policy π with parameter θ. It receives the DRL state $\boldsymbol{s_i^t}$ as the input and outputs a probability distribution over actions. The control signal u_i^t is then sampled from the distribution. For the network structure, there is one hidden layer with 200 neurons, and the ReLu function is adopted as the activation function for the output. The actor network is updated by maximizing the objective function $L^{CLIP}(\theta)$:

$$L^{CLIP}(\theta) = \hat{E}_t[\min(p_t(\theta)\hat{A}_t, clip(p_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t], \tag{5-20}$$

where $p_t(\theta) = \frac{\pi_\theta(a_t|\mathbf{s_t})}{\pi_{\theta_{old}}(a_t|\mathbf{s_t})}$ identifies the probability ratio of the new policy $\pi_\theta(a_t|\mathbf{s_t})$ and old policy $\pi_{old}(a_t|\mathbf{s_t})$. The clipping function $clip(p_t(\theta), 1-\varepsilon, 1+\varepsilon)$ function restricts $p_t(\theta)$ between $1-\varepsilon$ and $1+\varepsilon$ to limit the update range of new policy, making the policy gradient less sensitive to the step size and improving the convergence. Ɛ is the clipping parameter. $\hat{A}_t$ is the estimated advantage at state s_i^t, which is provided from the critic network:

$$\hat{A}_t = R_t - V_\Phi(\boldsymbol{s}_i^t), \tag{5-21}$$

where $V_\Phi(\boldsymbol{s}_i^t)$ is the value estimated from the critic network; R_t denotes the discounted sum of rewards in T steps at state s_i^t:

$$R_t = \sum_{m=0}^{T-1} \gamma^m r_i^{t+m} + \gamma^T V_\Phi(\boldsymbol{s}_i^{t+T}), \tag{5-22}$$

where T is the minibatch size; r_i^{t+m} is the immediate reward defined in Equation (5-18); γ is the discount factor. Therefore, the parameter θ of the actor network is updated based on the gradient of $L^{CLIP}(\theta)$ with learning rate α_θ:

$$\theta = \theta - \alpha_\theta \nabla L^{CLIP}(\theta). \tag{5-23}$$

Critic Network

On the other hand, the critic network with parameter Φ evaluates the decision u_i^t output by the actor network. The critic network receives the DRL state $\boldsymbol{s}_i^t$ as the input and outputs the estimated state value $V_\Phi(\boldsymbol{s}_i^t)$. For the network structure, there is one hidden layer with 100 neurons, and the ReLu function is used as the activation function for the output. The critic network is updated by minimizing the critic loss function $L_c(\Phi)$:

$$L_c(\Phi) = \hat{E}_t(V_\Phi(\boldsymbol{s}_i^t) - R_t)^2, \tag{5-24}$$

where Temporal Differences (TD) error δ_t is denoted as $\delta_t = V_\Phi(\boldsymbol{s}_i^t) - R_t$ in the loss function. The TD error δ_t estimates the advantage value $\hat{A}_t$ in actor since $\delta_t = -\hat{A}_t$. Thus, the parameter Φ is iteratively optimized based on the gradient $\nabla L_c(\Phi)$ with learning rate α_Φ: $\Phi = \Phi - \alpha_\Phi \nabla L_c(\Phi)$.

5.2.3 Procedure and Results

Based on the proposed DRL scheme (given in Section 5.2.1) and the adopted policy updating algorithm (given in Section 5.2.2), this section describes the detailed training procedure in which the DPPO agent continuously interacts with the simulation environment. The DPPO agent consists of one global agent for the actor-critic network updating and multiple parallel agents interacting with their independent simulation environments for data collecting (e.g., state, action, reward), which further improves the sampling efficiency and training speed. The detailed training process is demonstrated in Fig. 5-6. At timestep t, each parallel agent receives the CAV i's state $\boldsymbol{s}_i^t$ of its corresponding simulation environment and outputs the control signal u_i^t to update the longitudinal movement based on the current policy $\pi(a|s)$. Concurrently, the collecting data, including the calculated reward r_i^t, state $\boldsymbol{s}_i^t$, and action u_i^t, is sent to the memory buffer for storage. The update of the policy and actor-critic network is triggered after a certain batch of data is stored in the memory buffer.

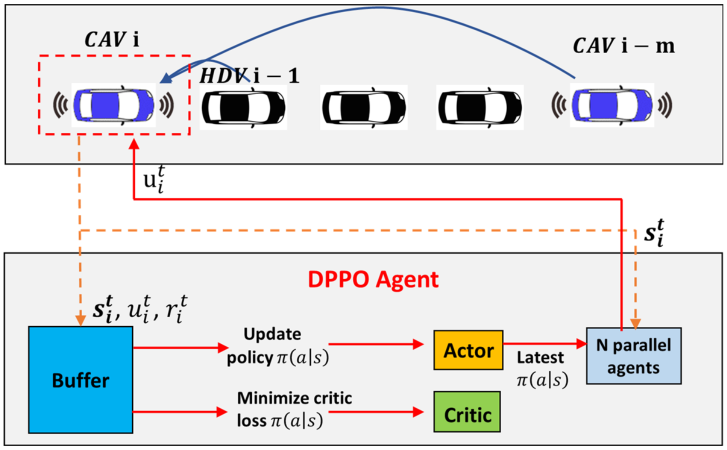


Fig. 5-6. The schematic diagram of training framework

Regarding the training environment settings, ten sets of five-vehicle ground-truth vehicular platoon trajectories with a time length of 334 timesteps from NGSIM datasets are embedded for the initial configuration. For each training episode, one of the ten platoon trajectories is randomly sampled and assigned for the trajectories of the leading vehicle CAV $i-m$ and the following aggregated HDVs. Fig. 5-7 below shows details of one platoon trajectory. The platoon trajectories that incorporate typical phases of acceleration, deceleration, and uniform speed are embedded in the training process for the DRL control model to better capture stochastic driving characteristics. It should be noted that the proposed "CAV-AHDV-CAV" structure is a generic unit in the mixed traffic flow, whose leading CAV's driving behavior could be varied and even stochastic due to the impact of the downstream traffic.

Thus, rather than setting a deterministic leading CAV, using NGSM data to represent the stochastic leading vehicle behavior should be rational.

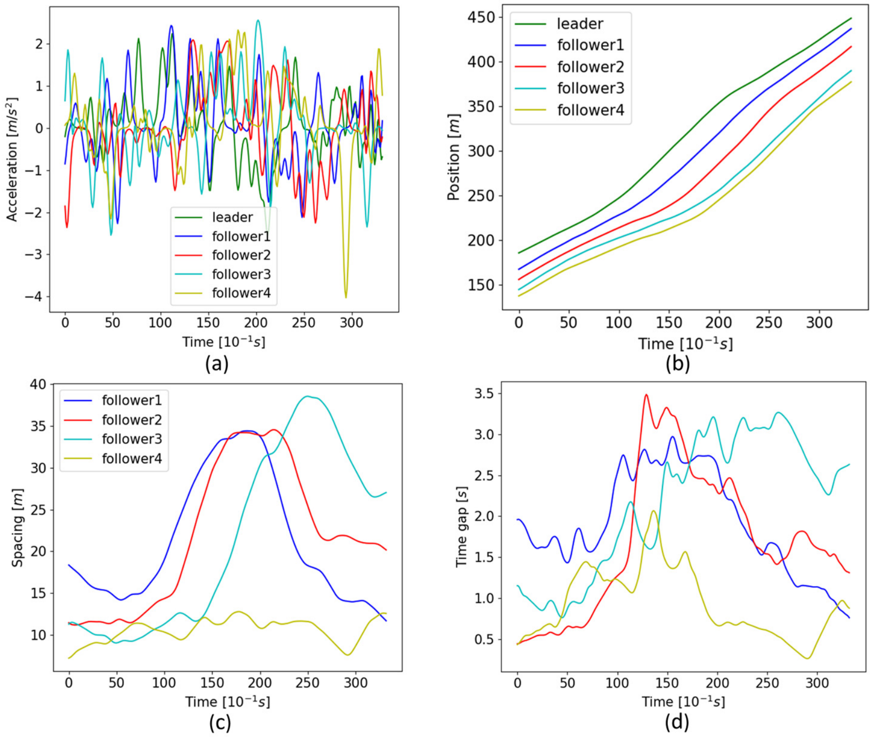
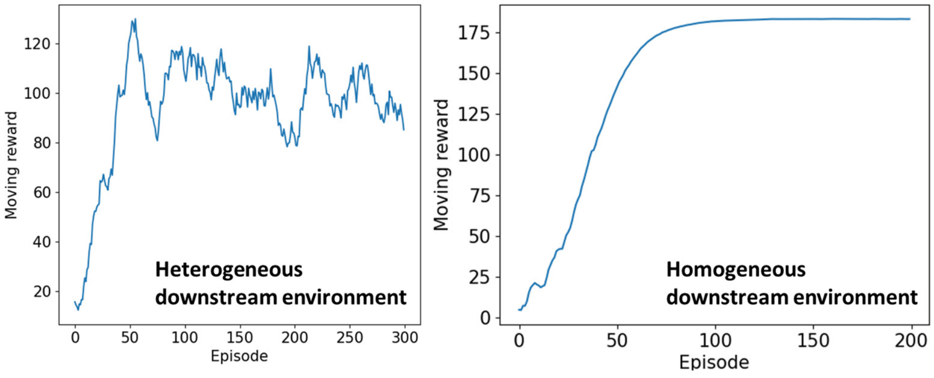


Fig. 5-7. Trajectories of the vehicular platoon in the training process regarding (a) acceleration; (b) position; (c) spacing; and (d) time gap ($\boldsymbol{g_i^t / v_i^t}$)

Specifically, the number of HDVs n $(1 \le n \le K - 1)$ between the two CAVs is randomly sampled in the simulation environment at each training episode to enhance the generalized capability for different topologies. Based on sampled topology, the trajectories of the leading CAV $i - m$ and following HDVs are assigned from the above NGSIM platoon data. Taking an example of the topology 'CAV-three HDVs-CAV', the NGSIM leader trajectory and three follower trajectories are assigned to CAV $i - m$ and three HDVs. For the controlled CAV i, it starts with the initial equilibrium state defined in Section 5.1 and is then controlled by the DRL learning agent without the loss of generality.

For the scenario where the local downstream environment is pure connected and automated (i.e., full CAVs), only the trajectory of the leading vehicle is from the NGSIM dataset, and the corresponding DRL-based models control the other downstream CAVs.

Training results are represented by the moving reward trajectory (Qu et al., 2020) demonstrated in Fig. 5-8. For the mixed (heterogeneous) traffic environment (Fig. 5-8(a)), the training platoon trajectories and number of HDVs are randomly sampled, which leads to a varied mixed environment. Thus, the rewards fluctuate in the converging area. On the other hand, the full CAV downstream environment leads to more stable reward trajectories, as presented in Fig. 5-8(b). In general, the rewards for both cases monotonically increase until convergence, indicating good converging performance.



(a) (b)

Fig. 5-8. The moving reward trajectory results for mixed traffic environment (a) and homogeneous traffic environment (b)

5.3 Simulation Experiments

5.3.1 Experiment Settings

After developing the DRL-based control models, we conduct several numerical experiments to evaluate the control approach using NGSIM datasets of I-80 in California. To remove the noises and handle the missing data, we reconstruct the datasets using a low-pass filter proposed by (Punzo et al., 2011) and

(Montanino & Punzo, 2015). The trajectories in Lane 2 from 4:00 pm to 4:15 pm are selected for experiments and analysis due to the frequent traffic congestions and oscillations and the fewer lane-changing maneuvers. The default experimental setting is shown in Table 5-2.

Table 5-2 Default parameters for the experimental setting

Parameters	**Value**
number of local downstream vehicles K	5
update interval Δt	0.1 s
vehicle length l_v	4.6 m
standstill spacing l_i	6.4 m
constant time gap τ_i^*	1 s
SINR threshold β	0.01
control demand ratio $K_{i,L}$	1
actuation time lag $I_{i,L}$	0.1
noise term $N(\mu, \sigma^2)$	$N(0, 0.1)$
$\alpha_{1,i}$, $\alpha_{2,i}$, $\alpha_{3,i}$	$1(\frac{1}{m^2})$, $0.5(\frac{s^2}{m^2})$, $0.5(\frac{s^4}{m^2})$
$[a_{i,min}, a_{i,max}]$	[-4 m/s^2 , 4 m/s^2]
free flow speed v_f	33.3 m/s (120 km/h)
wave speed w	4.4 m$/s$ (16 km/h)

The simulation experiments can be divided into three parts: (i) model performance analysis; (ii) application of the proposed model in a long vehicular platoon with different CAV penetration rates; (iii) generalization capability validation. Based on these experiments, the effectiveness, robustness, and generalization of the proposed control strategy are analyzed comprehensively. For the simulated platoons in these experiments, the leader's trajectory is picked from NGSIM datasets or the customized trajectory profile to reproduce traffic disturbances. The initial states of followers start with the predefined equilibrium states or start with the ground-truth NGSIM data. With the leader profile and followers' initial states, the vehicular platoon trajectory can be simulated based on the proposed model or other compared methods.

Evaluation Metrics

There are several performance indicators for quantitatively evaluating the control performance: cumulative dampening ratio $d_{p,i}$, local stability measured by $\Delta d_{i,i-1}^t$ and Δv_{i-1}^t, driving comfort g_i^t (given in Equation (5-16)), and average velocity $\bar{v}_i$. The cumulative dampening ratio $d_{p,i}$ quantifies the empirical string stability, an important property that measures the capability of the CAV controller in dampening traffic oscillations. The traffic oscillation magnitude is reduced or remains the same as it goes through a string stable CAV. Specifically, the l_2-norm acceleration dampening ratio $d_{p,i}$ (Ploeg et al., 2014) is specified as:

$$d_{p,i} = \frac{\|a_i^t\|_2}{\|a_0^t\|_2} = \frac{(\sum_{t=0}^N |a_i^t|^2)^{\frac{1}{2}}}{(\sum_{t=0}^N |a_0^t|^2)^{\frac{1}{2}}}, \tag{5-25}$$

where N denotes the time length; i is the vehicle index. Index 0 represents the leader of the whole vehicular platoon. The smaller dampening ratio $d_{p,i}$ indicates that the disturbances are dampened to a greater extent, leading to a more string stable driving behavior. The local stability is another important property of CAV longitudinal control, denoting a vehicle's ability to remain in the equilibrium state with its immediate preceding vehicle (Willems and Polderman, 2013). The deviations from equilibrium spacing $\Delta d_{i,i-1}^t$ and equilibrium speed Δv_{i-1}^t regarding vehicle $i-1$ are the indicators for local stability. Great local stability with low equilibrium deviations indirectly guarantees driving safety since it leads to large time-to-collision (TTC) (Minderhoud & Bovy, 2001). The average velocity $\bar{v}_i$ refers to the mean velocity of all timesteps ($\bar{v}_i = \frac{\sum_{t=0}^N v_i^t}{N}$).

HDV Modeling Method

Furthermore, the precise modeling of HDV driving behavior in the mixed traffic simulation contributes to a more realistic simulation environment and convincing results. This study uses a calibrated Intelligent Driver Model (IDM) (Kesting & Treiber, 2008; Treiber et al., 2000), which can be representative of HDV's string instability property, to model the HDV behaviors in the experiments. The IDM parameters are calibrated by (Kesting & Treiber, 2008) using ground-truth datasets of HDV behaviors. The calibrated datasets show complex situations of daily city traffic with several accelerations, decelerations, or standstill periods, which is quite similar to the adopted NGSIM datasets for experiments. Therefore, the calibrated IDM model can be applied in the experiments of this study. The calibrated parameters are presented in Table 5-3.

Table 5-3 Calibrated Parameters of IDM

Variable	Parameter	Values
V_0	Desired velocity	33.3 m/s
T	Safe time headway	1.12s
a	Maximum acceleration	1.23 m/s^2
b	Comfortable Deceleration	3.2 m/s^2
sigma	Acceleration exponent	4
S_0	Minimum distance	2.3 m

5.3.2 Control Performance Evaluation

For the first part of the experiments, this section analyzes the performance of the proposed distributed control strategy. The proposed distributed control performance is analyzed in the mixed local environment compared with the following state-of-art CAV controllers as comparisons:

- **Decentralized DRL-based controller.** The decentralized controller, also developed by the DRL, downgrades the CAV to the autonomous vehicle (AV) that can only receive its immediate preceding vehicle's information through onboard sensors. The decentralized DRL control model is developed using the same methodology (i.e., same reward design and training data) given in Section 5.2. The only difference lies in that the decentralized DRL state is defined as the local equilibrium deviations regarding the immediate preceding vehicle (i.e., $\boldsymbol{s_i^t} = [\Delta d_{i,i-1}^t$, $\Delta v_{i,i-1}^t])$.

- **Linear-based CACC controller.** The compared linear-based CACC controller (Zhou et al., 2019) is based on the constant time gap (CTG) policy, which has been proved to have excellent traffic oscillation dampening performances and guaranteed string stability performance.
- **MPC-based CACC controller.** The compared MPC-based CACC controller (Wang et al., 2016) has explicit constraints of velocity and acceleration to meet restrictions of the vehicle kinematics. The cost function is designed to achieve control efficiency and driving comfort criteria. Similarly, the CTG policy is incorporated into the control model to enhance the empirical string stability performances.

The simulated mixed platoon follows a topology $T_p = \{1', 0, 1, 0, 0, 1, 0, 0, 0, 1\}$, where $1'$ represents the leading CAV of the platoon with its trajectory from the NGSIM dataset; 0 denotes the simulated HDV follower; 1 denotes the simulated CAV follower. Each follower starts with the pre-defined equilibrium state. This topology provides the typically mixed traffic local environment, in which driving behaviors of CAV 2, CAV 5, and CAV 9 are determined by the proposed distributed control approach. The mixed platoons generated from the decentralized control and linear-based CACC strategy follow the same topology.

Fig. 5-9 presents the position, velocity, and acceleration trajectory results under DRL-based distributed control (Fig. 5-9(a)) and DRL-based decentralized control (Fig. 5-9(b)). The leading CAV's trajectory (black trajectory) shows frequent acceleration-deceleration waves and a short standstill period. For the mixed platoon under the decentralized control strategy (Fig. 5-9(b)), the HDV tends to amplify the traffic oscillations due to the long reaction time, aggravating the traffic jam. Compared with HDVs, the decentralized CAVs are more responsive to their leaders with smaller spacing, showing efficient car-following behaviors. However, the decentralized CAVs only slightly dampen the traffic disturbances since they are separated and distributed in the mixed platoon. The decentralized CAV makes the control decision only based on its preceding HDV, whose driving behavior is negatively affected by the propagated traffic disturbances. Thus, the decentralized CAV is hard to diminish the disturbances in this mixed traffic scenario.

On the other hand, the distributed CAVs, as presented in Fig. 5-9(a), also demonstrate responsive driving behaviors with smaller spacings compared with the HDVs. Moreover, the distributed CAV can dampen the traffic oscillation significantly, showing great string stability. The reason is that the downstream CAV's driving state and macroscopic traffic flow property of the aggregated HDVs are conveyed into the DRL control framework, which enhances the car following performance and better optimizes the entire mixed traffic flow. Fig. 5-9(c) gives the velocity portfolio of the linear-based CACC controller and MPC-based CACC controller. Compared with these two approaches, the distributed DRL-based control can alleviate the propagated oscillations to a more significant extent. The performance of these approaches can be differentiated around the inflection point of the acceleration-deceleration process (e.g., timestep 280, timestep 320). The underlying reason is that the DRL can better capture leading HDV characteristics and stochasticity with the proposed 'CAV-AHDV-CAV' structure and ground truth training dataset. Whereas Zhou et al. (2019a) focused on the frequency predominant range, Wang et al. (2016) focused on the formation and propagation of moving jams, which may lose some nuanced characteristics of leading HDV behaviors.

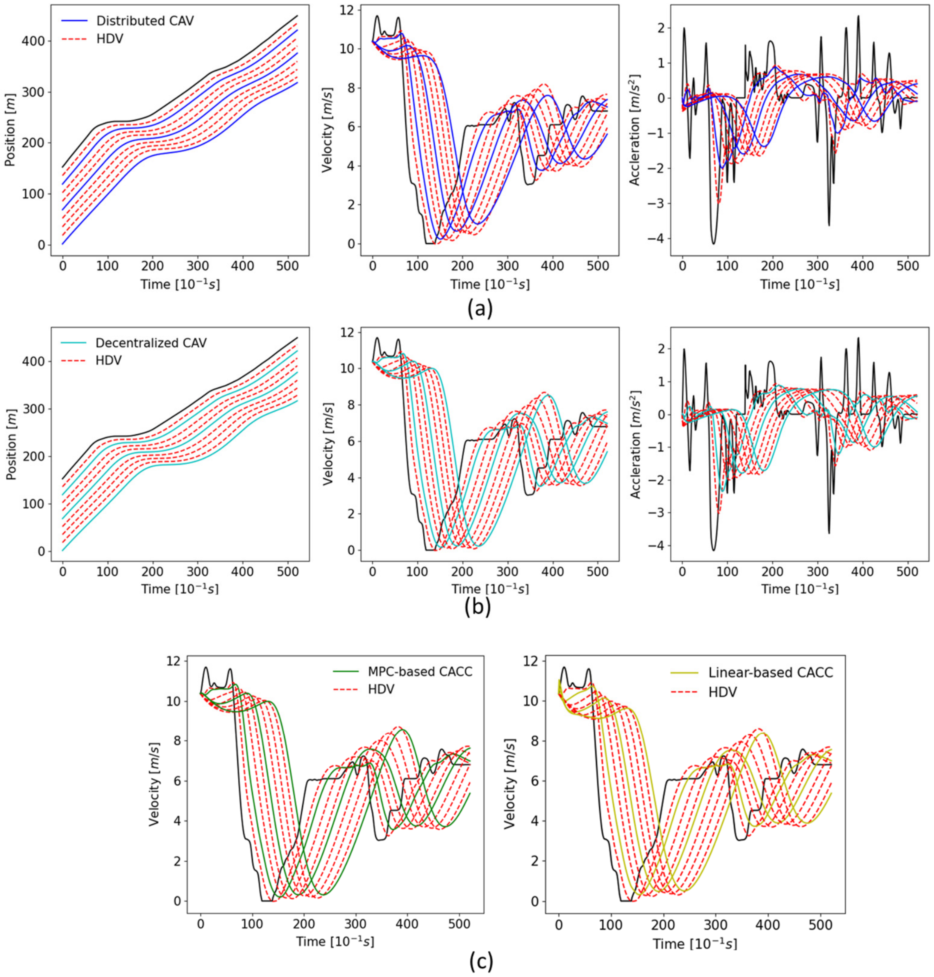


Fig. 5-9. The position, velocity and acceleration trajectory results comparison: (a) distributed control; (b) decentralized control; (c) MPC-based CACC and Linear-based CACC velocity trajectories

The quantified performance indicators of the nine followers from the distributed control, decentralized control, and CACC strategy are shown in Fig. 5-10, respectively, in which we focus on each vehicle's average performance. In general, the mixed platoon under distributed control framework greatly outperforms the decentralized control-based mixed platoon in terms of string stability and driving comfort. The performance of the linear-based CACC strategy is more akin to the proposed distributed DRL-based approach, while it scarifies certain performances in velocity. Particularly, the three distributed CAVs (CAV 2, CAV 5, and CAV 9) differentiate their performances from other strategies, in which the most upstream CAV 9 has the greatest advantage. This indicates that the more HDVs between the controlled CAV and the downstream CAV, the distributed control can better dampen traffic oscillations and have higher advantages than other strategies. Specifically, compared with the decentralized CAV 9, the distributed CAV 9 can reduce a 20.23% dampening ratio, 36.38% driving comfort cost, and increase by 0.52% average velocity. Regarding the local stability, Fig. 5-11 demonstrates the trajectories of equilibrium spacing $\Delta d_{i,i-1}^{t}$ and equilibrium speed $\Delta v_{i,i-1}^{t}$ of vehicle $i-1$. The equilibrium deviations of the CAVs are within a relatively small range (i.e., -2.5 m/s to 0.8 m/s for Δv_{i-1}^{t}; $-1.6\,m$ to $2.4\,m$ for $\Delta d_{i,i-1}^{t}$), which indicates that local stability is achieved empirically.

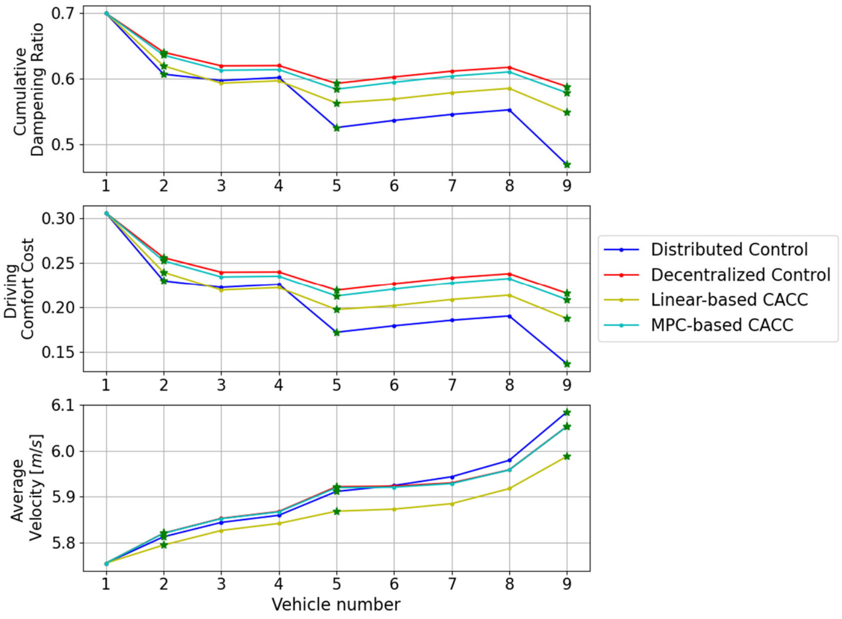


Fig. 5-10. The detailed performance indicators of the distributed control, decentralized control, linear-based CACC, and MPC-based CACC

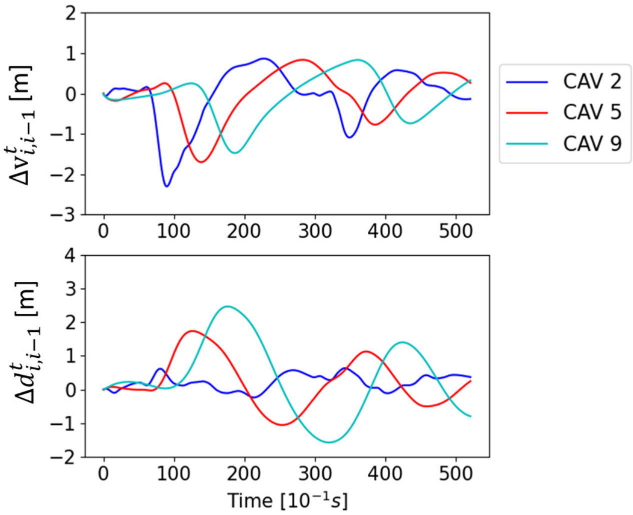
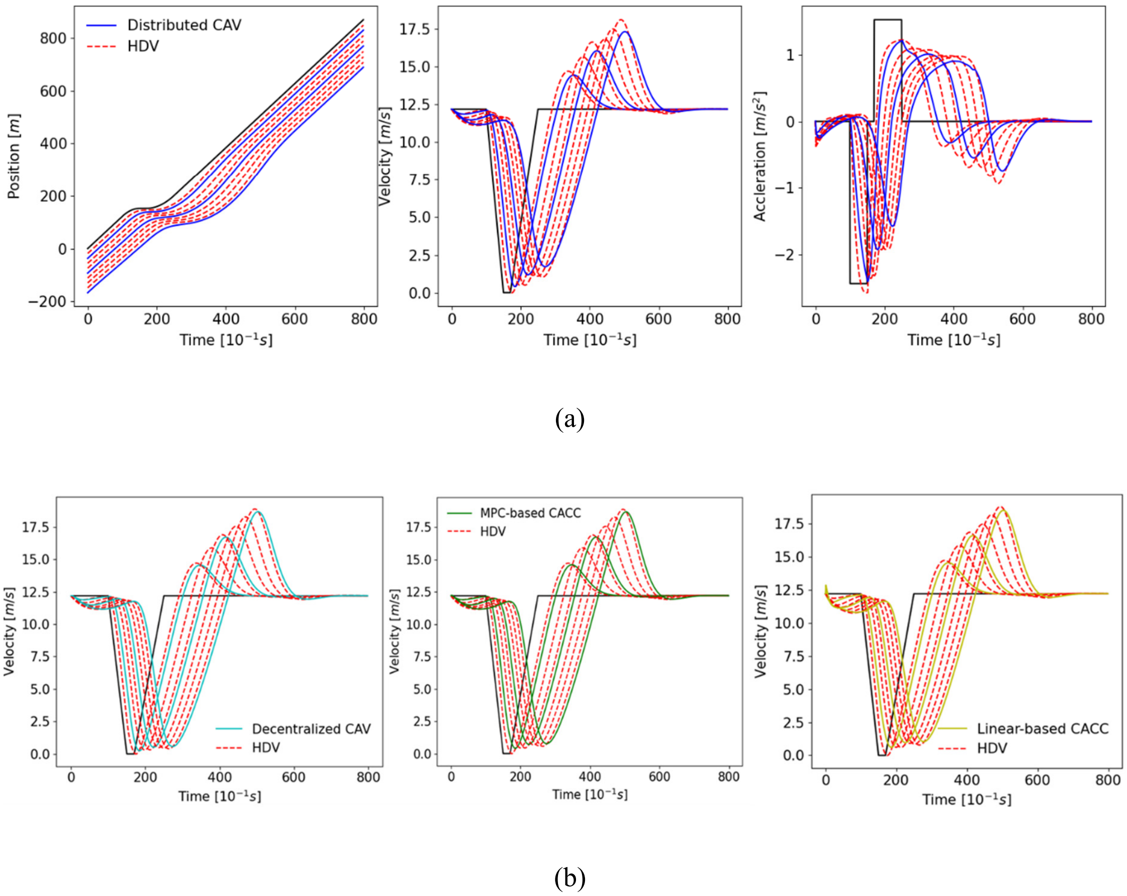


Fig. 5-11. The spacing equilibrium deviation $\boldsymbol{\Delta d_{i,i-1}^{t}}$ and speed equilibrium deviation $\boldsymbol{\Delta v_{i,i-1}^{t}}$ trajectories (equilibrium deviation with the preceding vehicle)

Evaluation under the Extreme Scenario

Furthermore, we conduct an experiment under an extreme traffic scenario to validate the robustness of the proposed controller. The vehicular platoon topology has the same topology as the previous experiment, while the leading vehicle profile is customized with one rapid deceleration-stall-acceleration cycle ($-2.4\,m/s^2 \to 0\,ft/s^2 \to 1.5\,m/s^2$) traffic oscillation. Fig. 5-12(a) gives the position, velocity, and acceleration for the proposed DRL-based controller, and Fig 5-12 (b) shows the velocity for the other three compared approaches. Similarly, the aggregated HDVs amplify traffic oscillations while the distributed CAVs greatly dampen traffic oscillations with stability-wise performances. The quantified results of indicators are presented in Fig. 5-13, which shows the proposed DRL-based control has manifest advantages over other approaches regarding string stability and driving comfort.



(a)

(b)

Fig. 5-12. The position, velocity and acceleration trajectory results under the extreme scenario: (a) Distributed DRL-based controller; (b) Velocity trajectories of decentralized CAV, MPC-based CACC, and linear-based CACC

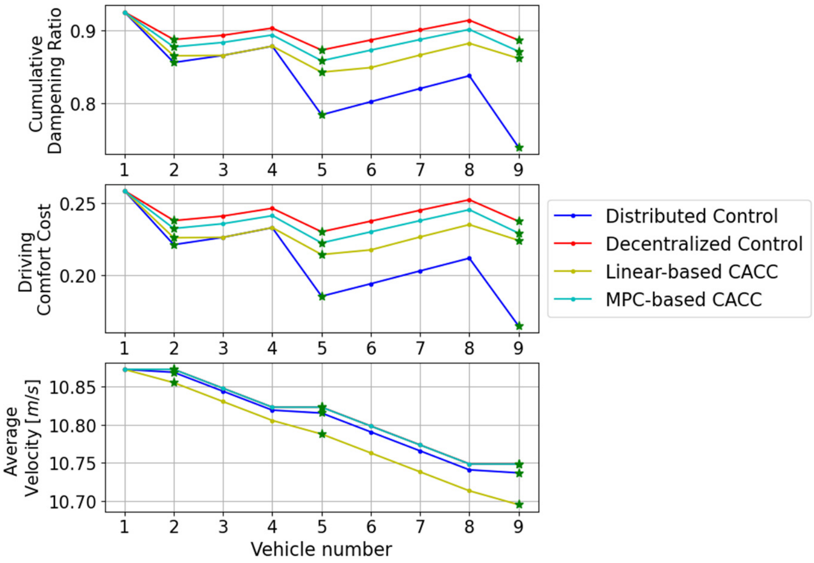
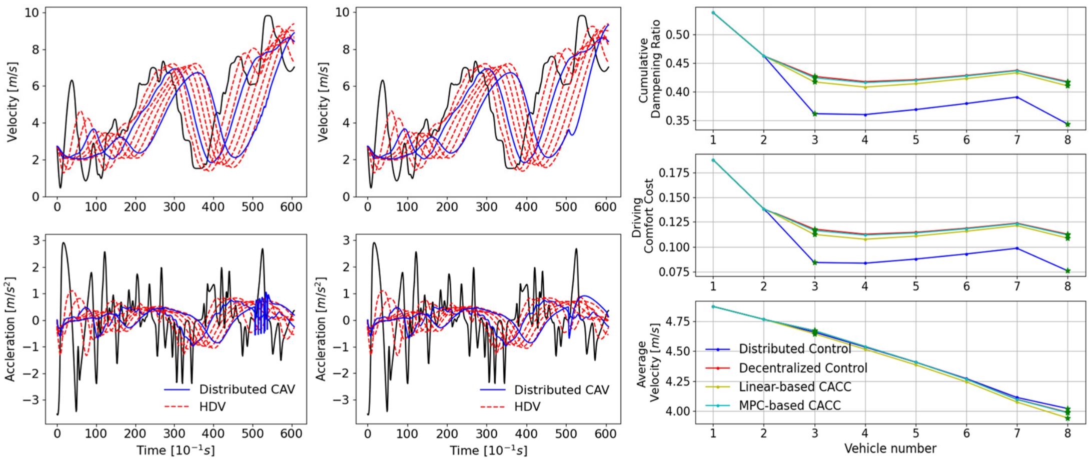


Fig. 5-13. The detailed performance indicators of the distributed control, decentralized control, linear-based CACC, and MPC-based CACC under the extreme scenarios

Evaluation under Communication Failure

Although we assume the communication quality within the communication range should be stable (i.e., communication rarely fails), the communication loss could happen at a certain period when the communication distance is relatively far (e.g., four aggregated HDVs in between CAV i and CAV $i-m$). If communication failures between CAV i and CAV $i-m$ happen, the IFT status $\eta_{i,i-m}^{t}$ (in Equation 5-14) frequently switches between 0 and 1, which makes the DRL state $\mathbf{s}_{\boldsymbol{i}}^{\boldsymbol{t}}$ fluctuate**.** This will lead to high-jerk accelerations since the DRL policy directly maps the DRL state to the control action, as presented in Fig. 5-14(a). Considering the issue, we adopted the 'dynamic information fusion mechanism' proposed by (Shi et al., 2022) to reduce the adverse impact caused by communication losses. The 'dynamic information fusion mechanism' adjusted the IFT status $\eta_{i,i-m}^{t}$ during communication failures to smooth the acceleration signal, alleviating the high-jerk DRL control issue.

The experiments are conducted to evaluate the control performances under communication failure, as presented in Fig. 5-14. The communication failure happens during 500 to 550 timesteps, where the IFT status $\eta_{i,i-m}^{t}$ between the receiver CAV 9 and transmitter CAV 5 switches frequently (Fig. 5-14(a)). With the adopted dynamic information fusion mechanism, the acceleration trajectory suddenly changes when communication failure happens and then performs smoothly without high jerks (Fig. 5-12b). The quantified results for indicators are presented in Fig. 5-14(c). Similar to the previous experiments, the proposed DRL-based control outperforms other approaches in oscillation dampening and driving comfort performances.

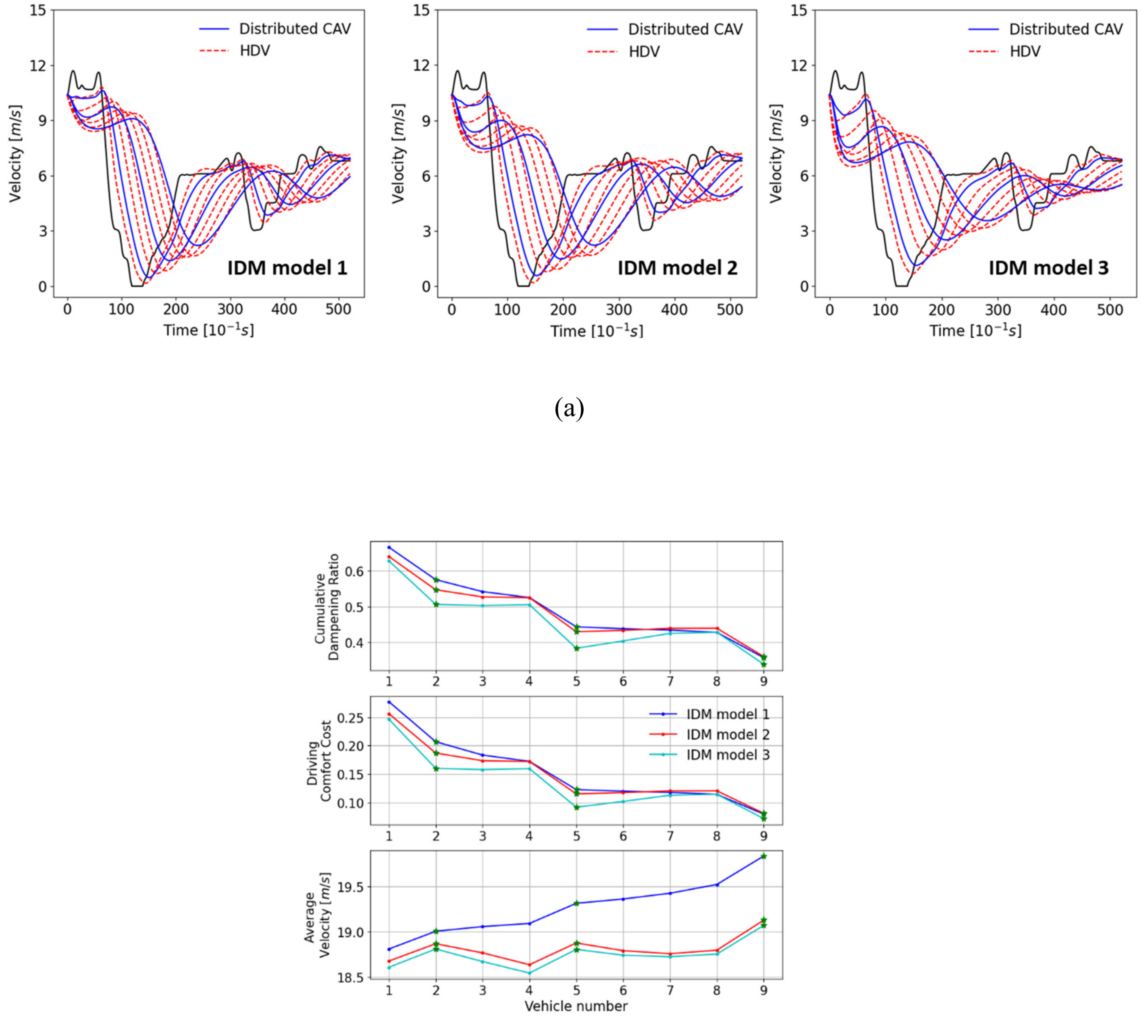


(a) (b) (c)

Fig. 5-14. The performance comparison under communication failure: (a) trajectory under communication failure (b) trajectory under communication failure with adjusted IFT by 'dynamic information fusion mechanism'; (c) performance indicator comparison under communication failure

Evaluation with different IDM model parameter settings

The adopted IDM model only reproduces one HDV driving pattern. To evaluate the CAV controller considering different HDV driving patterns, we further conducted an experiment using the IDM model with three sets of parameters calibrated based on the NGSIM datasets (Jiang et al., 2023). The trajectory and indicator results are illustrated in Fig. 5-15. The performances using different IDM models show a similar tendency of control performances, illustrating that the distributed CAV can markedly dampen traffic disturbances in the mixed traffic flow regarding different HDV driving patterns.



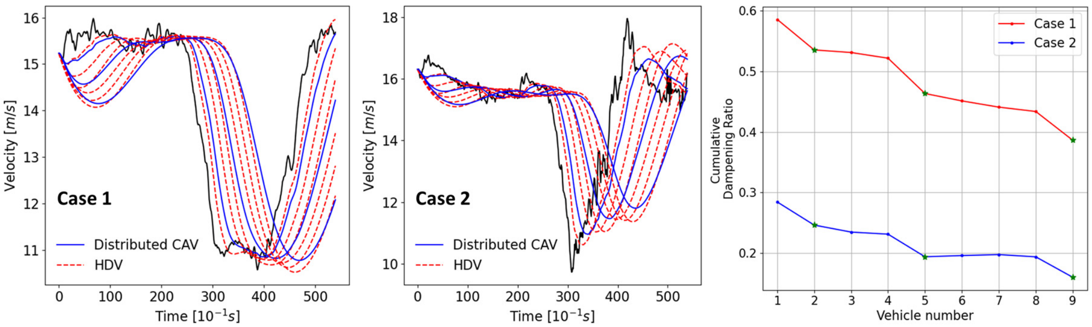
(a)

(b)

Fig. 5-15. Performance evaluation results using different calibrated IDM models: (a) trajectory results; (b) indicator results

Evaluation using ground-truth AV trajectory

The previous experiments use ground-truth NGSIM data (i.e., HDV trajectory) as the leading vehicle trajectory for evaluation. Furthermore, we evaluate our model using two ground-truth AV trajectories adopted from (Li et al., 2022) as the platoon leader trajectory, with results presented in Fig. 5-16 below. As suggested by the results, each CAV in the platoon can greatly stabilize the propagated traffic oscillations, suggesting similar control performances as the previous experiments.



(a) (b) (c)

Fig. 5-16. Two cases of AV ground-truth trajectories as the platoon leading trajectory for control performance evaluation: (a) case 1; (b) case 2; (c) indicator performance

5.3.3 Mixed Platoon with Different Penetration Rates

To further visually demonstrate the dampening performance of the proposed control strategy, we utilized the strategy to control CAVs in a 50-follower mixed platoon with different penetration rates (0%, 20%, 40%, 60%, 80%, 100%), where the CAVs are randomly sampled and distributed in the mixed traffic. The platoon leader experienced two typical deceleration-acceleration maneuvers. The followers start with the pre-defined equilibrium states. For results, Fig. 5-17 illustrates the mixed platoon's velocity and acceleration heat map, and the platoon trajectories under three typical penetration rates (0%, 60%, 100%) are demonstrated in Fig. 5-18. Generally, the traffic oscillations are dampened gradually with the increasing CAV penetration rate. When the followers are all HDVs (i.e., %0 CAV), the disturbances are propagated and amplified towards the end, seriously impairing the entire traffic flow. Furthermore, the oscillations are undermined to a large extent when the CAV penetration rate reaches 60%, where the upstream vehicles are much less affected by the dampened disturbances. Finally, the disturbances are quickly dissipated in the 100% CAV penetration rate, and the downstream CAVs can promptly recover from the disturbances, showing the controller's strong robustness and resilience. Therefore, the distributed CAV control strategy can effectively stabilize traffic oscillations and significantly improve the entire traffic flow.

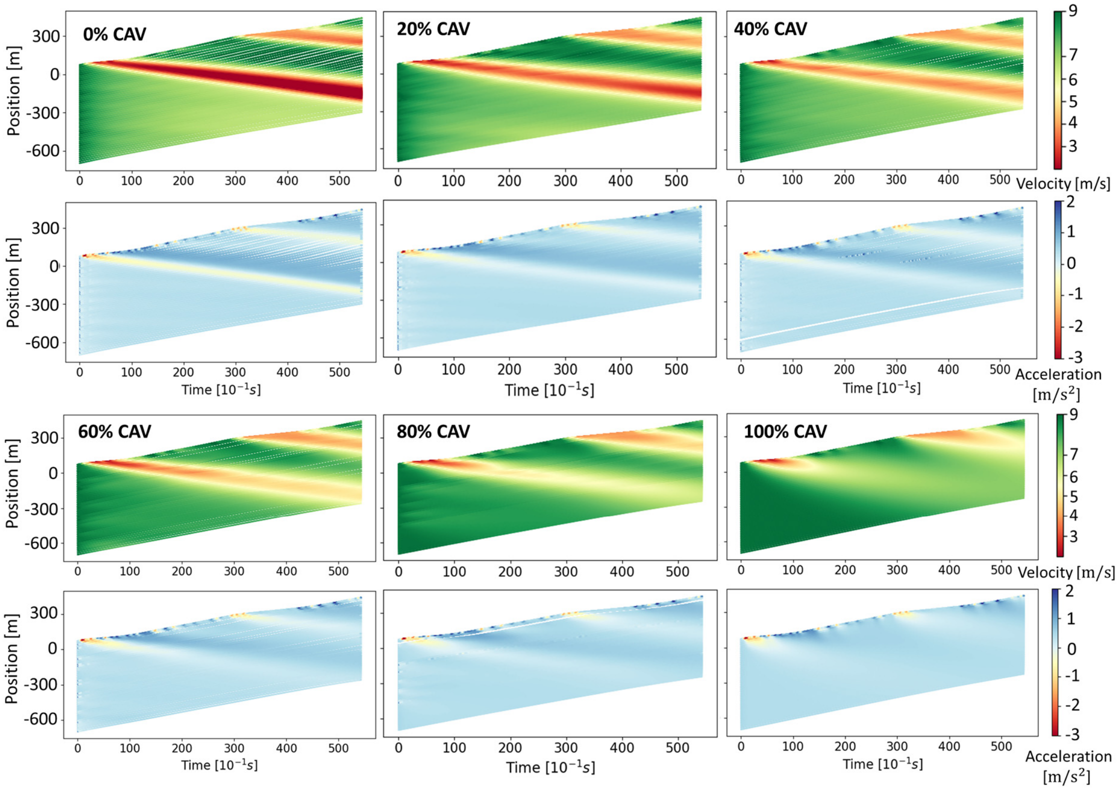
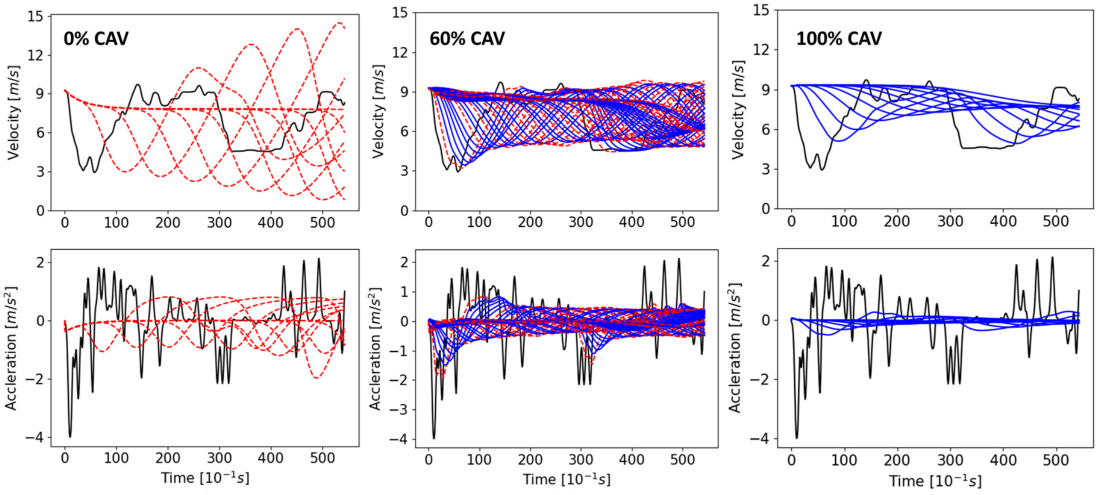


Fig. 5-17. Velocity and acceleration heat map of mixed platoon with different penetration rates



(a)

(b)

Fig. 5-18. Trajectories of velocity and acceleration under 0% (Fig. 5-18 (a)) , 60% (Fig. 5-18 (b)), and 100% (Fig. 5-18 (c)) penetration rates. (For illustration, only $\boldsymbol{vehicle\ 5, vehicle\ 10, vehicle\ 15 \dots vehicle\ 50}$ are plotted for 0% and 100% penetration rates)

5.3.4 Generalization Capability Validation

After evaluating the model performance, the generalization capability is validated in this section.

Statistical validation in mixed traffic

First, we use 150 NGSIM ground-truth trajectories excluded from training set, which is with a time length of over 50 seconds, to validate the statistical robustness of the proposed model's control performances. The experiment is configured with a 15-follower mixed platoon with different penetration rates (0%, 20%, 40%, 60%, 80%, 100%), where CAVs are randomly distributed in the mixed traffic. Each ground-truth trajectory of the 150 NGSIM datasets is assigned as the platoon leader's trajectory for each penetration rate, which means there are 150 simulated vehicular platoons for each penetration rate. The followers start with the initial equilibrium state and are then simulated by the corresponding control models (IDM for HDVs and DRL model for CAVs).

For each simulated platoon in the experiment, we first average the performance indicators of the 15 followers to represent the performance of the whole platoon. Then for each penetration rate, the mean indicator performance value over the 150 platoons is calculated, representing the generalized performance of the penetration rate. The results are demonstrated in Fig. 5-19, which illustrates that the traffic flow performance in terms of travel efficiency, string stability, and driving comfort is improved monotonically with the increasing CAV penetration rate. Specifically, compared with the HDV platoon, the platoon with a 100% CAV penetration rate reduces a 38.54% dampening ratio, 55.74% driving comfort cost, and increases 5.16% travel efficiency, respectively. These generalized results further validate the generalization capability of the proposed control strategy. To focus on the detailed performance of each vehicle in the platoon, we directly average the indicator performance for each vehicle over the 150 platoons under different CAV penetration rates, as illustrated in Fig. 5-20. As the CAV penetration rate increases, the traffic disturbances are dampened to a greater extent through the platoon, which optimizes the entire mixed traffic flow.

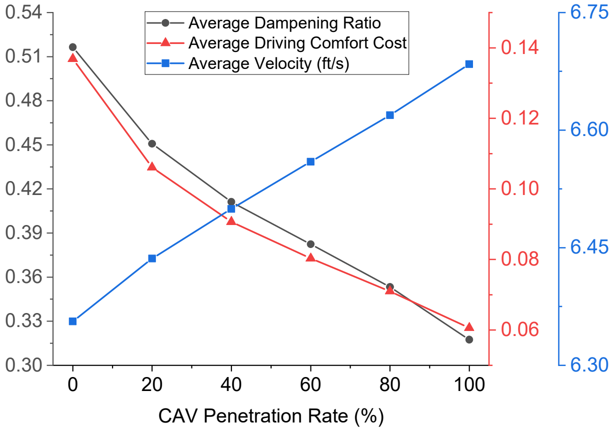


Fig. 5-19. The generalized statistical results of the mixed platoon with different penetration rates

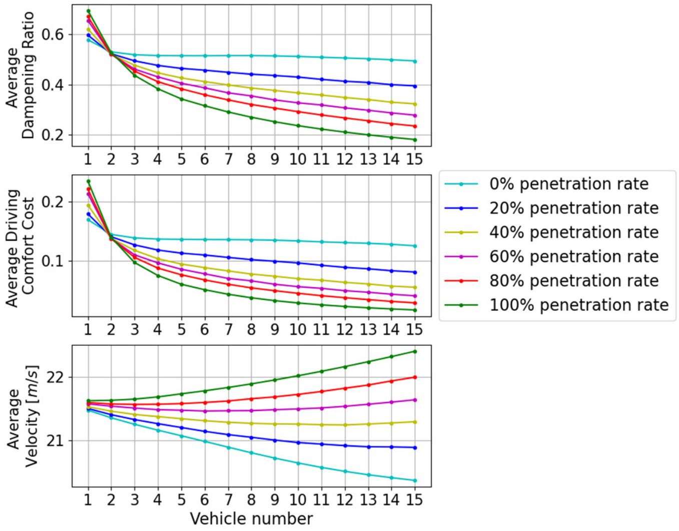


Fig. 5-20. The generalized indicator results of each vehicle in the mixed platoon under different penetration rates

Cases for irregular initial condition

The initial condition (e.g., initial velocity, acceleration, spacing) of a vehicular platoon has a great impact on the CAV controller (Li et al., 2016; Gao et al., 2019). Irregular initial conditions normally impair the control performances. To further validate the generalization capability, different NGSIM datasets are assigned for both the leader trajectory and the follower's initial states. The vehicular platoon has a topology {0, 1, 0, 0, 0, 1}, where '0' represents the HDV, and '1' represents the CAV. The results are presented in Fig. 5-21. Like the previous experiments, the DRL-based distributed CAV has responsive driving behaviors with great oscillation dampening performances even under the various initial conditions. With the equilibrium-based control philosophy, the CAV can quickly recover to the equilibrium state from the large initial spacing (Fig. 5-21 (c), (e)) or small initial spacing (Fig. 5-21 (b), (d)) and maintain close to the equilibrium, which stabilizes the traffic flow and alleviates the adverse impact brought by HDVs' stochasticity. The results validate the great robustness and resilience of the proposed controller.

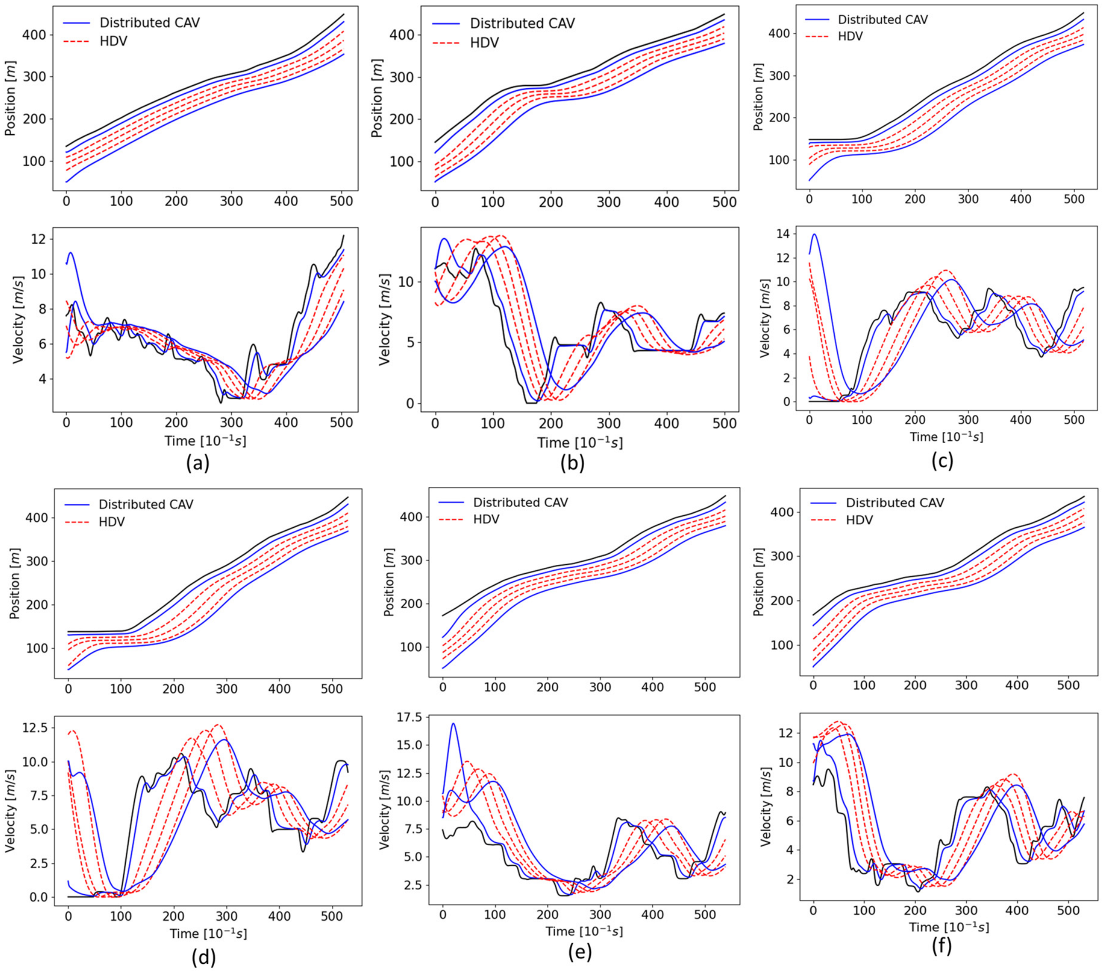


Fig. 5-21. The generalized mixed platoon trajectories with initial states from ground-truth NGSIM data

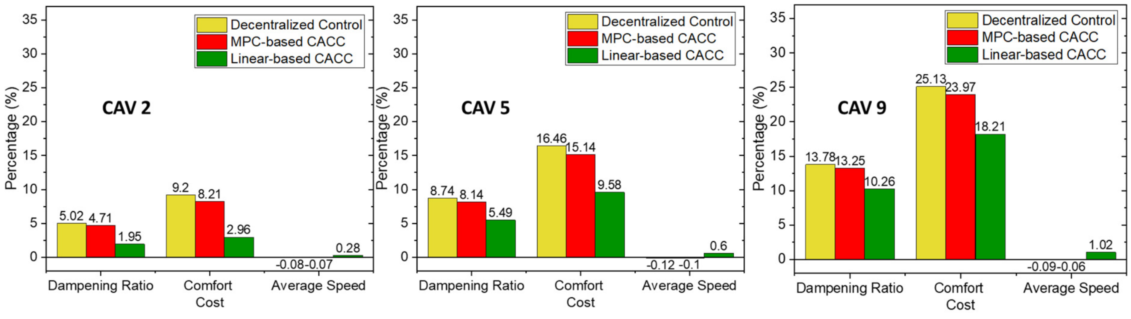
Generalization capability comparison with other approaches

Finally, multiple ground-truth trajectories from NGSIM datasets are used to further statistically validate the advantage of the proposed DRL-based controller over other compared control methods. The experiments comprise two parts, including (i) equilibrium initial condition and (ii) random initial condition. For each vehicular platoon in the experiments, the first part follows the same experiment configuration in Section 5.3.3, and the second part follows the same experiment configuration of 'Cases for Random Initial Condition' in Section 5.3.4. Specifically for each vehicular platoon, a superiority percentage P is defined to quantify the advantage of the proposed DRL-based method over other control approaches:

$$P = \frac{PI_o - PI_d}{PI_o} * 100\%, \tag{26}$$

where PI_o and PI_d represent the CAV's performance indicator value of the compared control approaches (PI_o) and the proposed DRL-based control approach (PI_d), respectively. Then, the average superiority percentage $\tilde{P}$ over multiple vehicular platoons is calculated as the final result.

For the first experiment, whose vehicular topology is $T_p = \{1, 0, 1, 0, 0, 1, 0, 0, 0, 1\}$, we focus on the performance of CAV 2, CAV 5, and CAV 9 in the platoon. The average superiority percentage $\tilde{P}$ is calculated over the 150 vehicular platoons with different NGSIM leading trajectories (i.e., same NGSIM data in Section 5.3.4). The results are presented in Fig. 5-22. As can be found in the Figure, the proposed control method markedly outperforms other control approaches regarding oscillation dampening and driving comfort. Moreover, the more HDVs between the controlled CAV and the immediate CAV downstream, the proposed DRL-based control shows higher advantages than other approaches. With more HDVs and a longer distance between the two CAVs, the proposed control can better capture the stochastic characteristics of the aggregated HDVs joint driving behaviors and stabilize traffic disturbances to a greater extent.



(a) (b) (c)

Fig. 5-22. Generalization capability comparisons for different CAVs in the mixed platoon (CAV 2 with one downstream HDVs (a); CAV 5 with two downstream HDVs (b); CAV 9 with three downstream CAVs (c); 0 line represents the performance of the proposed approach)

For the second experiment, whose vehicular topology is $T_p = \{0, 1, 0, 0, 0, 1\}$, we focus on the performance of CAV 5, which is the last CAV in the platoon. The average superiority percentage $\tilde{P}$ is calculated over thirty vehicular platoons with different NGSIM leading trajectories over 500 timesteps and irregular initial conditions for followers. The average superiority percentage $\tilde{P}$ is shown in Fig. 5-23. The MPC-based controller does not perform well in this case since the optimized control policy is more sensitive to the initial state. A large initial spacing may lead to a relatively aggressive control policy, which increases acceleration energy. Similarly, the proposed control method performs better in every aspect than other compared approaches for the cases with irregular initialized conditions.

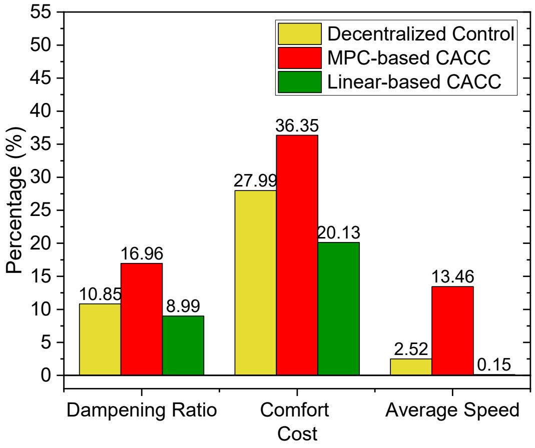


Fig. 5-23. Generalization capability comparisons with irregular initial states (i.e., CAV 5 in the mixed platoon; 0 line represents the performance of the proposed approach)

5.4 Comparison of Mixed Platoon with Different Combinations

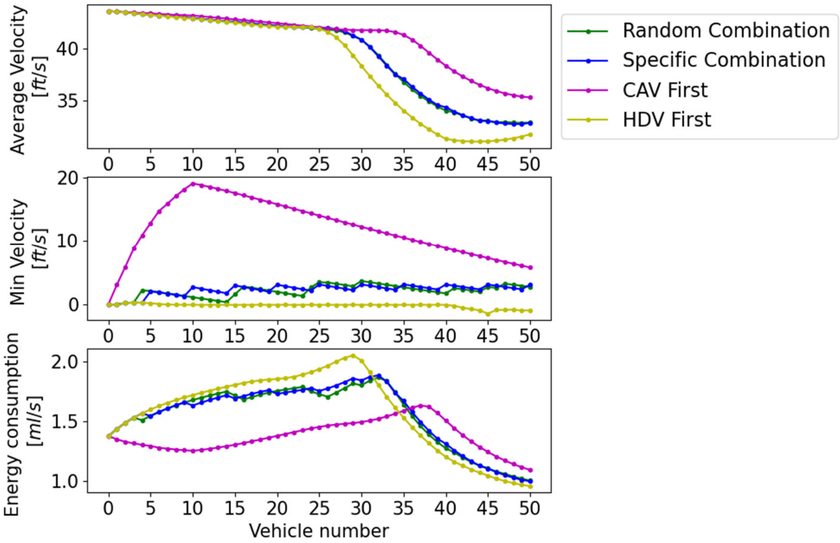
Although the proposed control strategy's performance has been validated in the mixed connected automated traffic environment, HDVs and CAVs are only combined randomly in previous experiments.

However, the combination of CAVs and HDVs can make a significant impact on traffic flow. To evaluate the impact of combinations on mixed traffic flow, four combinations of followers (random combination, specific combination, CAV first, HDV first) are analyzed and compared. The ‘Random Combination’ means CAVs are randomly sampled with a certain penetration rate in the platoon. The ‘Specific Combination’ means CAVs are evenly distributed in the platoon. 'CAV First' combination and 'HDV First' combination mean all CAVs in front of HDVs and all HDVs in front of CAVs, respectively. Platoon T_p (except for the leader) with a specific combination for different penetration rates η is defined as follows:

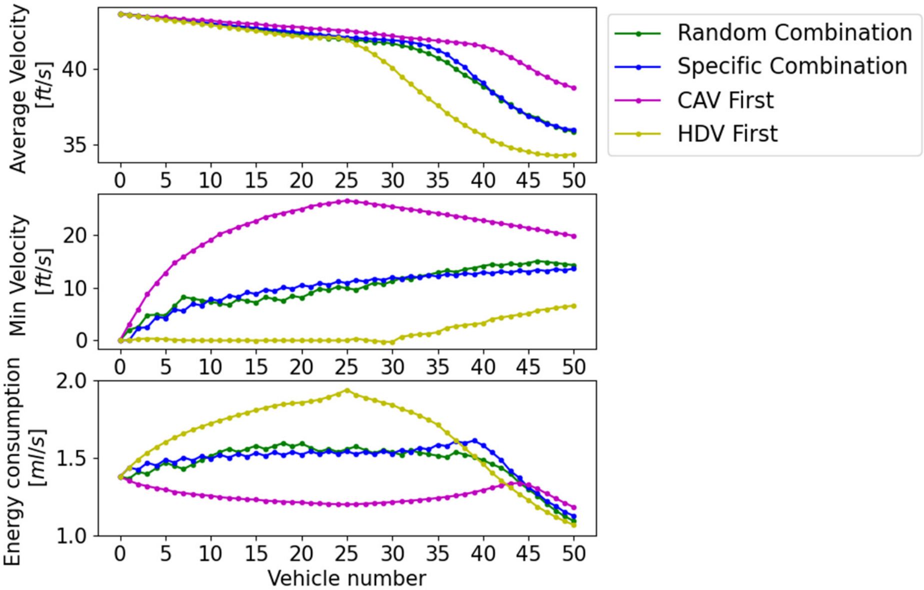
$$\overrightarrow{T_p} = \begin{cases} (\,1,1,1,1,0,1,1,1,1,0,\dots\dots,1,1,1,1,0\,), when\ \eta = \ 20\% \\ (1,0,1,0,\dots\dots,1,0), when\ \eta = 50\% \\ (1,0,0,0,0,1,0,0,0,0,\dots\dots,1,0,0,0,0), when\ \eta = 80\% \end{cases} \tag{28}$$

where η represents penetration rate; 1 represents HDV; 0 denotes CAV. In this experiment, we mainly focused on a non-cyclic traffic oscillation. The experiment is conducted with a leading vehicle trajectory incorporating one deceleration-acceleration cycle (-8 ft/s^2 - 6 ft/s^2) disturbance with a short period of standstill (2 seconds). The number of followers is set to 50 to enhance the diversity of the combinations.

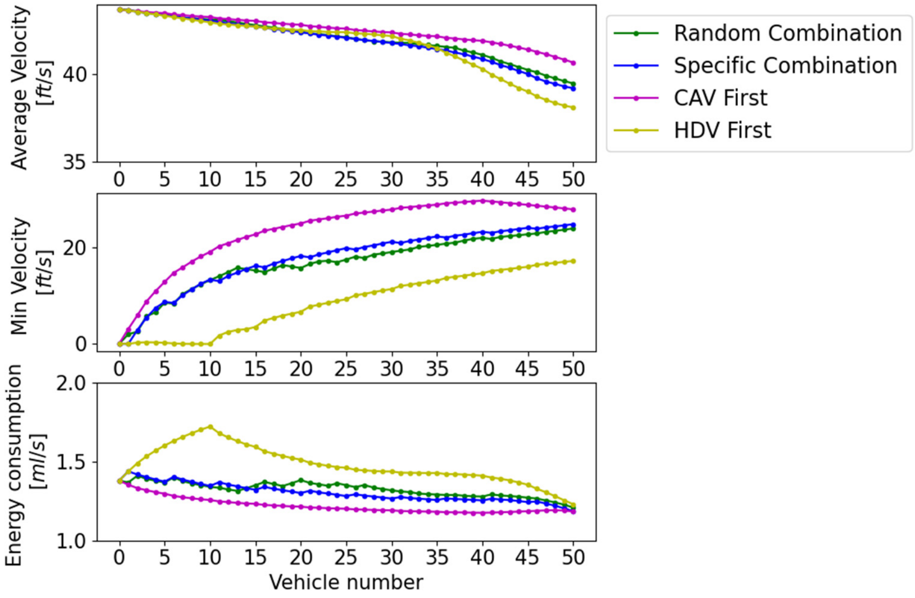
Fig. 5-24 demonstrates the detailed results. It is cleared that the platoon with all CAVs in front of HDVs outperforms the platoon with other combinations in all aspects (travel efficiency, string stability, energy efficiency) no matter what penetration rate. In contrast, platoon with all HDVs preceding CAVs takes the worst case. Results of "specific combination" and "random combination" show similar performance because HDVs and CAVs are scattered in the mixed platoon for both combinations. The results suggest that clustering leading CAVs can better optimize the entire mixed platoon's traffic flow because oscillations are dampened before they reach HDV followers. Thus, the behaviors of HDVs are optimized to the greatest extent, which mitigates the negative impact of oscillations. In contrast, if HDVs are in front of CAVs, oscillations from the platoon leader are amplified towards upstream, which makes it harder for CAVs to dampen them. Thus, the "HDV first" combination takes the worst case.



(a) 20% Penetration Rate



(b) 50% Penetration Rate



(c) 80% Penetration Rate

Fig. 5-24. Comparison of detailed indicators of each vehicle in mixed platoon with different combinations

The entire platoon's average performance under different combinations further validates the conclusion, shown in Fig. 5-25. To generalize the results of “random combination,” an average of 25 experiments was adopted for analysis. In addition, the Wilcoxon signed-rank test was conducted for each performance indicator to denote the significant level, with Table 5-4 below showing the p-value of the comparison between “random combination” and “specific combination.” As shown in Fig. 5-25, there is a great difference between the two extreme combinations ("CAV First" and "HDV First"), while "specific combination" and "random combination" show similar performance. From the statistics point of view, most of the p-value in Table 5-4 is greater than 0.05, demonstrating that the performance difference between “random combination” and “specific combination” is not significant. By contrast, the p-value of other comparisons is less than E-08, validating the significant difference. Particularly, the "CAV First" combination improves 5.77% and 11.91% in travel efficiency and energy efficiency compared with the 'HDV First' combination when the CAV penetration rate reaches 50%. Thus, CAVs can be guided to lead the mixed platoon with lane-changing maneuvers, which optimizes traffic flow to the greatest extent.

Table 5-4 P-value of Wilcoxon signed-rank test between “random combination” and “specific combination”

PI \ Penetration Rate	20%	32%	50%	66%	80%
Average Velocity	0.076	0.467	0.467	2.429E-08	2.429E-08
Minimum Velocity	0.090	1.333	0.290	0.0290	0.225
Energy Consumption	0.090	0.029	3.738E-05	0.0896	0.225

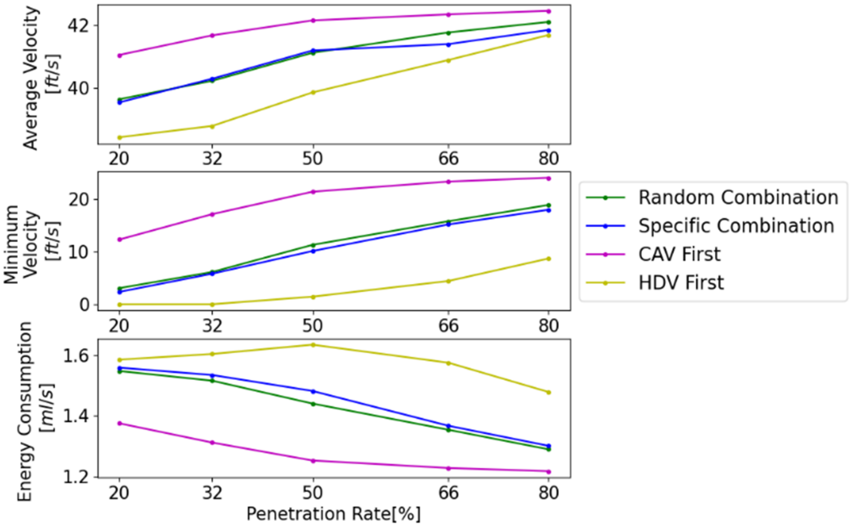


Fig. 5-25. Comparison of detailed results of mixed platoon with different combinations

5.5 Conclusion

This research proposes a DRL-based distributed CAV longitudinal control strategy for mixed traffic of CAVs and HDVs. In this generic distributed control framework, each CAV receives the fused real-time information of vehicles in the local downstream environment for longitudinal control. To generalize the diversified downstream topologies, any mixed local downstream environment is categorized as the CAV-HDVs-CAV pattern, which consists of a nearest downstream CAV followed by aggregated HDVs. For this local heterogeneous environment, we construct a novel vehicle-following structure 'CAV-AHDV-CAV' based on Newell's car-following model to capture the macroscopic traffic properties of the aggregated HDVs and embed them into the control framework. This approach efficiently attenuates the HDVs' stochasticity and enhances the car-following performances. With the philosophy, a novel DRL state fusion strategy based on the equilibrium concept is proposed to regulate each CAV close to the pre-defined equilibrium state and greatly stabilize traffic oscillations. For model development, NGSIM datasets are embedded in training to better incorporate the preceding vehicles' stochastic characteristics into control. The DPPO algorithm is adopted to enhance the convergence of control policy updated in the training process.

A series of simulated experiments are conducted with NGSIM datasets. The proposed strategy's control performance is evaluated regarding empirical string stability, travel efficiency, and driving comfort. Numerical results indicate that the proposed distributed control strategy can significantly dampen the traffic oscillation and outperform the decentralized strategy and linear-based CACC strategy in every aspect. Then, the dampening performance of the proposed control strategy is intuitively demonstrated in a 50-follower mixed platoon with different penetration rates, showing its strong robustness and resilience. Finally, the generalization capability of the proposed strategy is validated.

This study still has several limitations. The first point lies in that the proposed control method focuses on heavily congested traffic conditions, while it is not suitable for free flow conditions. Besides, this study does not consider the communication delay, which may lead to an over-optimistic performance. Moreover, the study only considers the longitudinal car-following movement, which is relatively limited for applications in more complex scenarios (e.g., lane-changing movement). Some future work can be conducted based on the research. For instance, the vehicle dynamics can be built more complex considering the internal vehicle components (e.g., pedal, steering wheel, brake). Moreover, lateral movement can be incorporated into the control framework to reproduce more complex traffic scenarios, such as lane-changing, merging, or diverging behaviors.

6 CONCLUSIONS

6.1 Summary of Chapters

Chapter 1 introduces the background of CAV control, mixed traffic environment, and the issues of traffic oscillations. This Chapter gives the identified gaps, research objectives and scope of work, research contributions, and organization of the dissertation.

Chapter 2 reviews relative literatures of CAV control approaches, the DRL algorithms, V2V communications, and equilibrium and consensus concepts. This Chapter summarized the pros and cons of current CAV longitudinal control approaches regarding linear-based controller, MPC-based controller, and DRL-based controller.

Chapter 3 discusses the current major challenges for CAV control in the mixed traffic flow, and introduces the general philosophy of the proposed CAV control strategy in the mixed traffic. To accommodate any possible CAV-HDV platoon configuration, we categorize the local downstream environment into two broad traffic scenarios based on the composition of CAVs and HDVs, and designed a generic DRL-based control framework for the two scenarios. The details control design for the two scenarios are discussed in Chapter 4 and Chapter 5, respectively.

Chapter 4 presents a DRL-based generic distributed CAV longitudinal control approach in a relatively realistic communication environment. To better capture stochastic characteristics of the preceding vehicles and communication loss, we embed the NGSIM datasets and the SINR based dynamic communication mechanism into the training framework. Each CAV in the framework receives its downstream CAVs' fused information as the DRL state for real-time control. The fused DRL state and reward function are specially designed to incorporate the merits of the equilibrium concept and consensus concept, which maintains CAVs around the predefined equilibrium point and achieves the system-level consensus to better dampen traffic oscillations. A dynamic information fusion mechanism is proposed to smooth the fluctuated DRL state and the high-jerk control signal caused by the dynamic communication loss. The simulated experiments validated the performances of the proposed controller.

Chapter 5 proposes presents a novel distributed longitudinal control strategy for connected automated vehicles (CAVs) in mixed traffic environments with human-driven vehicles (HDVs), utilizing high-dimensional platoon information. Traditional CAV control methods, which focus on microscopic trajectory information, struggle to address HDV stochasticity and mixed traffic heterogeneities efficiently. Our approach, for the first time, treats consecutive HDVs as a single entity (AHDV) to reduce stochasticity and leverages macroscopic features for controlling following CAVs. The new strategy anticipates disturbances and traffic features in mixed traffic scenarios, significantly outperforming traditional methods. The control algorithm employs deep reinforcement learning (DRL) to enhance car-following efficiency and addresses aggregated car-following behavior stochasticity by incorporating it into the training environment. Mixed traffic platoons are categorized as CAV-HDVs-CAV patterns, with macroscopic traffic properties based on the Newell car-following model to capture aggregated HDVs' joint behaviors. Simulation experiments validate the proposed strategy, demonstrating its superior performance in oscillation dampening, eco-driving, and generalization capability.

6.2 Limitation And Future Works

Some future studies can be investigated based on current results. The CAV lateral control can be incorporated in the control framework for merging, diverging or lane-changing maneuvers. In addition, other dynamic or validated communication models (Kim et al., 2017; Wang et al., 2019) or topologies (e.g., relay communication topology, V2I, V2C) can be embedded in the framework to conduct extended experiments. The dynamic communication delay can be considered to make the control framework more realistic. Moreover, the complex mixed traffic flow properties can be further studies and optimized based on this study by extending the control framework.

Furthermore, we can incorporate the prediction process (i.e., predicting the behavior of the surrounding vehicles) into the control framework to achieve more efficient control performance. The prediction process using advanced supervised machine learning algorithms (Ahmadlou & Adeli, 2010; Alam et al., 2020; Pereira et al., 2020; Rafiei & Adeli, 2017) are considered as the extension based on the current control framework.

REFERENCE

Absil, P. A., & Kurdyka, K. (2006). On the stable equilibrium points of gradient systems. *Systems and Control Letters*, *55*(7), 573–577. https://doi.org/10.1016/j.sysconle.2006.01.002

Ahmadlou, M., & Adeli, H. (2010). Enhanced probabilistic neural network with local decision circles: A robust classifier. *Integrated Computer-Aided Engineering*, *17*(3), 197–210. https://doi.org/10.3233/ICA-2010-0345

Alam, K. M. R., Siddique, N., & Adeli, H. (2020). A dynamic ensemble learning algorithm for neural networks. *Neural Computing and Applications*, *32*(12), 8675–8690. https://doi.org/10.1007/s00521-019-04359-7

Bando, M., Hasebe, K., Nakayama, A., Shibata, A., & Sugiyama, Y. (1995). Dynamical model of traffic congestion and numerical simulation. *Physical Review E*, *51*(2), 1035–1042. https://doi.org/10.1103/PhysRevE.51.1035

Bang, S., & Ahn, S. (2019). Mixed Traffic of Connected and Autonomous Vehicles and Human-Driven Vehicles: Traffic Evolution and Control using Spring-Mass-Damper System. *Transportation Research Record*, *2673*(7), 504–515. https://doi.org/10.1177/0361198119847618

Bian, Y., Zheng, Y., Ren, W., Li, S. E., Wang, J., & Li, K. (2019). Reducing time headway for platooning of connected vehicles via V2V communication. *Transportation Research Part C: Emerging Technologies*, *102*(August 2018), 87–105. https://doi.org/10.1016/j.trc.2019.03.002

Chen, D., Laval, J., Zheng, Z., & Ahn, S. (2012). A behavioral car-following model that captures traffic oscillations. *Transportation Research Part B: Methodological*, *46*(6), 744–761. https://doi.org/10.1016/j.trb.2012.01.009

Chen, D., Srivastava, A., Ahn, S., & Li, T. (2020). Traffic dynamics under speed disturbance in mixed traffic with automated and non-automated vehicles. *Transportation Research Part C: Emerging Technologies*, *113*(April 2019), 293–313. https://doi.org/10.1016/j.trc.2019.03.017

Chen, N., Wang, M., Alkim, T., & Van Arem, B. (2018). A Robust Longitudinal Control Strategy of Platoons under Model Uncertainties and Time Delays. *Journal of Advanced Transportation*, *2018*, 15–17. https://doi.org/10.1155/2018/9852721

Chen, T., Wang, M., Gong, S., Zhou, Y., & Ran, B. (2021). *Connected and Automated Vehicle Distributed Control for On-ramp Merging Scenario: A Virtual Rotation Approach*. http://arxiv.org/abs/2103.15047

Chong, L., Abbas, M. M., Medina Flintsch, A., & Higgs, B. (2013). A rule-based neural network approach to model driver naturalistic behavior in traffic. *Transportation Research Part C: Emerging Technologies*, *32*, 207–223. https://doi.org/10.1016/j.trc.2012.09.011

Du, L., & Dao, H. (2015). Information dissemination delay in vehicle-to-vehicle communication networks in a traffic stream. *IEEE Transactions on Intelligent Transportation Systems*, *16*(1), 66–80. https://doi.org/10.1109/TITS.2014.2326331

Du, R., Chen, S., Li, Y., Dong, J., Ha, P. Y. J., & Labi, S. (2020). *A Cooperative Control Framework for CAV Lane Change in a Mixed Traffic Environment*. 0–1. http://arxiv.org/abs/2010.05439

Duan, J., Li, S. E., Guan, Y., Sun, Q., & Cheng, B. (2020). Hierarchical reinforcement learning for self-driving decision-making without reliance on labelled driving data. *IET Intelligent Transport Systems*, *14*(5), 297–305. https://doi.org/10.1049/iet-its.2019.0317

Duret, A., Ahn, S., & Buisson, C. (2011). Passing rates to measure relaxation and impact of lane-changing in congestion. *Computer-Aided Civil and Infrastructure Engineering*, *26*(4), 285–297. https://doi.org/10.1111/j.1467-8667.2010.00675.x

Feng, S., Zhang, Y., Li, S. E., Cao, Z., Liu, H. X., & Li, L. (2019). String stability for vehicular platoon control: Definitions and analysis methods. *Annual Reviews in Control*, *47*, 81–97. https://doi.org/10.1016/j.arcontrol.2019.03.001

Fisher, J., & Bhattacharya, R. (2009). Linear quadratic regulation of systems with stochastic parameter uncertainties. *Automatica*, *45*(12), 2831–2841. https://doi.org/10.1016/j.automatica.2009.10.001

Gao, S., Dong, H., Song, H., & Zhou, M. (2019). On state feedback control and Lyapunov analysis of car-following model. *Physica A: Statistical Mechanics and Its Applications*, *534*, 122320. https://doi.org/10.1016/j.physa.2019.122320

Ge, J. I., & Orosz, G. (2014a). Dynamics of connected vehicle systems with delayed acceleration feedback. *Transportation Research Part C: Emerging Technologies*, *46*, 46–64. https://doi.org/10.1016/j.trc.2014.04.014

Ge, J. I., & Orosz, G. (2014b). Optimal control of connected vehicle systems. *Proceedings of the IEEE Conference on Decision and Control*, *2015-Febru*(February), 4107–4112. https://doi.org/10.1109/CDC.2014.7040028

Gong, S., & Du, L. (2018). Cooperative platoon control for a mixed traffic flow including human drive vehicles and connected and autonomous vehicles. *Transportation Research Part B: Methodological*, *116*, 25–61. https://doi.org/10.1016/j.trb.2018.07.005

Gong, S., Shen, J., & Du, L. (2016). Constrained optimization and distributed computation based car following control of a connected and autonomous vehicle platoon. *Transportation Research Part B: Methodological*, *94*, 314–334. https://doi.org/10.1016/j.trb.2016.09.016

Görges, D. (2017). Relations between Model Predictive Control and Reinforcement Learning. *IFAC-PapersOnLine*, *50*(1), 4920–4928. https://doi.org/10.1016/j.ifacol.2017.08.747

Grondman, I., Busoniu, L., Lopes, G. A. D., & Babuška, R. (2012). A survey of actor-critic reinforcement learning: Standard and natural policy gradients. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *42*(6), 1291–1307. https://doi.org/10.1109/TSMCC.2012.2218595

Guan, Y., Ren, Y., Li, S. E., Sun, Q., Luo, L., Taguchi, K., & Li, K. (2019). *Centralized Conflict-free Cooperation for Connected and Automated Vehicles at Intersections by Proximal Policy Optimization*. 1–9. http://arxiv.org/abs/1912.08410

Guo, H., Liu, J., Dai, Q., Chen, H., Wang, Y., & Zhao, W. (2020). A Distributed Adaptive

Triple-Step Nonlinear Control for a Connected Automated Vehicle Platoon with Dynamic Uncertainty. *IEEE Internet of Things Journal*, *7*(5), 3861–3871. https://doi.org/10.1109/JIOT.2020.2973977

Heess, N., TB, D., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., Erez, T., Wang, Z., Eslami, S. M. A., Riedmiller, M., & Silver, D. (2017). *Emergence of Locomotion Behaviours in Rich Environments*. http://arxiv.org/abs/1707.02286

Introduction to Mathematical Systems Theory : A Behavioral Approach by Jan Willem Polderman ; Jan C . Willems Review by : Babatunde A . Ogunnaike Published by : American Statistical Association Stable URL : http://www.jstor.org/stable/2670190 . Your use o. (2014). *94*(446), 651–652.

Jiménez, F., Naranjo, J. E., & García, F. (2013). An Improved Method to Calculate the Time-to-Collision of Two Vehicles. *International Journal of Intelligent Transportation Systems Research*, *11*(1), 34–42. https://doi.org/10.1007/s13177-012-0054-4

Karnchanachari, N., Valls, M. I., Hoeller, D., & Hutter, M. (2020). *Practical Reinforcement Learning For MPC: Learning from sparse objectives in under an hour on a real robot*. *xxx*, 1–14. http://arxiv.org/abs/2003.03200

Kesting, A., & Treiber, M. (2008). Calibrating car-following models by using trajectory data methodological study. *Transportation Research Record*, *2088*, 148–156. https://doi.org/10.3141/2088-16

Kim, Y. H., Peeta, S., & He, X. (2017). Modeling the information flow propagation wave under vehicle-to-vehicle communications. *Transportation Research Part C: Emerging Technologies*, *85*(October), 377–395. https://doi.org/10.1016/j.trc.2017.09.023

Kwak, S. G., & Kim, J. H. (2017). *cornerstone of modern statistics*.

Laval, J. A., & Leclercq, L. (2010). A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1928), 4519–4541. https://doi.org/10.1098/rsta.2010.0138

Li, M., Li, Z., Xu, C., & Liu, T. (2020). *A Deep Reinforcement Learning-based Vehicle Driving Strategy to Reduce Crash Risks in Traffic Oscillations*.

Li, S. E., Qin, X., Zheng, Y., Wang, J., Li, K., & Zhang, H. (2019). Distributed Platoon Control under Topologies with Complex Eigenvalues: Stability Analysis and Controller Synthesis. *IEEE Transactions on Control Systems Technology*, *27*(1), 206–220. https://doi.org/10.1109/TCST.2017.2768041

Li, S., Li, K., Rajamani, R., & Wang, J. (2011). Model predictive multi-objective vehicular adaptive cruise control. *IEEE Transactions on Control Systems Technology*, *19*(3), 556–566. https://doi.org/10.1109/TCST.2010.2049203

Li, T., Chen, D., Zhou, H., Xie, Y., & Laval, J. (2022). Fundamental diagrams of commercial adaptive cruise control: Worldwide experimental evidence. *Transportation Research Part C: Emerging Technologies*, *134*(October 2021), 103458. https://doi.org/10.1016/j.trc.2021.103458

Li, X., Peng, F., & Ouyang, Y. (2010). Measurement and estimation of traffic oscillation properties. *Transportation Research Part B: Methodological*, *44*(1), 1–14. https://doi.org/10.1016/j.trb.2009.05.003

Li, Y., Li, K., Zheng, T., Hu, X., Feng, H., & Li, Y. (2016). Evaluating the performance of vehicular platoon control under different network topologies of initial states. *Physica A: Statistical Mechanics and Its Applications*, *450*, 359–368. https://doi.org/10.1016/j.physa.2016.01.006

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2016). Continuous control with deep reinforcement learning. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*.

Lin, F., Fardad, M., & Jovanović, M. R. (2012). Optimal control of vehicular formations with nearest neighbor interactions. *IEEE Transactions on Automatic Control*, *57*(9), 2203–2218. https://doi.org/10.1109/TAC.2011.2181790

Lin, Y., Wang, P., Zhou, Y., Ding, F., Wang, C., & Tan, H. (2020). Platoon Trajectories Generation: A Unidirectional Interconnected LSTM-Based Car-Following Model. *IEEE Transactions on Intelligent Transportation Systems*, 1–11. https://doi.org/10.1109/TITS.2020.3031282

Lu, C., & Liu, C. (2021). Ecological control strategy for cooperative autonomous vehicle in mixed traffic considering linear stability. *Journal of Intelligent and Connected Vehicles*, *4*(3), 115–124. https://doi.org/10.1108/jicv-08-2021-0012

Marsden, G., McDonald, M., & Brackstone, M. (2001). Towards an understanding of adaptive cruise control. *Transportation Research Part C: Emerging Technologies*, *9*(1), 33–51. https://doi.org/10.1016/S0968-090X(00)00022-X

Meng, D., Song, G., Wu, Y., Zhai, Z., Yu, L., & Zhang, J. (2021). Modification of Newell's car-following model incorporating multidimensional stochastic parameters for emission estimation. *Transportation Research Part D: Transport and Environment*, *91*(January), 1–20. https://doi.org/10.1016/j.trd.2020.102692

Minderhoud, M. M., & Bovy, P. H. L. (2001). Extended time-to-collision measures for road traffic safety assessment. *Accident Analysis and Prevention*, *33*(1), 89–97. https://doi.org/10.1016/S0001-4575(00)00019-1

Mnih, V., Badia, A. P., Mirza, L., Graves, A., Harley, T., Lillicrap, T. P., Silver, D., & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. *33rd International Conference on Machine Learning, ICML 2016*, *4*, 2850–2869.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236

Montanino, M., & Punzo, V. (2015). Trajectory data reconstruction and simulation-based validation against macroscopic traffic patterns. *Transportation Research Part B: Methodological*, *80*, 82–106. https://doi.org/10.1016/j.trb.2015.06.010

Morbidi, F., Colaneri, P., & Stanger, T. (2013). Decentralized optimal control of a car platoon with guaranteed string stability. *2013 European Control Conference, ECC 2013*, 3494–

3499. https://doi.org/10.23919/ecc.2013.6669336

Naus, G. J.L., Ploeg, J., Van de Molengraft, M. J. G., Heemels, W. P. M. H., & Steinbuch, M. (2010). Design and implementation of parameterized adaptive cruise control: An explicit model predictive control approach. *Control Engineering Practice*, *18*(8), 882–892. https://doi.org/10.1016/j.conengprac.2010.03.012

Naus, Gerrit J.L., Vugts, R. P. A., Ploeg, J., Van De Molengraft, M. J. G., & Steinbuch, M. (2010). String-stable CACC design and experimental validation: A frequency-domain approach. *IEEE Transactions on Vehicular Technology*, *59*(9), 4268–4279. https://doi.org/10.1109/TVT.2010.2076320

Newell, G. F. (2002). A simplified car-following theory: a lower order model. *Transportation Research Part B: Methodological*, *36*(3), 195–205.

Noor-A-Rahim, M., Ali, G. G. M. N., Guan, Y. L., Ayalew, B., Chong, P. H. J., & Pesch, D. (2019). Broadcast Performance Analysis and Improvements of the LTE-V2V Autonomous Mode at Road Intersection. *IEEE Transactions on Vehicular Technology*, *68*(10), 9359–9369. https://doi.org/10.1109/TVT.2019.2936799

Orosz, G. (2016). Connected cruise control: modelling, delay effects, and nonlinear behaviour. *Vehicle System Dynamics*, *54*(8), 1147–1176. https://doi.org/10.1080/00423114.2016.1193209

Pereira, D. R., Piteri, M. A., Souza, A. N., Papa, J. P., & Adeli, H. (2020). FEMa: a finite element machine for fast learning. *Neural Computing and Applications*, *32*(10), 6393–6404. https://doi.org/10.1007/s00521-019-04146-4

Petrillo, A., Salvi, A., Santini, S., & Valente, A. S. (2018). Adaptive multi-agents synchronization for collaborative driving of autonomous vehicles with multiple communication delays. *Transportation Research Part C: Emerging Technologies*, *86*(October 2017), 372–392. https://doi.org/10.1016/j.trc.2017.11.009

Ploeg, J., Van De Wouw, N., & Nijmeijer, H. (2014). Lp string stability of cascaded systems: Application to vehicle platooning. *IEEE Transactions on Control Systems Technology*, *22*(2), 786–793. https://doi.org/10.1109/TCST.2013.2258346

Punzo, V., Borzacchiello, M. T., & Ciuffo, B. (2011). On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data. *Transportation Research Part C: Emerging Technologies*, *19*(6), 1243–1262. https://doi.org/10.1016/j.trc.2010.12.007

Qu, X., Yu, Y., Zhou, M., Lin, C. T., & Wang, X. (2020). Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: A reinforcement learning based approach. *Applied Energy*, *257*(September 2019), 114030. https://doi.org/10.1016/j.apenergy.2019.114030

Rafiei, M. H., & Adeli, H. (2017). A New Neural Dynamic Classification Algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(12), 3074–3083. https://doi.org/10.1109/TNNLS.2017.2682102

Richards, P. I. (2013). *Shock Waves on the Highway Author (s): Paul I . Richards Published by : INFORMS Stable URL : http://www.jstor.org/stable/167515 . SHOCK WAVES ON THE HIGHWAY **. *4*(1), 42–51.

Schulman, J. (2015). Trust Region Policy Optimization John. *Raisons Politiques*, *67*(3), 31–47. https://doi.org/10.3917/rai.067.0031

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). *Proximal Policy Optimization Algorithms*. 1–12. http://arxiv.org/abs/1707.06347

Shi, H., Zhou, Y., Wang, X., Fu, S., Gong, S., & Ran, B. (2022). A deep reinforcement learning-based distributed connected automated vehicle control under communication failure. *Computer-Aided Civil and Infrastructure Engineering*, 1–19. https://doi.org/10.1111/mice.12825

Shi, H., Zhou, Y., Wu, K., Wang, X., Lin, Y., & Ran, B. (2021). Connected automated vehicle cooperative control with a deep reinforcement learning approach in a mixed traffic environment. *Transportation Research Part C: Emerging Technologies*, *133*(August), 103421. https://doi.org/10.1016/j.trc.2021.103421

Shladover, S. E., Nowakowski, C., Lu, X. Y., & Ferlis, R. (2015). Cooperative adaptive cruise control: Definitions and operating concepts. *Transportation Research Record*, *2489*, 145–152. https://doi.org/10.3141/2489-17

Stipanović, D. M., Inalhan, G., Teo, R., & Tomlin, C. J. (2004). Decentralized overlapping control of a formation of unmanned aerial vehicles. *Automatica*, *40*(8), 1285–1296. https://doi.org/10.1016/j.automatica.2004.02.017

Swaroop, D., & Hedrick, J. K. (1996). String stability of interconnected systems. *IEEE Transactions on Automatic Control*, *41*(3), 349–357. https://doi.org/10.1109/9.486636

Tian, J., Zhu, C., Chen, D., Jiang, R., Wang, G., & Gao, Z. (2021). Car following behavioral stochasticity analysis and modeling: Perspective from wave travel time. *Transportation Research Part B: Methodological*, *143*, 160–176. https://doi.org/10.1016/j.trb.2020.11.008

Treiber, M., Hennecke, A., & Helbing, D. (2000). Congested traffic states in empirical observations and microscopic simulations. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *62*(2), 1805–1824. https://doi.org/10.1103/PhysRevE.62.1805

Van Arem, B., Van Driel, C. J. G., & Visser, R. (2006). The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Transactions on Intelligent Transportation Systems*, *7*(4), 429–436. https://doi.org/10.1109/TITS.2006.884615

Van Otterlo, M., & Wiering, M. (2012). Reinforcement learning and markov decision processes. *Adaptation, Learning, and Optimization*, *12*, 3–42. https://doi.org/10.1007/978-3-642-27645-3_1

Wang, C., Gong, S., Zhou, A., Li, T., & Peeta, S. (2019). Cooperative adaptive cruise control for connected autonomous vehicles by factoring communication-related constraints ☆. *Transportation Research Part C*, *April*, 1–22. https://doi.org/10.1016/j.trc.2019.04.010

Wang, C., Gong, S., Zhou, A., Li, T., & Peeta, S. (2020). Cooperative adaptive cruise control for connected autonomous vehicles by factoring communication-related constraints. *Transportation Research Part C: Emerging Technologies*, *113*(March), 124–145. https://doi.org/10.1016/j.trc.2019.04.010

Wang, Jian, Peeta, S., Lu, L., & Li, T. (2019). Multiclass information flow propagation control under vehicle-to-vehicle communication environments. *Transportation Research Part B: Methodological*, *129*, 96–121. https://doi.org/10.1016/j.trb.2019.09.005

Wang, Jiawei, Zheng, Y., Xu, Q., Wang, J., & Li, K. (2020). Controllability Analysis and Optimal Control of Mixed Traffic Flow With Human-Driven and Autonomous Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 1–15. https://doi.org/10.1109/tits.2020.3002965

Wang, M. (2018a). Infrastructure assisted adaptive driving to stabilise heterogeneous vehicle strings. *Transportation Research Part C: Emerging Technologies*, *91*, 276–295. https://doi.org/10.1016/j.trc.2018.04.010

Wang, M. (2018b). Infrastructure assisted adaptive driving to stabilise heterogeneous vehicle strings. *Transportation Research Part C: Emerging Technologies*, *91*(April 2017), 276–295. https://doi.org/10.1016/j.trc.2018.04.010

Wang, M., Daamen, W., Hoogendoorn, S. P., & van Arem, B. (2014a). Rolling horizon control framework for driver assistance systems. Part I: Mathematical formulation and non-cooperative systems. *Transportation Research Part C: Emerging Technologies*, *40*, 271–289. https://doi.org/10.1016/j.trc.2013.11.023

Wang, M., Daamen, W., Hoogendoorn, S. P., & van Arem, B. (2014b). Rolling horizon control framework for driver assistance systems. Part II: Cooperative sensing and cooperative control. *Transportation Research Part C: Emerging Technologies*, *40*, 290–311. https://doi.org/10.1016/j.trc.2013.11.024

Wang, M., Daamen, W., Hoogendoorn, S. P., & Van Arem, B. (2016). Cooperative Car-Following Control: Distributed Algorithm and Impact on Moving Jam Features. *IEEE Transactions on Intelligent Transportation Systems*, *17*(5), 1459–1471. https://doi.org/10.1109/TITS.2015.2505674

Wang, Yipei, Hou, S., & Wang, X. (2019). Crossing Traffic Avoidance of Automated Vehicle Through Bird-View Control, a Reinforcement Learning Approach. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3495727

Wang, Yu, Li, X., Tian, J., & Jiang, R. (2020). Stability analysis of stochastic linear car-following models. *Transportation Science*, *54*(1), 274–297. https://doi.org/10.1287/trsc.2019.0932

Wang, Z., Bian, Y., Shladover, S. E., Wu, G., Li, S. E., & Barth, M. J. (2020). A Survey on Cooperative Longitudinal Motion Control of Multiple Connected and Automated Vehicles. *IEEE Intelligent Transportation Systems Magazine*, *12*(1), 4–24. https://doi.org/10.1109/MITS.2019.2953562

Wang, Z., Wu, G., & Barth, M. J. (2018). A Review on Cooperative Adaptive Cruise Control (CACC) Systems: Architectures, Controls, and Applications. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, *2018-Novem*, 2884–2891. https://doi.org/10.1109/ITSC.2018.8569947

Whitham, G. B. (1955). On kinematic waves I. Flood movement in long rivers. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *229*(1178), 281–316. https://doi.org/10.1098/rspa.1955.0088

Zhang, L., & Orosz, G. (2016). Motif-Based Design for Connected Vehicle Systems in Presence of Heterogeneous Connectivity Structures and Time Delays. *IEEE Transactions on Intelligent Transportation Systems*, *17*(6), 1638–1651. https://doi.org/10.1109/TITS.2015.2509782

Zhang, L., & Orosz, G. (2017). Consensus and disturbance attenuation in multi-agent chains with nonlinear control and time delays. *International Journal of Robust and Nonlinear Control*, *27*(5), 781–803. https://doi.org/10.1002/rnc.3600

Zhang, Z., & Yang, X. (Terry). (2021). Analysis of highway performance under mixed connected and regular vehicle environment. *Journal of Intelligent and Connected Vehicles*, *4*(2), 68–79. https://doi.org/10.1108/jicv-10-2020-0011

Zheng, F., Liu, C., Liu, X., Jabari, S. E., & Lu, L. (2020). Analyzing the impact of automated vehicles on uncertainty and stability of the mixed traffic flow. *Transportation Research Part C: Emerging Technologies*, *112*(January), 203–219. https://doi.org/10.1016/j.trc.2020.01.017

Zheng, Yang, Li, S. E., Wang, J., Cao, D., & Li, K. (2016). Stability and scalability of homogeneous vehicular platoon: Study on the influence of information flow topologies. *IEEE Transactions on Intelligent Transportation Systems*, *17*(1), 14–26. https://doi.org/10.1109/TITS.2015.2402153

Zheng, Yang, Li, S. E., Wang, J., Wang, L. Y., & Li, K. (2014). Influence of information flow topology on closed-loop stability of vehicle platoon with rigid formation. *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, 2094–2100. https://doi.org/10.1109/ITSC.2014.6958012

Zheng, Yuan, Zhang, Y., Ran, B., Xu, Y., & Qu, X. (2020). Cooperative control strategies to stabilise the freeway mixed traffic stability and improve traffic throughput in an intelligent roadside system environment. *IET Intelligent Transport Systems*, *14*(9), 1108–1115. https://doi.org/10.1049/iet-its.2019.0577

Zheng, Z., Ahn, S., Chen, D., & Laval, J. (2011). Freeway traffic oscillations: Microscopic analysis of formations and propagations using Wavelet Transform. *Transportation Research Part B: Methodological*, *45* `(9), 1378–1388. https://doi.org/10.1016/j.trb.2011.05.012

Zheng, Z., & Washington, S. (2012). On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C: Emerging Technologies*, *25*, 18–33. https://doi.org/10.1016/j.trc.2012.03.006

Zhou, A., Gong, S., Wang, C., & Peeta, S. (2020). Smooth-Switching Control-Based Cooperative Adaptive Cruise Control by Considering Dynamic Information Flow Topology. *Transportation Research Record*, *2674*(4), 444–458. https://doi.org/10.1177/0361198120910734

Zhou, M., Yu, Y., & Qu, X. (2020). Development of an Efficient Driving Strategy for Connected and Automated Vehicles at Signalized Intersections: A Reinforcement Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*, *21*(1), 433–443. https://doi.org/10.1109/TITS.2019.2942014

Zhou, Y., Ahn, S., Chitturi, M., & Noyce, D. A. (2017). Rolling horizon stochastic optimal control strategy for ACC and CACC under uncertainty. *Transportation Research Part C: Emerging Technologies*, *83*, 61–76. https://doi.org/10.1016/j.trc.2017.07.011

Zhou, Y., Ahn, S., Wang, M., & Hoogendoorn, S. (2019). Stabilizing mixed vehicular platoons with connected automated vehicles: An H-infinity approach. *Transportation Research Part B: Methodological*, *xxxx*. https://doi.org/10.1016/j.trb.2019.06.005

Zhou, Y., Wang, M., & Ahn, S. (2019). Distributed model predictive control approach for cooperative car-following with guaranteed local and string stability. *Transportation Research Part B: Methodological*, *128*, 69–86. https://doi.org/10.1016/j.trb.2019.07.001

Zhu, M., Wang, X., & Wang, Y. (2018). Human-like autonomous car-following model with deep reinforcement learning. *Transportation Research Part C: Emerging Technologies*, *97*(October), 348–368. https://doi.org/10.1016/j.trc.2018.10.024

Zhu, Y., Wu, J., & Su, H. (2020). V2V-Based Cooperative Control of Uncertain, Disturbed and Constrained Nonlinear CAVs Platoon. *IEEE Transactions on Intelligent Transportation Systems*, 1–11. https://doi.org/10.1109/tits.2020.3026877