
**Center for Advanced Multimodal Mobility Solutions and Education**

**Project ID: 2022 Project 16**

# DYNAMIC COORDINATED SPEED CONTROL AND SYNERGISTIC PERFORMANCE EVALUATION IN CONNECTED AND AUTOMATED VEHICLE ENVIRONMENT

**Final Report**

by

Wei Fan (ORCID ID: https://orcid.org/0000-0001-9815-710X)
Chengying Hua (ORCID ID: https://orcid.org/0000-0002-4691-7100)

Wei Fan, Ph.D., P.E.
Director, USDOT CAMMSE University Transportation Center
Professor, Department of Civil and Environmental Engineering
The University of North Carolina at Charlotte
EPIC Building, Room 3261, 9201 University City Blvd, Charlotte, NC 28223
Phone: 1-704-687-1222; Email: wfan7@uncc.edu
for

Center for Advanced Multimodal Mobility Solutions and Education
(CAMMSE @ UNC Charlotte)
The University of North Carolina at Charlotte
9201 University City Blvd
Charlotte, NC 28223

**September 2024**

# ACKNOWLEDGEMENTS

This project was funded by the Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE @ UNC Charlotte), one of the Tier I University Transportation Centers that were selected in this nationwide competition, by the Office of the Assistant Secretary for Research and Technology (OST-R), U.S. Department of Transportation (US DOT), under the FAST Act. The authors are also very grateful for all of the time and effort spent by DOT and industry professionals to provide project information that was critical for the successful completion of this study.

# DISCLAIMER

# Table of Contents

# List of Tables

# List of Figures

# EXECUTIVE SUMMARY

In the vicinity of weaving areas, freeway congestion is nearly unavoidable due to their negative effects on the continuous freeway mainline flow. The adverse impacts include increased collision risks, extended travel time, and excessive emissions and fuel consumption. Dynamic Speed Harmonization (DSH), which is also known as Variable Speed Limit (VSL), has the potential to dampen traffic oscillation during congestion. However, the effectiveness of this strategy is typically limited by the low compliance rates of drivers and potential delays in information transmission and dissemination, and that control strategies can only affect a small area. Fortunately, new opportunities are emerging with the development of Connected and Automated Vehicles (CAVs) that can completely comply with the control system. CAVs can greatly help complement the intelligent transportation systems to enhance a variety of Measures of Effectiveness (MOEs), such as safety, mobility, and environmental sustainability.

The objective of this study is to investigate the effects of DSH in mixed traffic flow involving Human-Driven Vehicles (HDVs) and CAVs on the freeway. A safety-oriented DSH strategy based on Deep Reinforcement Learning (DRL) is developed to better understand how CAVs can improve operational performance. A holistic performance evaluation is conducted to quantify the impacts under different Market Penetration Rates (MPRs) of CAVs in multiple simulated scenarios. The mixed traffic flow integrated with DSH highlights the synergies across different metrics. The results reveal that for the recurrent congestion, the proposed method can enhance mobility and achieve co-benefits with safety, and sustainability could be improved under higher MPRs. Spatiotemporal features of bottleneck speed demonstrate that DSH powered by CAVs can smooth the speed variations for partial areas. Sensitivity analysis of headways indicates that high-level CAVs can further improve performance. For the nonrecurrent congestion, the DSH can further improve safety and enhance mobility with increasing MPRs. While special events may exacerbate congestion, their impact can be mitigated to some extent through DSH. Spatiotemporal patterns of speed variations demonstrate that the DRL controller has the capability to dampen oscillations. A series of numerical experiments also indicate the adaptability of the agent under adverse weather scenarios, and the differences of surrogate safety measurements in response to various parametric thresholds. Moreover, a Multi-Agent Dynamic Speed Harmonization (MADSH) system prevents the proposed strategy from getting stuck in local optimization. This study provides essential insights to foster a deeper understanding of the transformative potential of the proposed technique in promoting intelligent transportation systems.

# Chapter 1. Introduction

## 1.1 Problem Statement

When the design capacity of a bottleneck is exceeded by the traffic demand, vehicles coming from the upstream will queue up and traffic congestion will be generated in the downstream bottleneck area (Ghiasi et al., 2019). In this circumstance, mainline traffic flow will become unstable and stop-and-go phenomena may occur, which could lead to dramatic speed oscillation on the freeway segment. During this process, any slight driving behavior change made by even a single driver may produce a distinct change to the current traffic condition. This change is known as a 'shock wave', which propagates upstream and may cause a capacity drop and speed breakdown (Vrbanić et al., 2021).

Constructing additional road infrastructure is one way to address the aforementioned issues. However, it is not always a practical solution since increased capacity will evoke induced traffic demand and result in an undesirable cycle. Another way to alleviate congestion is utilizing the active traffic management strategy. Two most commonly used traffic control strategies are speed harmonization (also known as Variable Speed Limit (VSL)) and ramp metering. These solutions can better leverage the current roadway resource compared to building more infrastructure.

This research focuses on freeway mainline management, in which the general approach is to use speed harmonization to reduce spatiotemporal variations (Ma et al., 2016). The primary goal of this technique was also used to improve safety in work zones during inclement weather (Lu et al., 2010). Given that it could enhance headway and reduce speed oscillation (Ha et al., 2003), it would decrease the frequency and severity of crashes (Smulders, 1990). It can also smooth traffic flow to lessen lane-changing maneuvers which are risky in congestion. Dynamic Speed Harmonization (DSH) strategy can be implemented based on current traffic states by adjusting speed limits that are displayed on Variable Message Signs (VMS) for all lanes (Papageorgiou et al., 2008). Leveraging speed rather than volume is justified since it is easier to detect. Figure 1 demonstrates how DSH works. The downstream detectors collect the traffic parameters of the weaving area and send the information to the VSL controller. The controller then uses predefined strategies to optimize the speed limit of the control section, and the VMS will display new speed limits for the upstream traffic.

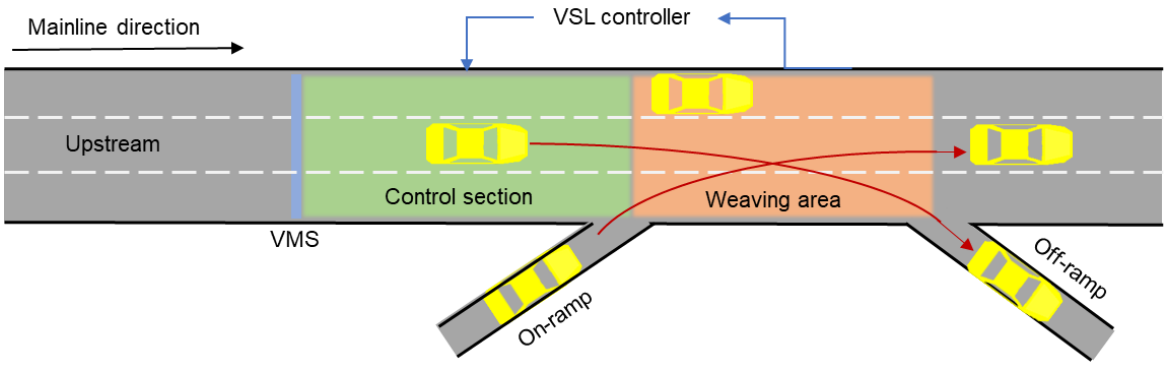


**Figure 1. The Mechanism of DSH on the Freeway**

However, there are a series of problems related to the conventional strategy. Firstly, the effectiveness of DSH is highly associated with the compliance rates of drivers and even fails due to unexpected human behaviors. Additionally, there may be delays in the collection of information, and it can only affect a small area. Fortunately, the emerging technology such as Connected and Automated Vehicles (CAVs) powered by DSH bring new opportunities to solve these problems through Vehicle-to-Everything (V2X) technology (Tajalli and Hajbabaie. 2018; Talebpour et al., 2013; Wang et al. 2016). Another limitation in previous studies is the requirement to continuously tune the dynamic traffic parameters in the fundamental diagram (Kušić et al. 2020). This can be overcome by Reinforcement Learning (RL) methods since it can learn and interact with various traffic conditions without exhibiting explicit traffic dynamics. Most importantly, it is unnecessary to set the same speed limit across all lanes (Wu et al. 2020). As is shown in Figure 1, when two different traffic flows (red lines) interfere with each other in the right two lanes in the weaving area, the left overtaking lane is actually not affected. Implementing a homogeneous speed limit may degrade the operational efficiency of the roadway.

When assessing the effectiveness of CAVs application, safety, mobility, and environmental sustainability are often the core elements. Several performance indicators can be used to quantify these Measures of Effectiveness (MOEs), such as the number of emergency braking, collision probability (Fang et al. 2015), and speed variations (Hegyi et al. 2002) for safety issues; Total Travel Time (TTT) (Alessandri et al. 1999), and time delay for mobility issues; fuel consumption and emission (Barth and Boriboonsomsin 2009; Vahidi and Sciarretta 2018) for sustainability issues. The majority of studies focused on one or two MOEs. While Tian et al. (2018) indicated that safety-oriented considerations can be included in mobility-based maneuvers as well to achieve sustainability, and co-benefits or trade-offs between them can be explored.

To investigate the operational performance of DSH in CAV environment, a holistic evaluation approach should be formulated. Besides, it is still in the initial stage of the

development of the vehicle-road synergy system, and CAVs will coexist with Human-Driven Vehicles (HDVs) for a long time. It is necessary to implement an effective strategy to control mixed flows under different Market Penetration Rates (MPRs) of CAVs in various scenarios. Establishing a simulation-based test environment also contributes to exploring the potential interactions existed in multiple MOEs.

## 1.2 Study Objectives

The main goal of this research is to investigate the effects of coordinated speed control in mixed traffic flow involving HDVs and CAVs on the freeway. The proposed work in this project is intended to complete the following objectives:

(1) To conduct a comprehensive review of the state-of-the-art and state-of-the-practice on DSH techniques, CAV technologies, DL-based traffic prediction models, DRL-based control methods, and their impacts on the freeway.

(2) To develop a DSH strategy based on DRL and to better understand how CAVs can improve operational performance.

(3) To evaluate and quantify the impact on mobility, safety, and sustainability, a comprehensive performance evaluation framework is formulated. A series of numerical experiments are conducted under different MPRs through various simulated scenarios.

(4) To explore the potential interactions between MOEs in mixed traffic flow.

## 1.3 Expected Contributions

This research aims to investigate state-of-the-art dynamic speed control strategies in mixed traffic flow that contains HDVs and CAVs on the freeway. The outcomes from this project are expected as follows:

(1) A comprehensive review of dynamic speed limit control is conducted.

(2) A DRL-based dynamic speed limit strategy is developed to control the vehicles on the congested freeway, and a MADSH system is further developed to prevent getting stuck in local optimization.

(3) A microscopic simulation environment is established for mixed traffic flows to evaluate the performance under various scenarios with different MPRs.

(4) A holistic performance framework is established by a variety of indicators to explore the co-benefits and trade-offs between MOEs.

(5) Sensitivity analysis under multiple traffic scenarios is conducted to verify the adaptation of the model.

## 1.4 Report Overview

In this chapter, the motivation of the research has been explained, followed by the study objectives and expected outcomes.

Chapter 2 summarizes a comprehensive literature review of the classical DSH Techniques for HDVs and CAVs, respectively. Previous methods that were conducted to implement DSH are classified into two categories: reactive and proactive methods. Considering the requirement to predict traffic patterns in a proactive approach, an overview of traffic prediction methods for intelligent vehicles is also summarized. Moreover, DSH based on state-of-art reinforcement learning technologies is also introduced.

Chapter 3 develops a single-agent DRL-based DSH strategy on the freeway recurrent bottleneck. In order to reduce the complexity and meet the time computation requirements, Intelligent Driver Model (IDM) is used in the DRL framework to model the mixed traffic flow. Meanwhile, CAVs are introduced to assess the effects of DSH in mixed traffic flow. A holistic performance evaluation is conducted to quantify the performance under different MPRs of CAVs in multiple simulated scenarios.

Chapter 4 employs the aforementioned DSH strategy at the nonrecurrent bottleneck including incidents on the freeway. To highlight the capabilities of DSH in safety improvements, a safety-oriented measurement is utilized, and comparative experiments are performed in a more indexed environment. A series of sensitivity analyses are conducted under different traffic demands, which include MPRs of CAVs, and time-to-collision thresholds.

Chapter 5 conducts an extension of the previous single-agent strategy. A distributed multi-agent DRL method is proposed when the central controller breaks down. A lane-based strategy is developed to verify the feasibility of setting differential speed limits for each lane. Moreover, the adaptation of the learning-based models is tested in the case of new scenarios and asymmetry driving behaviors.

Chapter 6 concludes the report by summarizing the proposed models and research results. Suggestions for future research directions are also provided.

# Chapter 2. Literature Review

## 2.1 Introduction

Due to the requirement of specific infrastructure and facilities for DSH systems, only a few institutions have conducted field tests. Many studies relied on the simulation-based approach, which can be classified into reactive and proactive methods. Reactive methods activate the VSL controller after congestion is detected (Malikopoulos et al., 2018). It is an offline algorithm based on the classical feedback control theory and can be used in stationary traffic. However, there is often a delay to manage the congestion, and fundamental diagrams are required to adjust the controller settings. Proactive methods are proposed to resolve this problem by anticipating future patterns at the beginning of congestion and conducting appropriate measures (Khondaker and Kattan, 2015). The following reviews the VSL algorithms in the DSH strategies.

The reactive and proactive methods used to implement DSH on the HDVs are introduced in Section 2.2. The problems of traditional ways and corresponding methods that take CAVs into account are described in Section 2.3. Section 2.4 presents the definition of reinforcement learning techniques, followed by state-of-art methodologies for HDVs and CAVs.

## 2.2 DSH Techniques for HDVs

The DSH has been utilized in the UK for safety reasons since the 1960s. At the very beginning, simulation practice was chosen considering the cost of the field test. It was not until the 21$^{st}$ century, this technique was widely implemented in Europe such as Germany, the Netherlands, France, and Sweden. The United States project started in Washington State in 2009 for the main purpose of safety. Prior to field testing, algorithms should always be developed and evaluated in the simulation since field testing may be expensive and, if done incorrectly, can have unintended consequences and detrimental effects on public traffic (Lu and Shladover, 2014).

### 2.2.1 Reactive Methods for HDVs

#### 2.2.1.1 Rule-based Control

For the first category, the rationale for determining speed limits is based on predetermined thresholds for a certain traffic flow situation with the objective of improving safety by reducing speed differences and stabilizing traffic flow. The rules are usually developed based on human experience rather than classical traffic flow theory. Weather and highway geometry may also be considered in the human-made rules activating DSH. The limitation of

rule-based strategies can be mainly attributed to the reaction delay. When control measures are taken, traffic conditions may already be in a state of breakdown, and this strategy has little ability to change the situation.

Based on microscopic modeling, Piao and McDonald (2008) explored the safety effects of in-vehicle changeable speed limit information. The simulation results indicated that the use of VSL on highways has a significant potential to improve traffic safety through reduced speed disparities between and within lanes, small time headways, tiny TTC, and a reduction in the frequency of lane changes. Nevertheless, deployment of in-vehicle VSL may potentially include certain safety issues as compared to roadside VSL. For instance, when showing information on on-board unit (OBU) in complex scenarios, it forces VSL to compete with other information providers. In addition, implementations of in-vehicle VSL may cause significant speed differences and frequent lane changes at low penetration rates. Last but not least, if VSL is mandated, roadside VSL are required temporarily while not all vehicles have an in-vehicle device.

#### 2.2.1.2 Local Feedback Control

Popov et al. (2008) designed a speed limit control method based on a distributed controller strategy for resolving shockwaves. The controller was dispersed in that there was a separate controller for each speed limit sign. Numerical optimization was used to improve the controller settings on the assumption that each controller has the same structure and set of parameters. With regard to the controller order and the degree to which the upstream and downstream traffic conditions were employed as inputs for the controller, the ensuing performances were contrasted for a number of configurations. Additional controllers identified in the literature were centralized model-based controllers with significant computational demands or switching systems employing just local information. The suggested approach offered a methodical approach for creating distributed controllers with the right quantity of upstream and downstream traffic data. Due to their high efficiency, the resultant controllers were desirable from an implementation standpoint. They merely employed data from the area and did not necessitate costly online computations. When compared to the uncontrolled case for the design scenario, the controller effectively addressed the shockwave and reduced the overall time spent by around 20%, which is equivalent to the performance of the best controllers documented in the literature.

Waller et al. (2009) tested a variety of variable speed limit and shoulder usage tactics and evaluated how they affected freeway traffic flow and safety. These tactics were shown to homogenize traffic and improve driving conditions, but they had little effect on the system's throughput. Furthermore, ITS tools that are needed to implement these methods,

enforcement concerns, potential roadblocks to their adoption, and a methodology for a cost-benefit analysis to establish their practicality, were also presented.

Iordanidou et al. (2015) suggested an expanded feedback control technique for Mainstream traffic flow control (MTFC) enabled by VSLs, taking numerous bottleneck sites into consideration. For the assessment of the controller using a verified macroscopic model, feedback-based outcomes were compared with optimum control results. Despite the feedback controller also taking into account a number of practical and safety limits, it was demonstrated that the performance of the feedback controller approaches the results of optimum control.

Müller et al. (2015) used a local feedback MTFC in a microscopic simulation of an on-ramp merge bottleneck. Important details that were not previously captured in macroscopic modeling were revealed by traffic behavior. The slower traffic reaction to speed limit adjustments was mostly caused by the more realistic VSL application at particular locations rather than along an entire highway segment. Furthermore, compared to what was shown at the macroscopic level, the nonlinear capacity flow/speed limit connection was more evident in the microscopic model. Significant improvements in traffic conditions were attained when the control law was modified as necessary.

Table 1. **Reactive DSH Methods for HDVs** summarizes the reactive DSH methods for HDVs.

**Table 1. Reactive DSH Methods for HDVs**

| Author | Year | Control algorithms | Metrics | Results |
|---|---|---|---|---|
| Piao and McDonald | 2008 | - | Safety (speed variance) | Reduced speed differences between and within lanes and number of small headways. |
| Popov et al. | 2008 | Distributed controller | Mobility (travel time) | Prevented the generation of shock waves by applying VSL, Total Time Spent (TTS) was reduced by 20% compared to the uncontrolled case. |
| Waller et al. | 2009 | Decision-tree based | Safety (speed variance) | Reduced speed variance but not throughput. |
| Iordanidou et al. | 2015 | Generic integrated | Mobility (travel time and time delay) | The feedback control was able to come close to the optimal control results. |
| Müller et al. | 2015 | - | Mobility (throughput) | Improvements on the 40% TTS. |

### 2.2.2 Proactive Methods for HDVs

#### 2.2.2.1 Classical Control

Recent research has mostly focused on more complex control logic that operates in a proactive manner with the aim of improving mobility while guaranteeing safety, in other words, proactively avoiding a problem before it occurs. Controlling the speed limit is essentially a process of optimization. A typical one is Model Predictive Control (MPC) (Hegyi et al. 2005). In the optimization, decision variables contain speed limit values (Hadiuzzaman and Qiu, 2013) and control locations (Zhang et al., 2015). Constraints include relevant traffic regulations. Objective functions entail multiple indicators, such as travel time, vehicle miles traveled (Lu et al., 2010), crash probabilities, throughput, queue length, and emission (Lin et al., 2010).

The traditional proactive method usually needs a prediction model on a macroscopic level to forecast traffic movement. Most previous prediction models are extensions of Payne's (1971) second-order model and need a fundamental diagram to describe the traffic flow. The main difference among the various models depends on the expression of drivers' desired speed. The MPC method highly relies on the accuracy of traffic state prediction, which is difficult to achieve due to the complexity of the transportation system. In addition, the use of a macroscopic traffic flow prediction model cannot completely reflect shockwaves caused by changes in driving behaviors. On the other hand, microscopic traffic models can more precisely describe disturbances in detail.

Lin et al. (2004) developed two online algorithms for VSL controls at highway work zones that may simultaneously meet the goals of queue reduction or throughput maximization and fully use all dynamic functionalities. This work carried out comprehensive tests based on virtual roadway systems that have been calibrated using field data to assess the efficacy of these suggested algorithms. Simulation analysis findings demonstrated that VSL algorithms can result in significant increases in work-zone throughputs and decreases in overall vehicle delays. Furthermore, compared to other non-controlled traffic scenarios, traffic flows that use VSL controls often showed fewer speed fluctuations. The decrease in speed variation may indirectly improve traffic safety in work zones as a whole.

By enforcing lower speed limits upstream and higher speed limits downstream of the point where collision risk is being tracked in real-time, VSL deployment increased safety. This improvement was shown in medium-to-high-speed motorway regimes, but no advantage was seen in low-speed circumstances. The suggestions for implementing VSL by Abdel-Aty et al. (2006) were as follows: decreasing speed limits upstream and raising speed limits downstream of a site of interest; altering the speed limit abruptly in space (no gap distance);

introducing adjustments to the speed limit gradually (5 mph every 10 minutes); and the speed restriction increases up- and downstream should be significant (15 mph) and executed close to the point of interest (within 2 miles).

To lower the risk of crashes on instrumented motorways, Lee et al. (2006) analyzed automated control systems for changing speed restrictions. Based on short-term variations in traffic flow parameters, a real-time accident prediction model was created to evaluate crash potential. In order to evaluate control logic, a crash prediction model was integrated with a microscopic traffic simulation model to mimic changes in traffic circumstances as a result of changeable speed restrictions realistically. The study examined how strategy control parameters affected the overall trip duration and crash potential reduction within this integrated assessment framework. The study's findings suggested that variable speed limits, which temporarily lower them in hazardous traffic situations when collision possibility exceeds a predetermined threshold, might reduce crash potential by 5–17%.

Hadiuzzaman and Qiu (2013) suggested a novel VSL management technique that explicitly took into account the fundamental diagrams (FDs) at active bottlenecks and their upstream-downstream segments. An analytical model based on the cell transmission model (CTM) was built in order to comprehend the efficiency of the VSL regulation in great detail. It was suggested to represent two changes to the FD: (1) an active bottleneck with a capacity loss once feeding flow reaches its limit; and (2) varying free-flow speeds for cells controlled by VSL. The local demand-supply technique was used to modify the CTM's boundary condition in order to accommodate these changes. The suggested VSL control model was applied in a North American urban highway corridor as a component of the VSL control algorithm using the model predictive control technique. This simulation research showed that VSL works best for traffic mobility when there is congestion.

Islam et al. (2013) examined the safety and mobility effects of a model predictive VSL control approach. In order to forecast traffic conditions and give speed for improving corridor operational performance, the approach used second-order traffic flow models. The optimal scenario was determined by doing a sensitivity analysis of the VSL update frequency and the safety limitations of the VSL approach. In Edmonton, Alberta, Canada, a section of Whitemud Drive, an urban highway corridor, was chosen as the research location. A unique software module was used to implement the suggested VSL technique in the microsimulation platform. By estimating the collision probability for each scenario using a matched case-control logistic regression approach, a real-time collision prediction model was created for the same research area. The findings suggested that the suggested VSL control approach can increase safety by around 50% and mobility by about 30%. The best results were obtained with a VSL update frequency of 5 min and a maximum speed variation

of 10 km/h between succeeding time steps. The field implementation of VSL control may benefit from this discovery.

Talebpour et al. (2013) established a method to explore the impacts of SH on traffic flow features and safety that relied on a cognitive risk-based microscopic simulation model capable of endogenously accounting for occurrences. In order to accomplish speed harmonization inside the microscopic simulation model, a wavelet transform-based approach to identify shock wave formation was paired with a reactive speed limit selection algorithm. There were three different sets of simulations. The application of the speed harmonization management approach in crowded situations resulted in a considerable improvement in traffic flow characteristics. An ideal spot to execute the adjustments to the speed restriction was upstream of the place where shock waves are detected, according to the analysis of an FD. The investigation also demonstrated the need of adhering to speed limits for speed harmonization to be successful.

Li et al. (2014) created a VSL control technique to lessen the danger of secondary accidents during bad weather. The VSL technique was suggested to dynamically alter the speed restrictions in accordance with the current traffic and weather circumstances by assessing the occurrence condition of a secondary collision. In order to replicate vehicle moves with the VSL control, a car-following model was modified. To assess the control effects of VSL, two surrogate safety metrics based on the time-to-collision were utilized. In a simulation, five weather possibilities were assessed. The outcomes demonstrated that the VSL technique successfully lowers the probability of secondary crashes in a variety of meteorological conditions. Both the time integrated time-to-collision (TIT) and the time exposed time-to-collision (TET) were shortened by 38.19% to 41.19% and 41.45% to 50.74%, respectively. The impacts on safety were contrasted with those of an earlier VSL method. The outcomes demonstrated that their technique typically works better than the prior one. They also assessed the impact of the driver's adherence to the speed restriction on the efficiency of VSL control.

Zhang et al. (2015) examined VSL systems and aims to improve system designs with moving variable message signs (VMSs). The number of VMSs to be deployed, their positions, and the speed limits posted on the VMSs were the decision variables for the optimization issue, which was written as a large mixed-integer nonlinear programming problem. One goal was to limit the negative environmental effects of highway traffic, and the other was to smooth the flow propagation. In addition, a genetic algorithm was suggested to resolve the challenging issue. The use of numerical examples on a real motorway stretch demonstrated the effectiveness of VSL in achieving smooth flow and minimizing the impact of freeway traffic on the environment.

In previous studies, prediction-based optimum VSL management has been carried out using the macroscopic traffic flow model. One of the most important factors in the development of the prediction is how the drivers react to the recommended VSL. Yet, this impact was either ignored or poorly modelled in prior studies (by assuming that a constant proportion of drivers will follow the VSL, regardless of various traffic conditions). Fang et al. (2015) suggested a dynamic driver response model as a solution to this issue. To represent the link between drivers' intended speed, the recommended VSL value, and actual traffic status factors, the model was developed and calibrated using field data. This model was used to quantitatively define the drivers' dynamic response to different VSL levels while taking into account the present traffic circumstances. Moreover, it was demonstrated through a simulation using real-world data that the proposed VSL control algorithm with improved driver reaction modeling accurately forecasts traffic conditions and significantly lowers crash probability in the traffic network.

Wang et al. (2017) used microsimulations to assess several active traffic management (ATM) techniques to increase the safety of a busy highway weaving stretch. ATM strategy effects on traffic safety were assessed using crash probabilities and the Surrogate Safety Assessment Model. Based on the real-time safety analysis model for weaving segments, the crash probabilities were determined. The techniques included ramp metering (RM), variable speed limit (VSL), and combined RM and VSL (RM-VSL). Overall, the findings indicated that the ATM techniques enhanced the safety of the weaving section. The modified ALINEA RM algorithms surpassed the original ALINEA algorithm in terms of safety since they took lane occupancy and other factors into account. The 45 mph VSLs, which were situated upstream of the investigated weaving portion, greatly improved safety without appreciably increasing average travel time. A combined RM-VSL approach was also suggested with the intention of enhancing traffic safety through the application of RM and VSL. To avoid lengthy waits on ramps, the modified ALINEA RM was changed in the combined RM-VSL strategy in accordance with the duration of the queue. The findings demonstrated that the combined RM-VSL technique reduced conflicts by 16.8% and crash chances by 6.0%.

#### 2.2.2.2 Open-loop Control

Alessandri et al. (1999) examined a traffic control issue with a dynamic macroscopic model by simulated analysis. In order to enhance traffic behavior near congestion, an optimal control problem was formulated for variable-speed signaling. With the help of real-time estimations of the traffic density, a speed signaling system was activated using a traffic state estimator based on the extended Kalman filter. A performance criterion was minimized (or maximized) in order to determine the closed-loop variable-speed signaling control law. The Powell's method-based optimization process was computationally tractable for off-line

execution on inexpensive machines as well. Results from simulations showed how effective the suggested strategy is in reducing congestion.

Hegyi et al. (2005) provided a model predictive control method for coordinating changeable speed restrictions for highway traffic in an effort to reduce shock waves. To reduce overall journey time, they started by optimizing continuous valued speed restrictions. Then, they added a safety restriction that stops cars from encountering speed limit decreases greater than 10 km/h. Moreover, they took into consideration discrete speed limitations to improve the congruence between the computed and applied control signals. A benchmark problem served as an illustration of their strategy.

Yang et al. (2013) presented two methods for proactive VSL on motorway portions with recurrent congestion. The suggested fundamental model calculated the speed limit while using embedded traffic flow relations to forecast the evolution of the congestion pattern over the anticipated time horizon. In order to address the challenge of collecting driver responses to VSL control, this work also suggested an enhanced model that further utilized Kalman Filter to boost the precision of traffic state prediction. With various traffic scenarios and various control goals, both models were studied. Their thorough simulation research using a VISSIM simulator, calibrated with field data from prior VSL demonstration sites, demonstrated the advantages of the suggested VSL control model when compared to the situation without VSL. The outcomes also showed that both proactive models may surpass the basic models and greatly cut down on travel time as well as the number of pauses over the places where bottlenecks frequently occur. The model with the control of reducing speed variation was found to perform much better than other models for a variety of chosen MOEs, including average number of stops and average trip duration.

Table 2 summarizes the reactive DSH methods for HDVs.

**Table 2. Proactive DSH Methods for HDVs**

| Author | Year | Control algorithms | Metrics | Results |
|---|---|---|---|---|
| Lin et al. | 2004 | Two online algorithms | Mobility, environmental impact | Reduced work-zone time delay and increase throughputs, the former performs better in speed variances. |
| Abdel-Aty et al. | 2006 | Crash prediction model | Safety | Improvement in medium-to-high-speed regimes, but no benefit in congested situations. |

| | | | | |
|---|---|---|---|---|
| Lee et al. | 2006 | Crash prediction model | Safety, mobility | Reduced crash potential, but higher travel time. |
| Hadiuzzaman and Qiu | 2013 | Cell transmission model (CTM) | Mobility (throughput, travel time) | Increased the flow volume by 5–7% and reduce the total travel time by 9–11% in the simulation. |
| Islam et al. | 2013 | MPC | Safety, mobility | Improved both safety and mobility by approximately 50% and 30%. |
| Talebpour et al, | 2013 | Wavelet transform | Safety, mobility | Significant improvements in flow and safety. Analyze the optimal location and time for the VSL transition. |
| Li et al. | 2014 | Modified car-following models | Safety | Reduced the risks of secondary collisions in various weather types. 41.45%–50.74% less TET, 38.19%–41.19% less TIT. |
| Zhang et al. | 2015 | Mixed-integer nonlinear programming | Safety, environmental impact | VSL can effectively improve the safety and environmental impact of freeway traffic. |
| Fang et al. | 2015 | Dynamic driver response model | Safety | Predicted traffic states more precisely, and effectively reduced the crash probabilities. |
| Wang et al. | 2017 | Consolidated RM-VSL, the modified ALINEA RM | Safety | Reduced the number of conflicts by 16.8% and decreased the crash odds by 6.0%. |
| Alessandri et al. | 1999 | Second-order METANET (Extended Kalman filter) | Mobility | Simulation results demonstrated the efficacy of the proposed approach for preventing and reducing congestion. |
| Hegyi et al. | 2005 | Second-order METANET | Mobility (travel time) | The system travel time can be reduced by up to 20.1%. |
| Yang et al. | 2013 | Kalman filter | Mobility (stop times, travel time) | Reduced vehicle stops by up to 42.4% and the travel time by up to 17.6% in the simulation. |

## 2.3 DSH Techniques for CAVs

### 2.3.1 The Problems of Traditional Ways

The primary drawback with traditional DSH algorithms' effectiveness in CAVs is their inability to adjust their control strategy to a new traffic circumstance, in which case they perform less ideally. The majority of current studies can only suggest drivers adjust vehicles' speed. These strategies rely on the compliance rates of drivers and even fail due to unexpected human behavior. Additionally, most studies depend on a few fixed sensors with low resolution and geographic limitations. There are delays in the collection of information, and control strategies can only affect a small area. Fortunately, emerging CAVs bring opportunities to address the problems of traditional DSH techniques through vehicle-to-everything (V2X) technology. Congestion can be alleviated with DSH powered by CAVs (Tajalli and Hajbabaie., 2018).

The disadvantage of DSH techniques in CAVs is the presumption that the communication network is error-free and that information is transferred to the cars without delay or information loss. DSH for CAVs also has the drawback of being obsolete at extremely high penetration rates. The advantages of CAVs in future mixed traffic flows are apparent in enhancing the macroscopic traffic characteristics of highways and eliminating the requirement for separate control, regardless of the penetration rate. This subsection gives an overview of research in this field, starting with the earliest ones. It categorizes the strategies in the same way as DSH in HDVs did.

### 2.3.2 Reactive Methods for CAVs

Most methods in this category are rule-based, in which the rationale is adjusting speed limits by predetermined thresholds for a certain traffic condition. It gathers the current traffic data from the downstream congestion area, and maintains traffic states at critical density.

Li et al. (2017) created a control approach that combines a cooperative adaptive cruise control (CACC) system with a variable speed limit (VSL) to lower the probability of rear-end collisions near motorway bottlenecks. First, a testbed for microscopic simulation was built, in which the precise PATH CACC models and substitute safety parameters of the time-exposed time-to-collision (TET) and time-integrated time-to-collision (TIT) were used. For the proposed vehicle-to-infrastructure system of CACC and VSL, a feedback control algorithm was subsequently devised. According to the simulation findings, the suggested integration system with 100% CACC penetration rate may significantly lower the risks of rear-end collisions, with a 98% drop in TIT and TET. When compared to manual, uncontrolled cars, the average trip time was likewise reduced by 33%. Moreover, the proposed integrated system's safety benefits exhibited a good degree of stability at a variety of bottlenecks with varying degrees of speed

decreases. According to the results of the sensitivity analysis, the safety performance was significantly impacted by the CACC penetration rate. Since the penetration rate of the CACC was low, the VSL control was crucial in lowering the likelihood of a rear-end accident. The mixed traffic flow of manual and CACC cars was less harmful when combined with VSL controls.

Li and Wanger (2019) conducted a thorough evaluation based on simulation utilizing a 5.3 km section of the Auckland Motorway and traffic data given by New Zealand Traffic Agent to explore the possible gains or losses due to the introduction of AVs into current highway systems. On mobility, safety, pollutants, and fuel usage, they examined the effects of various AV shares. The highway was evaluated both with and without traffic management under four different traffic scenarios: heavy traffic (>0.95*capacity), light traffic (0.7*capacity), free-flow traffic (0.5*capacity), and future traffic (3*heavy traffic volume).

Wu et al. (2020) created a control approach to lessen the likelihood of a rear-end collision during bottlenecks on the motorway while it is foggy. With consideration of the various correlations between the gap and visibility distance, a VSL control algorithm was created. Moreover, the VSL approach was evaluated in a fully CV setting. To merge the VSL with CV control, a framework for feedback control was created. Using the use of the microsimulation VISSIM and the IDM, which was used to account for cars following in the CV environment, the suggested VSL approach was put into practice and evaluated for a highway segment with a bottleneck. Ultimately, two measurements—total travel time (TTT) and time-to-collision at braking (TTC brake)—were used to assess how well the suggested control approach worked. The findings showed that the VSL control significantly decreased the likelihood of a rear-end collision and that compliance rates might have an impact on the control's effectiveness. It was also discovered that the CV environment might increase traffic efficiency and safety.

### 2.3.3 Proactive Methods for CAVs

Khondaker and Kattan (2015) introduced a control algorithm for maximizing mobility, safety, and environmental benefit at the same time in a connected vehicle environment. They concentrated on individual driver behavior (such as acceleration and deceleration) using a Model Predictive Control (MPC) method. Total Travel Time (TTT) was calculated using a microscopic traffic flow prediction model; immediate safety was assessed using a surrogate safety measure called Time-To-Collision (TTC); and the environmental effect was assessed using a microscopic fuel consumption model called VT-Micro. The ideal speed limit values were modified based on real-time driver adherence to the imposed speed limit. To assess the effectiveness of the established technique for various weights in the objective function and for two various percentages of CV, a sensitivity analysis was carried out. The findings showed that, with 100% penetration rate, the proposed VSL strategy consistently outperformed the uncontrolled scenario,

resulting in up to 20% reductions in total trip time, increases in safety of 6–11%, and fuel savings of 5–16%. The scenario that focused just on safety produced more optimal gains than the multi-criteria optimization. One may thus contend that in the event of 100% CV penetration rates, optimizing alone for safety was sufficient to obtain simultaneous and ideal improvements in all measurements. Mixed findings were achieved when optimizing for simply mobility or fuel consumption, which demonstrated a higher accident risk for lower penetration rates. This showed that with such a high penetration rate, multi-criteria optimization is essential to get the best and most balanced results.

Müller et al. (2015) investigated the impact of various autonomous vehicle penetration rates on MTFC-VSL using VSL as actuators. The findings of the simulations demonstrated that greater performance is correlated with higher penetration rates, with a considerable effect up to 30% penetration rate and very small improvements above that, and that mixed VSL application methods may also be harmful to traffic.

Malikopoulos et al. (2016) focused on the issue of speed regulation of a number of AVs before they reach a highway speed zone. They defined the control issue and offer an analytical, closed-form, real-time implementation-ready solution. Under the strict safety constraint of avoiding rear-end collisions, the solution produced the best acceleration and deceleration for each vehicle. A tiny simulation testbed was used to assess the solution's performance, and it demonstrated that the suggested strategy considerably decreases both fuel consumption and travel time. Fuel consumption was decreased for three different traffic volume levels by 12–17% for the VSL algorithm, by 18–34% for the vehicular-based speed harmonization (SPD-HARM) algorithm, and by 19–22% compared to the baseline scenario, which takes human-driven vehicles into account. Comparable improvements in travel time were made in comparison to the baseline scenario, the VSL algorithm, and the vehicular-based SPD-HARM algorithm, which ranged from 26% to 30%, 3% to 19%, and 31%-39%, respectively.

Yu and Fan (2016) illustrated the optimal VSL technique for a motorway stretch with several bottlenecks in a CAV environment. An enhanced cell transmission model (CTM) that accounted for capacity reduction and mixed traffic flow, including conventional human-driven automobiles and heavy vehicles, as well as AVs, was used to build the VSL control. A multiple-objective function was developed with the intention of enhancing operational effectiveness and facilitating speed transitions. In order to resolve the integrated VSL control problem, a genetic algorithm (GA) was used. The planned control structure was put to the test on a real-world length of roadway. Sensitivity analyses were carried out to examine the effects of the communication range and CAV penetration rate. The designed VSL control not only increased overall efficiency but also lowered emission. The simulation findings also surpassed the VSL control alone when it integrates vehicle-to-vehicle (V2V), vehicle-to-infrastructure

(V2I), and infrastructure-to-vehicle (I2V) communication. Better performance can also be attained when the penetration rate of CAVs rises.

Li et al. (2019) presented a revolutionary VSL-MPC technique employing a chain of CAVs. In this work, the influence of the innovative VSL control was added to a discrete first-order model that took the capacity decrease of jam waves into account. They cast an MPC scheme into a multi-layer control structure based on the extended model. On a fictitious multi-lane motorway with transient jam waves, a series of microscopic simulation experiments were carried out to validate the expanded model and assess the suggested control technique. The traffic flow model can replicate the development of traffic under VSL management, and a 3.7% decrease in the overall delay time of mainline traffic may be achieved.

Table 3 summarizes the reactive and proactive DSH methods for CAVs.

**Table 3. Reactive and Proactive DSH Methods for CAVs**

| Author | Year | Model | Scenarios | MPRs | Comparisons | Results |
|---|---|---|---|---|---|---|
| **Li et al.** | 2017 | ACC-VSL | 10 km, traffic demand 1600 veh/h/lane | 0%–100% Avs | ACC only, VSL only | 80% lower Time-to-collision (TTC), 77% lower Time Exposed Time-to-collision (TET). |
| **Li and Wagner** | 2019 | Rule-based | 5.3 km, three demand | 0%–100% Avs | no control | 83% higher throughputs, 88% higher maximum volume, 26% lower travel time against 0% Avs, 31% lower fuel consumption at 70% Avs. |
| **Wu et al.** | 2020 | Feedback CV-VSL | 9.3 mile, low and high volume | 0% and 100% CVs | no VSL or CV | Reduced rear-end crashes affected by compliance rates. Enhancing safety and efficiency. |
| **Khondaker and Kattan** | 2015 | MPC | 8km includes incidents, 2000 veh/h | 50% and 100% CAVs | no control | 20% lower TTT,<br>11% improved TTC,<br>16% lower fuel consumption |
| **Müller et al.** | 2016 | Feedback cooperative model | 4.3km, congestion forms once the ramp demand increases | 0%–100% Avs | Point-VSL | 49.5% lower time delay at a 40% AV, 47.9% lower delay at a 90% AV |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Malikopoulos et al.** | 2016 | Hamiltonian analysis | 2km, three traffic demands | 100% Avs | no control, VSL, SPD-HARM | 22% lower fuel consumption, 30% lower TT |
| **Yu and Fan** | 2019 | Genetic Algorithm | 8km includes multiple bottlenecks | 0%–10% CAVs | no control, VSL-only, VSL-V2X | 36% lower TTT, 68% lower delays, 66% lower number of stops, 7.6% lower emissions |
| **Li et al.** | 2019 | MPC | 10 km, three lanes 5400 veh/h | 20% CAV | no-control | Total delay reduction of 3.7% |

## 2.4 Reinforcement Learning (RL) Methodologies

The previous studies require information on the traffic flow dynamics in terms of the fundamental diagram (relation between flow, speed, and density) to tune the controller parameters. Recently, RL derived from machine learning technology has provided another efficient solution to the optimal control problems in intelligent transportation systems. Optimization of DSH requires the determination of an optimal policy for displaying speed limits as actions on VMS or delivering them directly to vehicles in the CAV environment. RL also outperforms classical optimization methods since it excludes the requirement to comprehend the explicit model of traffic flow dynamics and how VSL achieves (Kušić et al., 2020). The main problem with using RL is that the learning process is not well explained. It is unable to provide a precise reason for why the agent discovers a certain course of action.

### 2.4.1 The Classification of RL

The following figure contains a series of classical RL algorithms, and these algorithms can be classified from different aspects, such as model-based and model-free; value-based and policy-based; round update and single-step update; on-policy and off-policy. More details can be found in *Spinning up* of OpenAI.

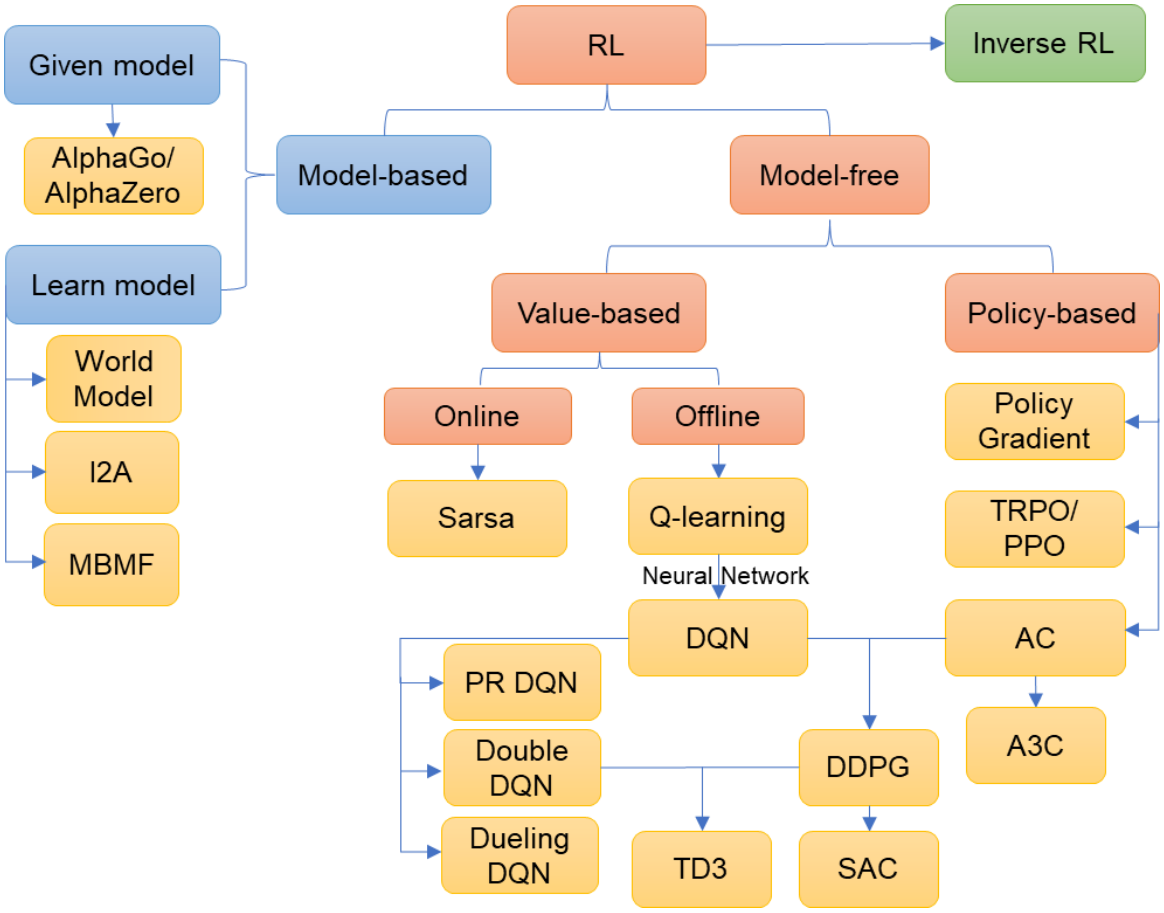


**Figure 2. Taxonomy of RL algorithms**

The Model-free will not learn and understand the environment. The Model-Based is the opposite, in which it uses a model to simulate the environment and get feedback. Common methods are AlphaGo (Silver et al., 2016); World Models (Ha and Schmidhuber, 2018); I2A (Imagination-Augmented Agents) (Weber et al., 2017); MBMF (Model-Based RL with Model-Free Fine-Tuning) (Nagabandi et al., 2017). Compared with Model-Free, Model-Based has one more step to simulate the environment, and all the situations that will happen are predicted, and then the best situation is selected.

The Policy-Based method directly outputs the probability of the next action, and the action is selected according to the probability. However, the action may not always be selected with the highest probability, and it still needs to be considered as a whole, which is applicable to discontinuous and continuous actions. Common methods are Policy Gradient (Sutton et al.,

2000); TRPO (Trust Region Policy Optimization) (Schulman et al., 2015); PPO (Proximal Policy Optimization) (Schulman et al., 2017). The Value-Based method outputs the value of the action and selects the action with the highest value. It applies to non-sequential actions. Common methods are Sarsa and DQN (Deep Q-Networks) (Mnih et al, 2015). The Actor-Critic is a combination of the two. The actor makes actions based on probability, and the critic gives values based on actions, thereby accelerating the learning process. Common methods are A2C / A3C (Asynchronous Advantage Actor-Critic) (Mnih et al., 2016); DDPG (Deep Deterministic Policy Gradient) (Lillicrap et al., 2015); TD3 (Twin Delayed DDPG) (Fujimoto et al., 2018); SAC (Soft Actor-Critic) (Haarnoja et al., 2018).

The round update method refers to updating after the entire learning process is over. Common methods include Monte-Carlo learning and the basic policy gradients. The single-step update method means that each step in the learning process is updated. Common methods include Q-learning (Watkins et al., 1992), Sarsa, and upgraded policy gradients. In comparison, the single-step update method is more efficient.

On-policy means that the agent must participate in the learning process, and the typical algorithms are Sarsa, PPO, and A3C. Off-policy refers to not only participating in it by itself, but also learning according to the learning process of others. Typical methods are DQN and DDPG.

RL is a known reward function. However, when the task is very complex, the reward function is often difficult to determine. Inverse RL is used to solve this problem and find the best reward function according to the expert strategy.

### 2.4.2 RL for HDVs and CAVs

When traffic on the urban freeway is steady in a spatial-temporal environment, DSH can give acceptable performance. Nevertheless, it loses some of its efficacy when the capacity of a highway is reduced or when the traffic circumstances are subject to abrupt oscillations in traffic demand. Being one of the ML methods, RL offers an ideal balance between the complexity and effectiveness of the many data-driven traffic control techniques (Kušić et al., 2020). It is beneficial to use RL-VSL because it has continual self-adaptation capabilities that can deal with control issues caused by brand-new, unanticipated traffic situations.

Zhu and Ukkusuri (2014) built a link-based dynamic network loading model to mimic the propagation of traffic flow permitting dynamic speed limits. By examining the difference between the queue-forming end and the dissipation end, shockwave propagation was clearly identified and recorded. Second, a real-time control mechanism was used to handle the Markov Decision Process (MDP) problem. In order to set time-dependent link-based speed limitations, the controller was described as an intelligent agent interacting with the stochastic network

environment. Using the R-Markov Average Reward Technique (R-MART) based reinforcement learning algorithm, the optimal speed limit scheme was obtained based on several metrics, such as total throughput, delay time, and vehicle emission. Compared to the basic scenario of no speed limit enforcement, the overall travel time and emissions (in terms of CO) were reduced by around 18% and 20%, respectively.

Walraven et al. (2016) provided an innovative reinforcement learning-based approach to traffic flow optimization. In order to alleviate traffic congestion, they employed Q-learning to understand the rules governing the maximum speed that is permitted on a roadway. The difference between their work and previous methodologies was traffic estimates. A number of simulation studies demonstrated that the resultant policies greatly decrease traffic congestion when there is a large demand and that the quality was enhanced by traffic predictions. Also, the policies were strong enough to handle erroneous density and speed observations.

A VSL control technique based on Q-learning was suggested by Li et al. (2017). A QL-based offline agent and an online VSL controller were both incorporated into the controller. To accomplish a long-term objective of system optimization, the VSL controller was taught to discover the optimal speed limits for diverse traffic conditions. Using a modified cell transmission model for a highway recurring bottleneck, the control effects of the VSL were assessed. The cell transmission model had a new parameter that takes into consideration the drivers' excessive speed in light traffic. Two scenarios that took into account both constant and varying traffic demands were assessed. The outcomes demonstrated that the suggested QL-based VSL technique worked better than the feedback-based VSL strategy. More precisely, the suggested VSL management technique decreased system travel time by 49.34% in a steady demand and by 21.84% in a fluctuating demand.

Similarly, Li et al. (2020) recommended a VSL control method based on RL to lessen crash risks brought by oscillations. The state, action, and reward were thoughtfully created to increase safety. To evaluate the safety close to motorway bottlenecks, a rear-end crash risk model was utilized. The simulation framework was modified from the cell transmission model. The outcomes demonstrated that the suggested RL-based VSL control effectively decreased the accident risks by 19.4%. An online learning function was built to increase stability and managed well under poor driver compliance with the aid of continual learning.

The issue of the exponential increase of the state space dimension and the high number of learning iterations are both disadvantages of RL. Using approaches for function approximation, this can be resolved. Kušić et al. (2018) examined three distinct feature-based state representation techniques in terms of the convergence of Total Time Spent. The competing methods were assessed using the VISSIM with a representative traffic model. The findings demonstrated that function approximation approaches surpassed RL-based VSLC developed with

a lookup table by an average increase of 10%, while feature extraction methods (Coarse and Tile) coding showed a slightly quicker learning rate.

To resolve the limitation of state representation and state-action space explosion, Greguri´c et al. (2020) introduced a Deep Q-Network to mimic the Q-function, and a unique learning technique was presented for the VSLC application with the potential to monitor vehicles at the microscopic level. The suggested reward function directed learning in the direction of reward enhancement and the avoidance of oscillation between successive speed limits.

Table 4 summarizes the RL-based DSH methods for HDVs.

**Table 4. RL-based DSH Methods for HDVs**

| Author | Year | Control algorithms | Metrics | Results |
|---|---|---|---|---|
| **Zhu and Ukkusuri** | 2014 | Reinforcement Markov Average Reward Technique (R-MART) | Mobility (travel time), environmental impact | Reduced TTS by 18% and almost 20% less CO2 emissions compared to the case without VSL. |
| **Walraven et al.** | 2016 | Q-Learning | Mobility (speed variance, travel time) | A decrease in TTS by approximately 30%. |
| **Li et al.** | 2017 | Q-Learning (kNN-TD) | Mobility (travel time) | The QL-VSL approach significantly outperformed the feedback-based with an improvement of TTT up to 21.84%. |
| **Kušić et al.** | 2018 | Q-Learning linear approximation | Mobility (travel time) | Function approximation methods outperform RL-based VSLC by an average improvement of 10 %, where feature extraction methods (Coarse and Tile) coding showed a slightly faster learning rate. |
| **Greguri´c et al.** | 2020 | Deep QL | Mobility (throughput), safety | Increased the average mainline speed and reduce traffic density. The oscillations between the posted speed limits and the measured speeds were prevented. |
| **Li et al.** | 2020 | Q-Learning | Safety | Reduced the crash risks by 19.4% while only increase the TTT by 1.5%; perform well under lower compliance rate with continuous online learning. |

To increase the throughput of a bottleneck after the San Francisco-Oakland Bay Bridge, Vinitsky et al. (2018) created unique control strategies for AVs. A two-stage bottleneck forms where four lanes drop to two and subsequently to one lane, a new library for applying DRL to traffic micro-simulators. First, they described the bottleneck's uncontrolled inflow-outflow curve and brought an inflow of AVs with Lagrangian control. They constructed a parametrization of the controller with per-lane changing speed limits to manage the flow. It showed that a 10% penetration rate may increase the bottleneck's throughput by 200 vehicles per hour, or 25% at high inflows. Lastly, the controller offered equivalent performance to additional ramp metering by comparing to feedback ramp metering.

For differential variable speed limit (DVSL) control, a DRL model was created by Wu et al. (2020), allowing for the imposition of dynamic and unique speed limits in various lanes. The suggested model learned a large number of discrete speed limits in a continuous action space using new actor-critic architecture. The DVSL controller was trained using a variety of reward signals, including total travel time, bottleneck speed, emergency braking, and vehicle emissions. The simulation findings demonstrated that the DRL-based DVSL control approach was capable of enhancing the freeway's safety, efficiency, and environmental friendliness. The generalization of the controllers with various driving behavior features allowed for observation of the DRL agent's resilience.

Seliman et al. (2020) provided a real-time, optimum control system to assist CAVs in a lane-drop location on a motorway (e.g. work zones). The Deep Q-Network (DQN) was used to identify the driving speed and lane change with the least time delay. The agent was trained using VISSIM. The performance was compared to that of a human-driven vehicle without intelligent control in terms of travel time. It showed how DQN-RL can help the CAV traverse the lane drop location wisely. In particular, the travel time decreased by almost 96% compared to the basic case. Further tests of the agent's resilience were conducted. The mean and standard deviation of the travel time reduction were around 31% and 61%, respectively.

Ko et al. (2020) suggested speed harmonization and merging control using CAVs. They used two deep Q networks correspondingly to save fuel and reduce traffic congestion. They also analyzed the trained Q-networks under various scenarios in terms of vehicle arrival rates and CAV market penetration rates. In comparison to the late merge control without speed harmonization, the simulation results showed that the suggested technique enhanced the mixed traffic flow by boosting throughput up to 30% and lowering fuel consumption up to 20%.

Vrbanić et al. (2021) integrated the two-step Temporal Difference target with the Q-Learning algorithm to improve the algorithm's ability for mixed traffic flows. Analyzing various CAV penetration rates, the outcomes were compared with a rule-based VSL and the no-control situation. The findings demonstrated that Q-Learning can adapt to changing

penetration rates and learn the policy while reducing Total Travel Time and Mean Travel Time. There was further evidence that the unnecessity of separate VSL as the penetration rate rises.

Xiao et al. (2022) explored the mainline VSL adjustment of the upstream off-ramp under the CV environment based on Q-learning. Three schemes, free control, rule-based, and Q-learning were designed by Python and VISSIM. The findings showed that mainline dynamic VSL adjustment of off-ramp upstream based on Q-learning algorithm performed well. The findings may offer useful information for reducing traffic congestion and managing traffic flow in the context of CAVs.

Gregurić et al. (2022) introduced spatially dynamic speed restriction zones. An innovative traffic state representation based on a series of sequential matrices that encode each vehicle's position and speed on the controlled road during the control time step was necessary for the spatial layout of speed restriction zones. The Deep Deterministic Policy Gradient (DDPG) architecture was used to calculate the actions for each proposed VSL strategy. ConvLSTM layers, which integrated Convolution and Long Short-Term Memory (LSTM), as well as Convolution and Fully Connected layers, were included in the DDPG learning models. The proposed VSL techniques outperformed baseline and static speed limit zones in terms of throughput (no-control and Simple Proportional Speed Controller algorithm). They simultaneously boosted the average headway while just slightly increasing the amount of severe braking.

Table 5 summarizes the RL-based DSH methods for CAVs.

**Table 5. RL-based DSH Methods for CAVs**

| **Author** | Year | Model | Scenarios | MPRs | Comparisons | Results |
|---|---|---|---|---|---|---|
| **Vinitsky et al.** | 2018 | TRPO (GRU) | 941m includes lane drop, 1500 veh/h | 100% AVs | No control, feedback ramp metering | 25% higher outflow for medium traffic demand, no control performed better for lower traffic demand. |
| **Wu et al.** | 2020 | Priority Replay-DDPG | 875.51m includes on-ramp and off-ramp | 100% CAVs | No control, Q-Learning, DQN, Actor-Critic | 8.1% lower ATT in incidents scenario, 5.8% lower in scenarios without incidents. DQN is safer than DDPG, Actor-Critic is the worst. |
| **Seliman et al.** | 2020 | DQN | 2km includes lane-drop | 100% CAVs | No control | The reduction in travel time is around 96 %. |
| Ko et al. | 2020 | DQN | 3.4 km includes two control areas, | Mixed CAVs, CVs, and | Late merge control without speed harmonization | Increased the 30% throughput and reducing the 20% fuel consumption. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | 2000-4000 veh/h | HDVs | | |
| **Vrbanić et al.** | 2021 | Q-Learning | 8 km includes two on-ramps and one off-ramp | 0-100% CAVs | No control, rule-based | Improved the TTT and ATT. The results are most obvious in low MPRs. |
| **Xiao et al.** | 2022 | Q-Learning | 1.6 km, exit to intersection is 300 m | 100% CAVs | No control, rule-based | 37.60% better travel efficiency, 27.49% lower average delay. |
| **Gregurić et al.** | 2022 | DDPG (ConvLSTM) | 8 km includes two on-ramps and one off-ramp | 10 % buses, 10 % trucks, 80 % personal vehicles | No-control, Simple Proportional Speed Controller | Higher overall throughput compared to static speed limit zones, the average headway is increased. |

# Chapter 3. Deep Reinforcement Learning-Based Dynamic Speed Harmonization at Recurrent Bottleneck

## 3.1 Introduction

This chapter investigates the effects of dynamic speed harmonization in mixed traffic flow involving Human-Driven Vehicles (HDVs) and CAVs on the freeway. To be more specific, (a) it conducts a comprehensive review of DSH on HDVs and CAVs, respectively, and state-of-the-art methodology to implement this technique; (b) it develops a DSH strategy based on deep reinforcement learning, and better understands how CAVs can improve operational performance; (c) considering that CAVs will coexist with HDVs for a long time, a series of numerical experiments are conducted under different Market Penetration Rates (MPRs) through various simulated scenarios; (d) a holistic performance evaluation framework is formulated to evaluate the impacts on MOEs, and potential interactions between MOEs are explored.

The remainder of this chapter is organized as follows. The methodology is introduced in Section 3.2. The simulation environment and experimental settings are described in Section 3.3. The results and discussions are illustrated in Section 3.4.

## 3.2 Methodology

### 3.2.1 DRL framework of DSH

The DSH issue can be regarded as a Markov decision process (MDP), which is also the core of DRL. It is composed of ($S$, $A$, $P$, $R$), where $S$ represents a set of states $s$, $A$ denotes a set of actions $a$, $P$ is the transition probability from the last step $a$ in $s$ based on policy $\pi$ that leads to next state $s'$, and $R$ is the immediate reward with a discount factor $\gamma$ from 0 to 1 for the agent after the transition. The objective of the agent is to maximize the cumulative rewards that interact with the environment. Figure 3 shows how DRL is implemented in DSH. In the DRL framework, the agent, which is the VSL controller managed by a DRL algorithm, receives the state represented by traffic parameters in the downstream congestion area. The agent takes actions that are variable speed limits and returns them to the environment. Then, the environment sends a new state decided by a defined policy mapping from the previous state and action. The feedback from the environment to the agent is to maximize the cumulative rewards represented by performance indicators. The configurations of environment, agent, state, action, and reward function are as follows:

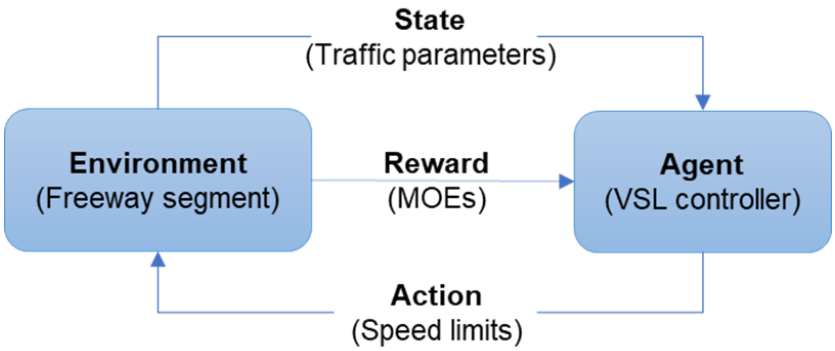


**Figure 3. The Control Scheme of DRL in DSH**

**Environment and agent:** The environment is a freeway mainline segment that includes a weaving area between an on- and off-ramp, and recurrent congestion occurs in the weaving area. More road configuration details will be described in Section 4.3. The agent is the VSL controller, which can interact with CAVs by sending and displaying differential speed limits for each lane. This is easy to implement in CAVs environment with V2X technology. Due to the complexity that regards each CAV as a VSL controller, this study mainly focuses on the single-agent system.

**State:** It consists of traffic states collected by detectors, and can be transformed into a vector or an image as input for deep neural networks in DRL. State representation is a complicated issue due to the difficulty to describe the state transition process. It usually considers traffic parameters such as density, average speed, or traffic volume of the congestion area. In this research, immediate occupancy rates of each lane in the upstream mainline, on-ramp, and downstream weaving area (Li et al. 2017) by specific detectors are collected. The occupancy rates of detectors are used as the input of the states.

**Action:** This study uses differential speed limits for each lane. The actions are set as discrete values attributed from 0 to $A$-$1$, where $A$ is the action space. The speed limit equals $V_0 + M(A\text{-}1)$, where $V_0$ is the minimum speed limit, and $M$ is an integer increment and usually takes 5 or 10 mph. Given that the shock-wave occurs by the encounter of two traffic states under different speeds, it contends that this issue can be efficiently resolved by maintaining a constant speed limit. The controller dynamically configures the speed limit based on different traffic congestion states, which is how the DSH implemented.

**Reward function:** Similar to the objective function in optimization, the goal of DRL is to maximize the reward. Defining an effective reward function is a tricky issue, and there is currently no unified statement to determine which reward is optimal, and the choice of reward function will largely affect the performance of DRL. It often depends on the MOEs the study wants to improve. Generally, real-time indicators, such as total travel time, number of emergency braking, emissions or fuel consumption, are selected. Although a complex reward function can

more accurately reflect the traffic state, an overly complex reward will make the algorithm difficult to converge. Considering that the primary goal of DSH is to ensure safety, this research defines a safety-oriented function as $r$=- $\theta_t$, where $\theta_t$ is the cumulative emergency deceleration that is above the default value (4.5 m/s$^2$) in the simulation, in which it reckons that the risk of a collision increases as the cumulative value increases. Here the value is considered as a threshold for the reward function. During the harmonization, the vehicle can follow this deceleration rate or not according to the traffic congestion state, as long as the final reward reaches the optimal point.

3.2.2 DRL Algorithm of VSL Controller

The DRL algorithm selected to control the speed limit is Deep Deterministic Policy Gradients (DDPG) (Lillicrap et al. 2015). It is a model-free, combination of off-policy Q-learning and on-policy gradients-based algorithm. The action dimension could be an exponential increase if each lane is set differential speed limits, and the discrete Q-learning method may encounter a space explosion problem. Therefore, an algorithm capable of handling continuous action without enumerating all values should be considered. The main advantage of DDPG is its ability to choose continuous actions, which allows for a lot of flexibility when developing different DSH strategies (Gregurić, Kušić, and Ivanjko 2022). The policy is built on an Actor-Critic framework that provides both value- and policy-based function approximations in the same deep neural networks. The actor takes actions, and then the critic evaluates the policy $\pi$ characterized by the actor and predicts the target $Q$-value function:

$$Q(s,a) = r(s,a) + \gamma Q(s',a) \tag{1}$$

Where $s$ denotes the current state, $s'$ denotes the next state, $r$ is the reward function, and $\gamma$ is the discount factor which is set to 0.9. The objective of the actor is to maximize the $Q$ predicted by the critic through a trial-and-error interaction.

The DDPG utilizes a stochastic policy for action exploration, but the target policy for action is deterministic. It updates the parameters of the actor $\theta^a$ and the critic $\theta^c$ in a bi-level optimization pattern. The critic uses the Adam optimizer to reduce its loss, which is defined as the difference between two sides of the *Bellman equation*. The weights of $\theta^c$ and $\theta^a$ can be updated with the gradients in the loss function $L$ ($Q$, $\theta^c$) and $L$ ($\pi$, $\theta^a$), which can be expressed as Temporal Difference (TD) error:

$$\theta^c = argmin_{\theta^c} Q_{\theta^c}(s,a) - (r(s,a) + Q_{\theta^c}(s', \pi(s'))) \tag{2}$$

$$L(Q,\theta^c) = \frac{\sum_1^N (\theta^c)^2}{N_{\text{sample}}} \tag{3}$$

$$L(\pi,\theta^a) = -\frac{\sum_1^N Q(s,\pi_{\theta^a}(s))}{N_{\text{sample}}} \tag{4}$$

The weights of $\theta^a$ is updated with the deterministic policy gradient:

$$\nabla_{\theta^a} = \frac{1}{N_{\text{sample}}} \sum_1^N \nabla_a Q_{\theta^c}(s,a) \mid_{a=\pi_{\theta^a}(s)} \nabla_{\theta^a} \pi_{\theta^a}(s) \tag{5}$$

The target actor and critic models are softy replaced by using $\tau$=0.01. Experience replay is used to store useful experience and discard useless experience through a reply memory, and the memory capacity is set to 30000 for the agent to sample. This research uses light-weight neural networks for both the actor and the critic with two layers, and each layer has 30 neurons. The batch size is set to 32, the learning rate of the actor is 0.0001 and the critic is 0.0002. The Adam optimizer is used to adapt the learning rate for each weight of the neural network. The actor $\theta^a$ and critic $\theta^c$ can be expressed by:

$$h_t^a = relu(W_1^a + b_1^a) \tag{6}$$
$$a_t = A * sigmoid(W_2^a h_t^a + b_2^a) \tag{7}$$
$$h_t^c = relu(W_s^c s_t + W_a^c a_t + b_1^c) \tag{8}$$
$$Q_t = W_2^c h_t^c + b_2^c \tag{9}$$

Where state dimension *S* is 10, action dimension *A* is 6, the parameters for each network is $W_1^a, W_s^c \in R^{30*10}$, $W_2^a, W_2^c \in R^{1*30}$, $b_1^a, b_1^c \in R^{30}$, $b_2^a, b_2^c \in R^1$, *relu* and *sigmoid* are utilized as activate function. All hyperparameters in this work were fine-tuned after many trials. The steps of the DDPG algorithm for DSH are summarized as follows:

Step 1: Initialize parameters and set target weights for the actor network $\theta^a$ and the critic network $\theta^c$. Then empty the replay buffer.

Step 2: Load the environment and observe state *s*. Repeat steps 2 to 6 until the episode reaches its maximum.

Step 3: During time steps in the simulation, explore action *a* based on the current policy $\pi$, and execute the noise decay.

Step 4: Select variable speed limits, receive reward *r* and new state *s'*. Store this MDP in the reply buffer.

Step 5: Randomly sample a batch of transitions from the reply buffer, computer the target function Q.

Step 6: Update the Q function by minimizing the *L* ($Q$, $\theta^c$), update the policy with the deterministic policy gradient $\nabla_{\theta^a}$

Step 7: Update target actor $\theta^a$ and critic $\theta^a$ networks by $\tau\theta^a + (1-\tau)\dot{\theta}^a$ and $\tau\theta^c + (1-\tau)\dot{\theta}^c$ until convergence.

## 3.3 Case study

### 3.3.1 Roadway Configuration and Traffic Scenario

A 1.4-mile freeway segment of I-80 northbound in District 4 of California, U.S. is selected. The length of the on- and off-ramp is 600 feet and 1100 feet, respectively. The weaving area between the on- and off-ramp is about 2100 feet, which is not abundant space, and congestion can easily occur during peak hours. The original speed limit for the mainline traffic is 65 mph, and 50 mph for both the on- and off-ramps, respectively.

Figure 4 shows the traffic states in the study area on one weekday, and the traffic direction of northbound is from northeast to southwest. The weaving area is located near the real changeable message sign. The traffic was still free-flow at 6 am, and there was light congestion starting at 7 am. It reached the worst state at 8 am, the average speed upstream was even lower than 40 mph, and the data obtained from the Caltrans Performance Measurement System (PeMS) showed that it was actually only 20 mph. The congestion was not able to dissipate until 9 am, and it returned to the normal state at 10 am. It can be found that though there is a changeable message sign in reality, the current measures cannot relieve congestion.

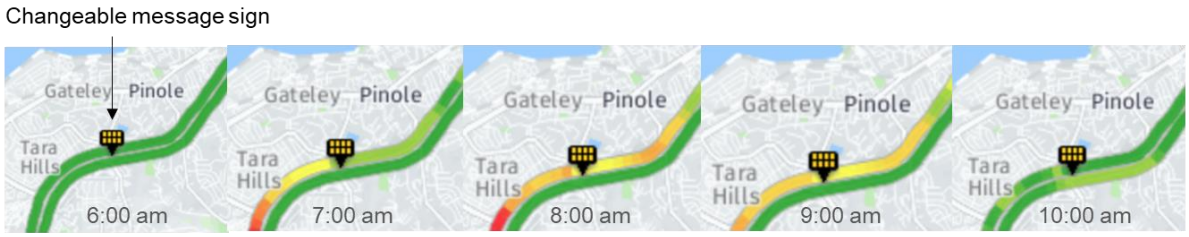


**Figure 4. Traffic States in the Study Area (Source: PeMS)**

### 3.3.2 Experimental Settings in the Simulation

The data source is from the PeMS for generating the traffic demand set by the detector data that were collected from the stations. The OpenStreetMap (OSM) is used to export the network in the simulation. Each simulation lasts for 4 h from 6 am to 10 am, and the demand is randomly generated for each round. There are three routes: mainline to mainline [4587, 4194, 4440, 4249], mainline to off-ramp [1529, 1398, 1480, 1416], and on-ramp to mainline [461, 771, 888, 744], in which the values in square brackets are the average value of the Poisson distribution for the demand. The proposed DSH algorithm is implemented in the Simulation of Urban Mobility (SUMO), which is an open-access simulation tool that provides an API - Traffic Control Interface (TraCI) package. To model the mixed flow, this study defines two types of vehicles. The Krauss car-following model is used to model HDVs. Given the compliance rate of

HDVs, the driver imperfection is set to 0.5, which means that half of HDVs will not comply with the display speed limits. The Intelligent Driver Model (IDM) car-following model is used to model CAVs. Considering the different features of mixed vehicles, the headway is set 1.1 s for HDVs and 0.9 s for CAVs, respectively. The values of all the parameters are adapted from Treiber, Hennecke, and Helbing (2000), Hua and Fan (2022). The maximum acceleration rate equals 0.73 $m/s^2$, comfortable deceleration rate equals 1.67 $m/s^2$, desired velocity equals the speed limit, acceleration exponent equals 4, linear jam gap equals 2 m, and non-linear jam gap equals 3 m. Due to the fact that this study mainly focuses on the mainline longitudinal control, the lang-changing model uses the default LC2013 in SUMO. To investigate the effects of DSH in mixed flow under different Market Penetration Rates (MPRs) of CAVs, the MPRs range from 0% to 100% with an increment of 25%. This study mainly utilizes the built-in parameters in SUMO, and more details related to the calibration of the micro-simulation model, involving the autonomous vehicles' adaptive cruise motion can be found in Silgu et al. (2021), Sadat and Celikoglu (2017), and Göksu et al. (2021).

Figure 5 demonstrates how DSH is implemented in SUMO. The traffic direction is from right to left, and the mainline section has 4 lanes and the weaving area has 5 lanes. The state dimension is 10, including 4 lanes in the upstream mainline, 5 lanes in the downstream weaving area, and 1 lane in the on-ramp. The OSM and traffic demand are the inputs of the simulation. The same environment is built in SUMO, and there are a series of induction loop detectors (red arrow) distributed through the freeway segment. The VSL controller will select a speed limit based on the current policy, and the action dimension is 6 from 50 to 75 mph with an increment of 5 mph. Then, it executes the DDPG algorithm, and the output by TraCI will update the reward to the optimal value. The detectors also provide the average speed at the upstream, midstream, and downstream of the weaving area.

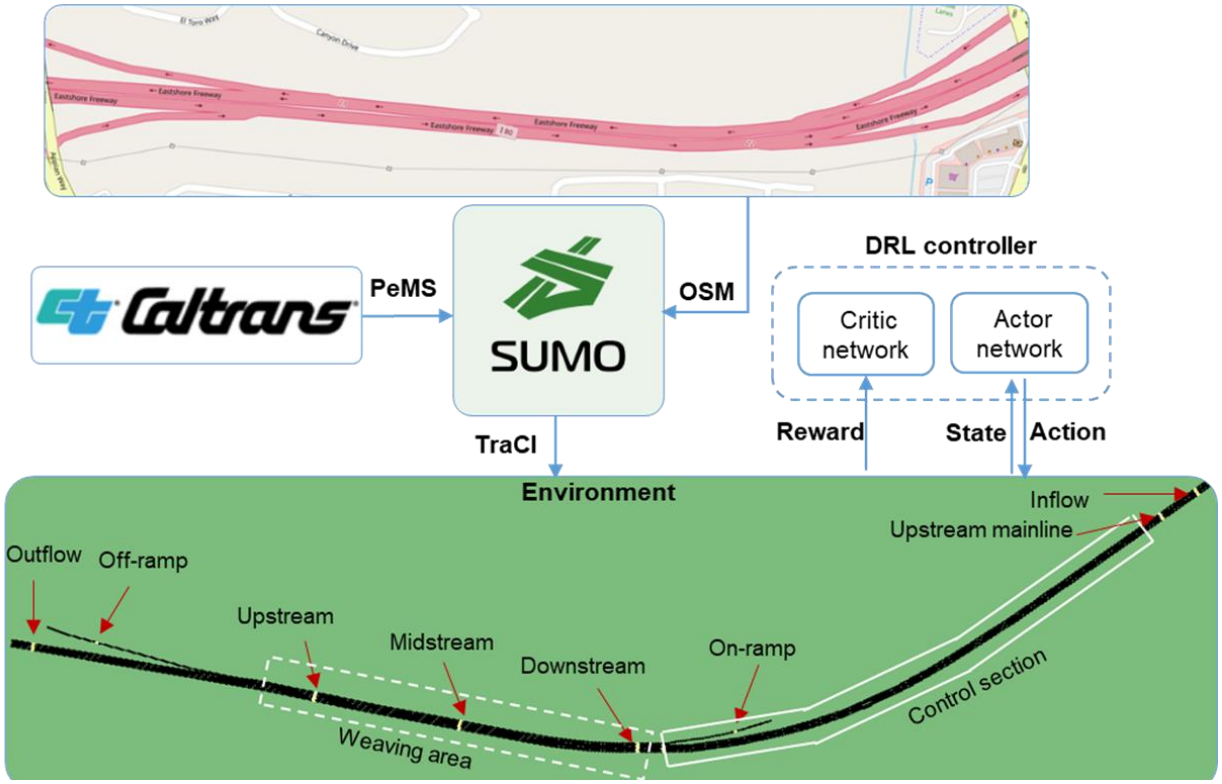


**Figure 5. Environment Configuration of DSH in SUMO**

## 3.4 Results and Discussions

### 3.4.1 Learning Process of DRL Agents

Figure 6 shows the cumulative rewards under different MPRs that are trained by 300 episodes. The reward achieved by no control (dash line) is expressed as the average value with various level of MPRs. Except in the case of pure HDV (0% MPR), the others' starting values are relatively low, but the learning curves afterward are roughly the same. It can be found that the VSL controller hardly learns useful experience before 130 episodes, and it hovers in a small range. The cumulative reward reaches its first peak around the 135th episode as more experience is learned. Later, more useful information is extracted, reaching the highest value in the 180th episode. After 200 episodes, the learning process has gradually stabilized and the reward has been greatly improved compared to the beginning of the implementation of DSH. More details about the difference under various MPRs will be discussed subsequently. Compared with DQN, the DDPG algorithm based on actor-critic architecture is more difficult to converge. It may not learn anything at the beginning, but it still can reach the optimal value in the end.

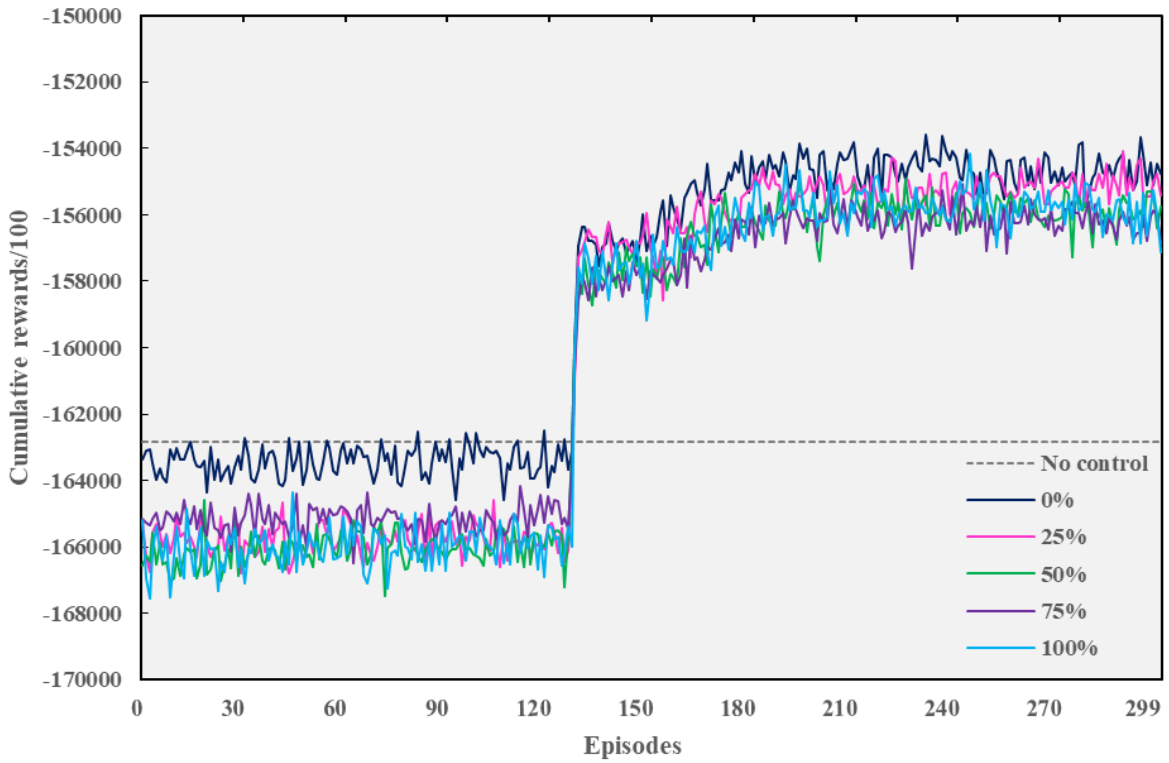


**Figure 6. Cumulative Rewards During the Learning Process**

3.4.2 Synergistic Evaluation of MOEs

The application of CAVs cannot only be assessed from a single perspective, and a comprehensive MOEs evaluation can better understand co-benefits and trade-offs between indicators. Each set of numerical experiments compares the average operational performances over 20 test episodes. The consideration of computing time is also an important part to reflect the efficiency of the algorithm. The average time for each episode under different MPRs before DSH is 153 s, and the time cost after DSH is only 156.8 s. Table 6 shows the operational performances under different MPRs of CAVs. MOEs mainly consider three aspects: a) safety issues represented by accumulated emergency deceleration; b) mobility issues represented by Mean Travel Time (MTT); and c) environmental sustainability issues represented by greenhouse gas (GHG) with $CO_2$, harmful gas with CO, and fuel consumption. The comparison is based on no control without CAVs under each indicator.

**Table 6. Operational Performances Between DSH and No Control**

| MOEs | Indicators | Control | Improvement by MPRs (%) | | | | |
|---|---|---|---|---|---|---|---|
| | | | 0% | 25% | 50% | 75% | 100% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Safety | Accumulated emergency deceleration($m/s^2$) ($10^7$) | No | 1.63 | 0.14% | -0.11% | -0.08% | 0.41% |
| | | DSH | 5.02% | 4.96% | 4.46% | 4.24% | 4.96% |
| Mobility | Mean Travel Time(s) | No | 82.48 | 0.18% | -0.05% | 0.12% | 0.75% |
| | | DSH | 5.21% | 4.92% | 4.50% | 4.55% | 4.62% |
| Environmental Sustainability | $CO_2$ (kg): | No | 10009.3 | 5.74% | 11.36% | 16.70% | 21.45% |
| | | DSH | -9.51% | -1.76% | 5.12% | 10.84% | 15.39% |
| | CO (kg): | No | 151.96 | 10.47% | 20.46% | 29.89% | 36.27% |
| | | DSH | -20.22% | -7.24% | 4.23% | 13.61% | 21.17% |
| | Fuel consumption (L): | No | 4302.53 | 5.74% | 11.36% | 16.70% | 21.44% |
| | | DSH | -9.51% | -1.76% | 5.12% | 10.84% | 15.39% |

In terms of safety, there is no significant improvement with or without control in the early stages (25% MPR) of CAVs implementation. There is a slight risk increase associated with more CAVs, and this phenomenon persists to higher MPRs. A possible reason is that CAVs prefer to adopting a more aggressive way considering smaller headway when interacting with HDVs. Maybe this phenomenon can be mitigated and performance can be further improved by introducing the platoon control of CAVs. This interaction is not obvious in lower MPRs, but is intensified with the increase of CAVs, fortunately, the increased risk is modest. When the road network is full of CAVs (100% MPR), the safety is improved again (0.41% for no control and 4.96% for DSH). Moreover, applying DSH can mitigate the detrimental effect of low CAV shares on safety (nearly 5% improvement). The co-benefits after DSH can be found in the mobility. It can be seen that, compared with no control, MTT has been improved to varying degrees (all above 4%). MTT increases slightly at 50% MPR, but it is still within an acceptable range.

From the environmental sustainability perspective, $CO_2$ emissions continue to decrease with the increase of CAVs under no control, and an improvement of 21.45% can be achieved at 100% CAVs. For DSH, the emission of greenhouse gases increases under a 25% MPR. However, when more CAVs are deployed, this situation is alleviated and finally reaches an enhancement level of 15.39%. The same changes can be found in fuel consumption, which reveals a clear connection with $CO_2$ emissions. In fact, using either one can reflect the impact on the environment. For the emission of harmful gases, CO is mainly selected due to the proportion of CO exceeding 96% in the simulation. This is much higher than other gases such as hydrocarbons, $NO_X$, and $PM_X$. Regarding the performance before and after DSH under different MPRs, the change of the harmful gas is similar to that of greenhouse gas. Although the DSH has an overall positive impact on safety and mobility, there are trade-offs with environmental sustainability. Actually, a certain difference can be observed between the outflow and inflow during the congestion phase. It is known that the introduction of CAVs can increase the throughput, but the DSH has learnt to assign lower speed limits during the congestion to ensure the safety. Therefore, there will be a balance between them. However, the impact of the fluctuations will eventually be offset, and the overall performance will be improved.

### 3.4.3 Spatiotemporal Variations of Bottleneck Speed

Figure 7 demonstrates the spatiotemporal changes of the average speed at the weaving area after DSH. The **U** denotes the upstream section (which is close to the off-ramp), the **M** represents the midstream section, and the **D** is the downstream section that is near the on-ramp. For each case, there are three sections on the vertical axis, and each section is divided into 5 lanes, with a total of 15 rows. The upper of the lane, the closer to the ramp, and the right lane is defined as the outermost lane, and vice versa. The 4-hour simulation time is divided into 48 intervals every 5 minutes. In all cases, the VSL controller has learned to assign differential speed limits to each lane, so that the vehicle can travel within the specified speed.

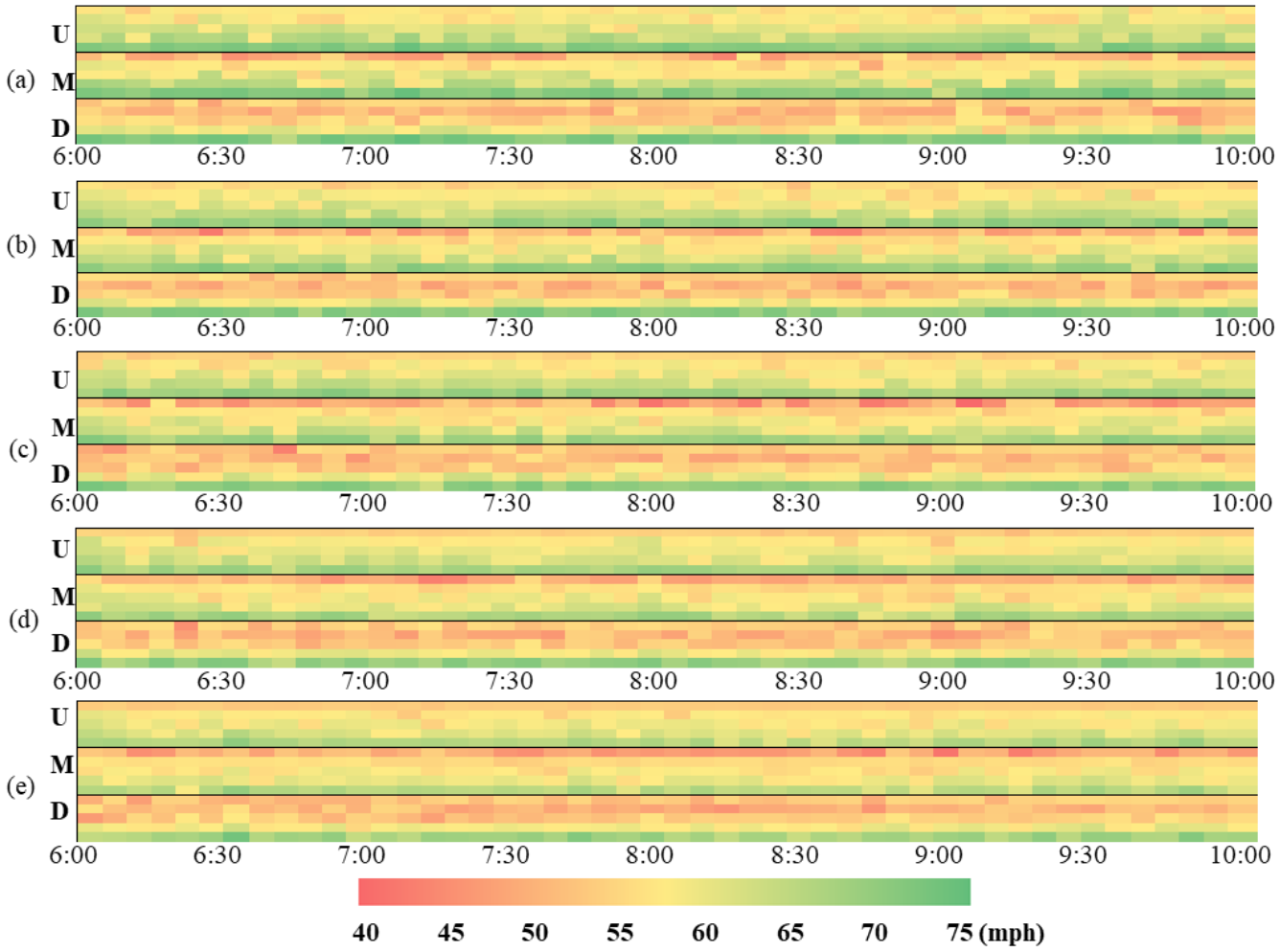


**Figure 7. Speed Variations by Time-of-Day Under Different MPRs. (a) 0%, (b) 25%, (c) 50%, (d) 75%, (e) 100%**

With regard to the upstream section, there is a small fluctuation in the right two lanes under low MPRs (<50%) at intervals during the congestion formation phase (6-7 am). This is

because the outer side needs to go down the ramp, and the disturbance caused by the deceleration will have an impact on the traffic flow stability. At higher MPRs, though, the speed variations become smaller. These two phenomena are also observed when congestion occurs (7-9 am) and dissipates (9-10 am), indicating that more CAVs powered by DSH can play a role in reducing speed variability during the deceleration process. For the middle two lanes, there is a slight speed oscillation due to the influence of the outer lanes, and this instability increases with the increase of MPRs. However, vehicles are still able to operate at a high speed. For the innermost lane, the agent always gives the highest speed limit regardless of the MPRs, and the highest speed can reach 71.9 mph. In fact, at any position of the weaving area, the innermost lane is not affected by other lanes owing to the differential speed limits, which means that the agent can recognize the leftmost lane and set it as the overtaking lane.

Concerning the midstream section, the right lane always suffers more severe speed oscillations due to the insufficient length of the weaving area, but the DSH can dampen speed variations under higher MPRs (>75%). A minimum of 40 mph is reached at 50% MPR, but it is still better than the original 20 mph before the DSH. The middle and the left lanes perform similarly to the situations in the upstream section.

As for the downstream section, the outer two lanes always show relatively larger speed variations during the congestion formation phase. Because this location is closest to the on-ramp, vehicles need to accelerate to match the mainline traffic, but the insertion of vehicles will disturb the adjacent lanes and the continuous lane change will affect the middle lane as well until it moves to the leftmost lane. However, during the congestion dissipation phase, the speed variations in the outer lane are alleviated with the increase of MPRs and the middle lane also benefits. During the congestion phase, the left two lanes are hardly affected all the way. The right three lanes encounter the most unstable situation compared to the upstream condition, indicating that the merging process causes more impact than the diverging process.

### 3.4.4 Sensitivity Analysis of CAVs Performance

While this study is based on the simulation, the real-world traffic flow may encounter more randomness and uncertainty. However, it is impractical to test in reality considering the cost. The operability of simulation allows us to explore the effects of parameters on vehicle control in different environments. The improvement of CAVs over MOEs is mainly attributed to the reduction of headway. The relatively conservative values of 1.1s for HDVs and 0.9s for CAVs are chosen to avoid the degradation of safety due to the large variation in headway, which makes the difference less pronounced at different MPRs. In fact, the headway can be reduced to 0.6s for higher-level CAVs. To investigate the sensitivity of MOEs to headway, this section only changes the headway from 0.9s to 0.6s while other variables remain fixed. Figure 8 illustrates the MOEs before and after DSH under the smaller headway. The three most representative metrics are selected: accumulated emergency decelerations for safety, MTT for mobility, and $CO_2$ emissions for sustainability. The colored bars denote the MOEs under the new headway. It is worth noting that the percentage changes in each MPR show the corresponding improvement before and after DSH compared to 0.9s.

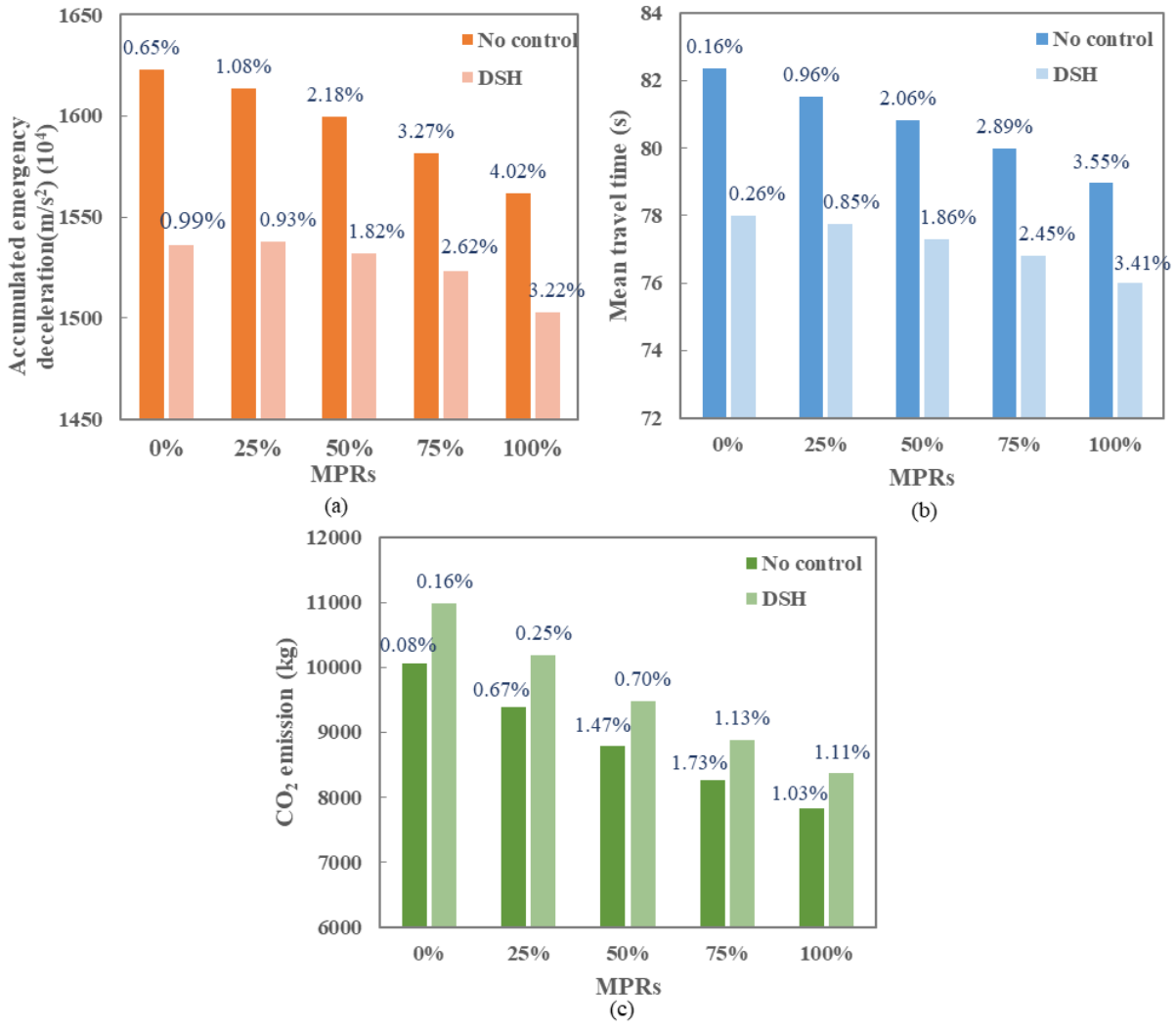


**Figure 8. MOEs Changes and Comparisons Under the Smaller Headway. (a) Safety, (b) Mobility, (c) Sustainability**

The change in the histogram shows that under the new headway, regardless of the before and after control, the trend in each indicator displays a roughly linear decrease with the increase in MPRs and is more stable. For the original headway, though, the changes in safety and mobility are not uniform and not distinct among MPRs, except for sustainability. This indicates that the performance of higher-level CAVs is more influenced by MPRs and the whole change process will be smoother. Indeed, when the gap between the headway of HDVs and CAVs increases, MPRs have more effects on MOEs, but too large a gap may lead to disturbances in the overall traffic flow and affect stability. The percentage changes show that the MOEs are further promoted by the higher-level CAVs compared to the original headway, and the safety improvements are approximately the same before and after control at lower MPRs (< 50%), but the enhancements before control are more significant at higher MPRs. The performance after DSH may be close to the limit, which is difficult to further improve. The same phenomenon occurs for the mobility, while the improvements in sustainability are not as obvious as them at higher MPRs. The above suggests that DSH powered by higher levels of CAVs can play a role in amplifying most MOEs depending on the pace of technology development and deployment.

# Chapter 4. Safety-oriented Dynamic Speed Harmonization at Nonrecurrent Bottleneck

## 4.1 Introduction

Optimizing traffic flow within congested areas poses a critical challenge in the transportation field. Congestion at the bottleneck emerges when the upstream demand exceeds the design capacity of the downstream roadway segment (Ghiasi et al., 2019), leading to heightened collision risks, prolonged travel times, and reduced fuel efficiency. This situation can be exacerbated by interruptions in traffic flow or sudden fluctuations in traffic volume. The Federal Highway Administration reported various types of congestion in the United States by their causes, revealing that 55% of congestion is attributed to nonrecurrent situations, including 25% incidents, 15% adverse weather, 10% work zones, and 5% special events. The remaining 45% of congestion falls under the recurrent scenarios, comprising 40% daily bottlenecks, and 5% poor signal timing.

Constructing additional road infrastructure is not always a practical solution due to the potential for induced traffic demand that could create an undesirable cycle. More effective and efficient utilization of existing infrastructure with active traffic management is an alternative approach. DSH, also known as VSL, is a widely employed control strategy to dampen traffic oscillations and smoothen traffic speed based on current traffic conditions (Ma et al., 2016). DSH primarily aims to enhance safety (Lu et al., 2010) by coordinating speeds and reducing their variations, mitigating stop-and-go motions and thus lessening crash frequency. DSH achieves this by adjusting speed limits displayed on variable message signs to indirectly manage traffic flow (Papageorgiou et al., 2008).

Nevertheless, the conventional DSH strategy confronts certain challenges. Firstly, its effectiveness is constrained by the low compliance rate of drivers. CAVs can be leveraged to facilitate coordinated decisions by employing the Vehicle-to-Everything technology (Talebpour et al., 2013). When assessing CAV deployment's performance, three typical MOEs, including Safety, Mobility, and Environmental Sustainability, are commonly used and evaluated through various metrics. Most studies conducted in the past focused on one or two MOEs, and only a few assessed all three MOEs simultaneously. Additionally, Tian et al. (2018) indicated that co-benefits or trade-offs between metrics can be explored. Therefore, a holistic framework needs to be established to evaluate MOE interactions (Hua and Fan, 2023). Figure 9 illustrates the synergistic performance evaluation framework with respective metrics in this research. Another limitation is the necessity of setting differential speed limits across all lanes for safety reasons (Khondaker and Kattan, 2015), an aspect less explored in CAV environments. RL and classical control models are two distinct approaches to implementing the vehicle speed control. Both methods aim to achieve a specific objective such as optimizing traffic flow, or reducing

congestion. Also, they involve a feedback loop where the system's state is observed, which allows the controller to take actions in response to changing conditions. While traditional methods are well-established in the speed harmonization problem, RL offers advantages in model dependency and adaptability. Classical methods often rely on a precise prediction model and require traffic dynamics with the fundamental diagram to tune the controller parameters (Kušić et al., 2020), which mitigates data efficiency. RL can interact with changing environments without explicit traffic dynamics and achieve comparable performance with less prior knowledge. In addition, RL can learn optimal control policies by exploration and exploitation of state-action space over time, where classical methods might struggle due to the curse of dimensionality. RL is particularly beneficial in terms of adaptability and can generalize learned policies to novel and unseen situations, making it excel in complex, dynamic, and uncertain scenarios.

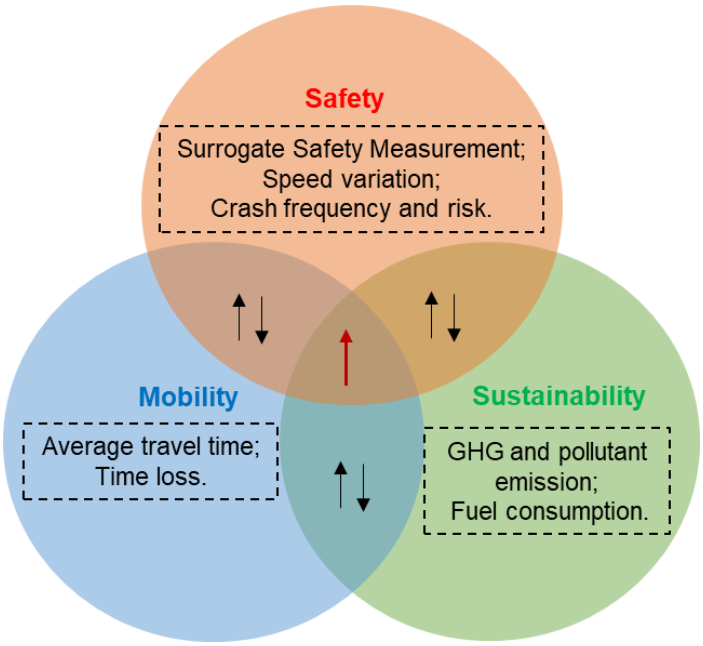


**Figure 9. Synergistic Performance Evaluation Framework**

The overall objective of this chapter can be highlighted as follows: a) given the frequency of nonrecurrent congestion caused by incidents, a safety-oriented DSH strategy is developed to optimize traffic flows. b) a DRL-based controller is designed to dampen speed oscillations and smoothen traffic. c) a comprehensive evaluation based on a variety of MOEs is conducted and potential interactions between multiple metrics can be found. d) experiments are carried out under different MPRs of CAVs, and reveal that CAVs powered by DSH can further improve operational performance. e) a series of sensitivity analyses in changing simulated scenarios are conducted to help gain deeper insights into the adaptability and generalization of the DRL agent.

The remainder of this chapter is organized as follows. Section 4.2 introduces the proposed DRL method and safety-oriented MOEs. Section 4.3 describes the experimental settings and microscopic simulation model. Section 4.4 presents the numerical results and discussions.

## 4.2 Methodology

### 4.2.1 DRL Scheme

DRL is a branch of machine learning that combines RL with deep learning to handle complex input spaces. RL is a paradigm where an agent learns to make decisions by interacting with an environment through trial and error. The agent takes actions, receives rewards or punishments as feedback, and adjusts to maximize cumulative reward over time. Deep learning involves the use of neural networks, which are composed of multiple layers of neurons. These networks can automatically learn hierarchical representations of state, making them well-suited for processing high-dimensional input spaces. However, traditional RL algorithms often encounter challenges when dealing with tasks characterized by a multitude of states or continuous action spaces. DRL addresses this limitation by using neural networks to approximate value functions or policies, allowing for more efficient and scalable learning in intricate and extensive environments. In contrast to traditional RL involving hand-engineering features or state representations, DRL focuses on end-to-end learning, where the entire learning process is integrated into a single, unified system.

In DRL, neural networks are used to approximate value function or policy. The value function estimates the expected cumulative reward of being in a certain state-action space. Policy determines the mapping from states to actions. DRL algorithms must balance exploration (discover new actions) and exploitation (choosing actions based on current knowledge), which is crucial for optimal learning to maximize reward. The configurations for the environment, agent, state, action, and reward functions are the same as chapter 3. The algorithm chosen to achieve the optimal speed limit is also the DDPG. The target actor and critic models are replaced by soft updating factor $\tau$=0.01. The agent is trained after 300 episodes to reach a stable and maximal reward. A reply memory with a capacity of 20000 is used for the experience replay, which is utilized to save useful experiences and dismiss useless experiences. This research employs two-layer, lightweight neural networks with 32 neurons per layer for the actor and critic to improve the efficiency of the learning process. The batch size is set to 32, the learning rate of the actor is 0.0001 and the critic is 0.0002. All hyperparameters are fine-tuned after many trials.

### 4.2.2 Surrogate Safety Measurement (SSM)

Due to the lack of mass deployment of CAVs in the real world and considering the scarcity of crashes, it is not yet feasible to perform safety assessments using historical crash data of mixed traffic flow. However, traffic conflicts occur considerably more frequently than car crashes. In light of this, SSM derived from traffic conflicts is proposed. Traffic conflicts are perceptible non-collision events that raise the collision risk if moving vehicles do not divert from their intended path. Traffic conflicts and crashes are thought to be related, and indicators used to

quantify the safety effects can be considered as SSM (Wang et al., 2022). Currently, most CAV studies involving SSM are conducted in a microscopic simulation environment.

In the simulation, a vehicle can be equipped with an SSM Device that logs the conflicts of participants. The criteria to qualify an encounter as a conflict (if their measurements exceed a threshold) can be customized by several generic parameters. The most typical SSM is Time-to-Collision (TTC), which was first proposed by Hayward (1972) and defined as "*time that remains until a crash between two vehicles would have occurred if the crash course and speed difference are maintained."* It is given as:

$$TTC = \frac{Space\ gap}{Speed\ difference} \tag{10}$$

Where the space gap is the distance between the following and leading vehicle minus the vehicle length (usually 5 meters). To connect the traffic conflict to the crash directly, the collision probability is used to evaluate the possible risk by comparing the calculated TTC and the threshold TTC (default value is 3s) shown as:

$$Collision\ probability = \frac{Number\ of\ TTC < Threshold\ TTC}{Total\ number\ of\ TTC} \tag{11}$$

Different from gauging time proximity, Cooper and Ferguson (1976) proposed Deceleration Rate to Avoid the Crash (DRAC) to measure the severity of the conflict. It is defined as:

$$DRAC = \frac{0.5 * Speed\ difference^2}{Speed\ gap} \tag{12}$$

Where the default value of DRAC is 3 m/s$^2$, and can be recorded at the time point of the maximal value as $DRAC_{max}$. A modified variant called MDRAC considering a Perception-Reaction-Time (PRT, default value: 1s) is defined as:

$$MDRAC = \frac{0.5 * Speed\ difference}{TTC - PRT} \tag{13}$$

It is noted that some SSMs only apply to a specific encounter or imply different calculation procedures for different encounters. This research is related to the freeway mainline control, and only lead/follow situations where vehicles are passing the same sequence of lanes before and after the conflict point is considered.

## 4.3 Experimental Settings

The testbed selects a busy northbound freeway segment located on I-80 in District 4 of California, U.S. It is a 1.4 miles long segment including a weaving area between the on- and off-ramps, and the length of the on- and off-ramp is 600 feet and 1100 feet. The traffic direction is from East to West, and the upstream mainline section has 4 lanes and the downstream weaving area has 5 lanes. The speed limit for the mainline is 65 mph, and 50 mph for both on- and off-ramps. The Caltrans Performance Measurement System (PeMS) database provides traffic parameters collected from roadside stations, and the OpenStreetMap is used to export the roadway network for simulation. The imported Map and traffic demand obtained from PeMS comprise the input of the simulation. Each simulation lasts for 3 hours from 7 am to 10 am, and the demand is randomly generated for each round. The average value for the demand of three routes in 3-hour are presented as follows: mainline [4791, 5172, 5045], off-ramp [801, 748, 620], and on-ramp [780, 879, 739]. The simulation is conducted in the Simulation of Urban Mobility (SUMO), providing an API - Traffic Control Interface (TraCI) package to realize interaction with external programs.

Figure 10 demonstrates the implementation of DRL-based DSH in SUMO. The testbed configuration is replicated as the environment for the DRL controller. E1 induction loop detectors (yellow mark with red arrow) are positioned along the segment, and their occupancy rates serve as input states. The state dimension is set as 10, including 4 lanes in the upstream mainline, 5 lanes in the downstream weaving area, and 1 lane in the on-ramp. In the DRL architecture, the agent (DRL controller) managed by the DDPG algorithm receives the state in the downstream congestion area, and takes actions (variable speed limits) returned to the environment. The action dimension is 6, from 50 to 75 mph with an increment of 5 mph. Then, the environment sends a new state decided by a defined policy mapping from the previous state and action. The interaction between the environment and the agent is updating the reward represented by MOE to the optimal value.

The incident is scheduled to occur at the downstream weaving area after the simulation begins. It is modeled by randomly stopping a vehicle for 5–10 min and taking 20 min to clear. Two types of vehicles are defined to simulate the mixed flow. The Krauss car-following model represents HDVs, with a driver compliance rate set at 0.7 (indicating 30% compliance with displayed speed limits). The Intelligent Driver Model (IDM) car-following model represents CAVs. Different headways are assigned for HDVs (1.2s) and CAVs (0.6s) based on their performance. All parameter values are adapted from previous studies (Hua and Fan, 2022). The lang-changing model uses the LC2013 in SUMO. To investigate the effects of DSH in mixed flow in various scenarios, a series of experiments are conducted under different traffic demands and penetration rates of CAVs. Sensitivity analyses of speed limit decrements in adverse weather and threshold of SSMs are also conducted.

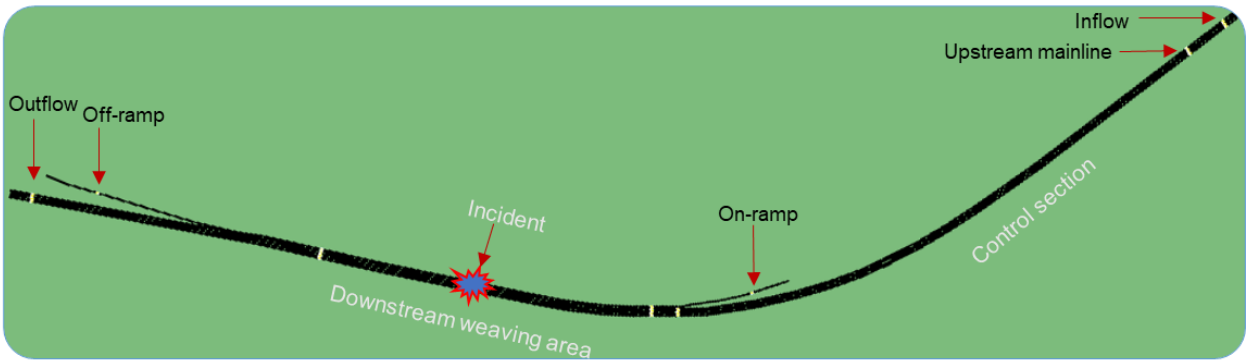


**Figure 10. Simulation Environment**

## 4.4 Results and Discussions

### 4.4.1 Evaluation of MOEs

The convergence of the cumulative reward during the learning process is shown in Figure 11. The DRL agent is trained under two situations, with no event and with special events, respectively. To test the performance during special events such as sports games and concerts, this study inflates the traffic demand of on-ramp by 1.5 times to simulate the traffic fluctuation caused by special events. The learning curves for various scenarios exhibit similarities and distinctions. Initially, both scenarios struggle to acquire valuable experiences, and the "with event" scenario starts from a lower point. However, the "no event" scenario reaches its peak around the 75$^{th}$ episode and maintains stability over time. While the "with event" scenario experiences two peaks, occurring around the 66$^{th}$ and 125$^{th}$ episodes, and the improvement is greater than that of no event at the end of training. Further details regarding the difference in terms of MOEs performance will be explored in subsequent discussions.

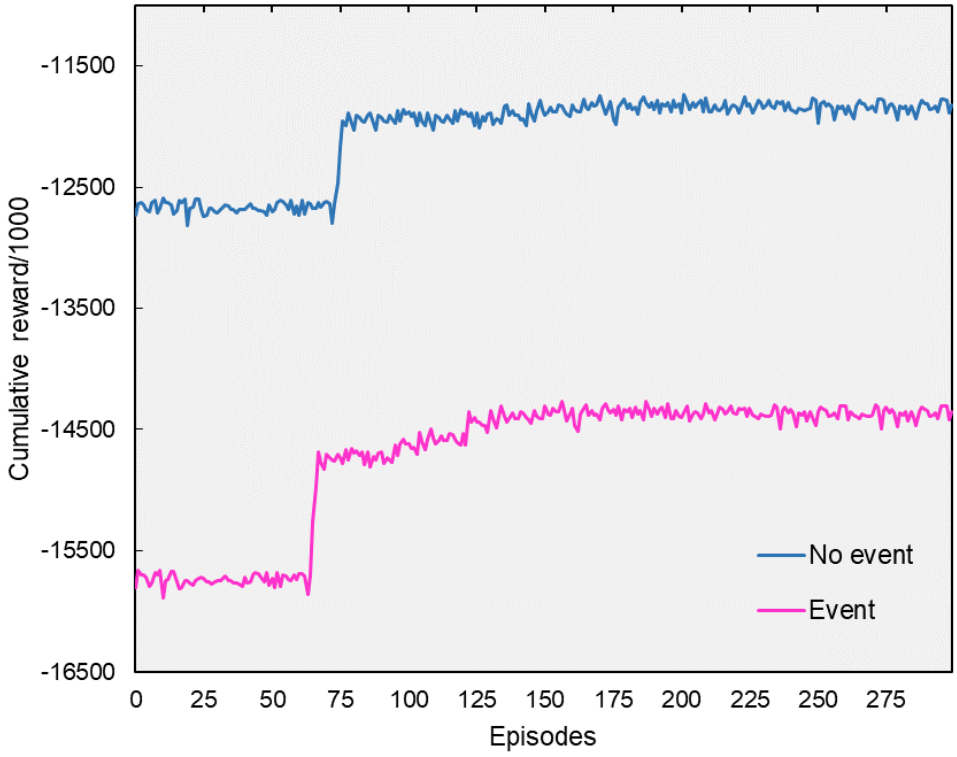


**Figure 11. The Learning Process of the Agent**

The holistic evaluation is done with 20 episodes for each MPR and takes the average value of MOE metrics. The time cost for pre- and post- control only increases from 260s to 269s. As depicted in Table 7, this study evaluates the MOEs from three aspects: safety represented by collision probability and DRAC, mobility represented by Average Travel Time (ATT) and time loss, and sustainability represented by greenhouse gas ($CO_2$) and pollutant gas (CO) emission (CO is mainly selected due to the fact that the proportion is much higher than others such as hydrocarbons, $NO_X$, and $PM_X$), and fuel consumption. To emphasize the effects of CAVs in different phases of mixed flow, 10% signifies early-stage deployment, 50% MPR represents mid-term adoption, and 90% MPR portrays future expectations. Key findings and potential co-benefits between MOEs are highlighted. Table 5.1 compares the performances of agents before and after DSH control under different MPRs of CAVs. The case of no control under 10% MPR (bold) serves as the baseline. All metrics, except collision probability (expressed as an absolute percentage), show enhancements relative to the baseline.

In no event situation, increased MPR leads to improvements in all MOE metrics, regardless of DSH activation. Compared to no control, the implementation of DSH further improves safety and mobility while slightly compromising sustainability. Specific to each metric, at the same MPR, a co-benefit is found between the collision probability and ATT after DSH. The collision probability can be reduced to 9.2%, and ATT can be reduced by 14.63% under 90%

MPR. Time loss, a consequence of not achieving optimal speed, diminishes with higher MPR, yet DSH has minimal impact at equivalent MPR levels. Sustainability fares well at 10% MPR but shows minor increments in emissions and fuel use at higher MPRs. To maximize the throughput, the CAVs will adopt a more aggressive way, leading to more severe cumulative decelerations and increased emissions. Notably, $CO_2$ emissions and fuel use exhibit a synchronous growth pattern. Generally, post-control scenarios outperform the base case with manageable emissions, and the safety priority for DSH is guaranteed.

Compared to no event, the time cost for no control escalates from 260s to 292s due to augmented demand. Remarkably, DSH becomes more time-efficient, costing only 293s. Concerning MOEs, all metrics demonstrate relatively inferior performance compared to the non-event scenario, indicating substantial event-induced traffic impacts. Focusing solely on pre- and post-control scenarios, all metrics except cumulative $DRAC_{max}$ improve with increasing MPR. This stems from the narrower headway adopted by CAVs, which introduces more vehicles, straining capacity and causing frequent deceleration. Unlike the no-event case, this phenomenon arises due to the volume nearing design capacity. MOE metric changes mirror those of the non-event scenario. The collision probability diminishes to 10.3%, and ATT can be reduced by 14.56% at 90% MPR after DSH. The increase in emissions is still acceptable. Given the protracted timeline for extensive CAV deployment, subsequent experiments will scrutinize post-DSH performance in mixed flow at a 50% MPR.

**Table 7. Comparisons of Performance Indicators Between No Control and DSH**

| | | | Improvement by percentage (%) | | | | | |
|---|---|---|---|---|---|---|---|---|
| **No event** | MPR (%) | Collision probability (%) | $DRAC_{max}$ Sum ($m/s^2$) | $CO_2$ (kg) | CO (kg) | Fuel use (L): | ATT (s) | Average time loss (s) |
| | **10%** | **30.9%** | **22577** | **8259** | **138** | **3550** | **90.68** | **19.26** |
| No control | 50% | 23.3% | 16.64 | 15.50 | 21.23 | 15.55 | 5.43 | 15.90 |
| | 90% | 15.0% | 33.49 | 23.53 | 38.10 | 23.53 | 9.82 | 25.00 |
| | 10% | 21.8% | 14.54 | 1.55 | 2.09 | 1.55 | 8.18 | 0.05 |
| DSH | 50% | 17.1% | 17.44 | 10.48 | 16.73 | 10.48 | 9.62 | 15.16 |
| | 90% | 9.2% | 20.43 | 20.17 | 28.35 | 20.17 | 14.63 | 23.13 |
| **Special event** | | | | | | | | |
| | **10%** | **31.8%** | **14892** | **9092** | **145** | **3908** | **94.09** | **25.81** |
| No control | 50% | 24.2% | -62.80 | 11.51 | 17.16 | 11.51 | 6.02 | 11.48 |
| | 90% | 13.9% | -93.13 | 21.30 | 34.42 | 21.30 | 12.50 | 18.90 |
| | 10% | 22.2% | -21.87 | -4.05 | -8.47 | -4.05 | 8.66 | 1.49 |
| DSH | 50% | 16.1% | -41.68 | 7.14 | 8.16 | 7.14 | 12.56 | 7.83 |
| | 90% | 10.3% | -84.39 | 15.77 | 19.32 | 15.77 | 14.56 | 19.04 |

### 4.4.2 Spatiotemporal Pattern of Speed Variations

Figure 12 demonstrates the spatiotemporal variations of the bottleneck speed. The z-axis of the surface plot is the average speed after DSH, the y-axis is the time of day, and the x-axis is the different locations at the bottleneck. The weaving area includes three sections: upstream and downstream near the on- and off- ramp, and the midstream portion. Each section is divided into 5 parts (#1-#5) that align with the five lanes present in the weaving area. #1 denotes the outermost lane near the ramp, and #5 denotes the innermost lane which is also the overtaking lane. Consequently, three parallel mountain-shaped plots are generated. The smoothness of the "mountain" reflects the degree of speed variations, and the flatter means less fluctuation.

From a spatial view, irrespective of location, the DSH controller consistently assigns the highest speed limit to the innermost overtaking lane (#5). This lane predominantly maintains speeds of around 65-70 mph, occasionally reaching 70 mph (indicated by red points). In contrast, speed decreases progressively as lanes approach the ramp. This behavior is motivated by the higher susceptibility of vehicles near the ramp, where even minor changes in driving behavior can exert pronounced effects on surrounding vehicles. However, even in the most severe downstream area, the lowest speed observed remains at 37 mph post-DSH, effectively preventing stop-and-go motions. When coupled with the MOE performance discussed earlier, this spatial pattern underscores the DSH agent's capacity to optimally allocate speed limits to various lanes based on their congestion levels. The isolation of the overtaking lane from the influence of other lanes further corroborates the efficacy of employing differential speed limits to ensure both safety and improved traffic flow.

In terms of temporal dimension, the midstream exhibits the least speed variability, consistently maintaining speeds exceeding 55 mph. The upstream follows suit, albeit with slightly more fluctuations. In contrast, the downstream section experiences the most speed variations. Additionally, the middle two lanes (#3 and #4) of each segment are the least affected, and the outer two lanes (#1 and #2) suffer the most speed variations. Comparing the conditions of the downstream and upstream outermost lanes, it can be seen that under the influence of the incident, the diverging process causes more impact than the merging process.

The heatmap in Figure 13 represents the projection of average vehicle speeds in both temporal and spatial dimensions, and the standard deviation of speeds is also included. Examining the evolving trends of adjacent color gradients over time reveals noteworthy insights. In the upstream bottleneck near the on-ramp, lane #4 which is closer to the overtaking lane exhibits the minimum standard deviation (0.91), indicating the most stable traffic flow. The maximum value (1.51) occurs in lane #2 near the outermost lane, raising safety concerns related to lane-changing behaviors during the merging process. Similar to the midstream bottleneck situation, lane #4 demonstrates higher stability with a standard deviation of 0.83. However, as

vehicles approach the downstream bottleneck, lane #1 in the outermost weaving area experiences significant speed oscillations. Throughout the entire period, the maximum value (8.13) is also observed where this lane is located near the off-ramp, emphasizing the imperative for improvement in this area to enhance safety in later stages.

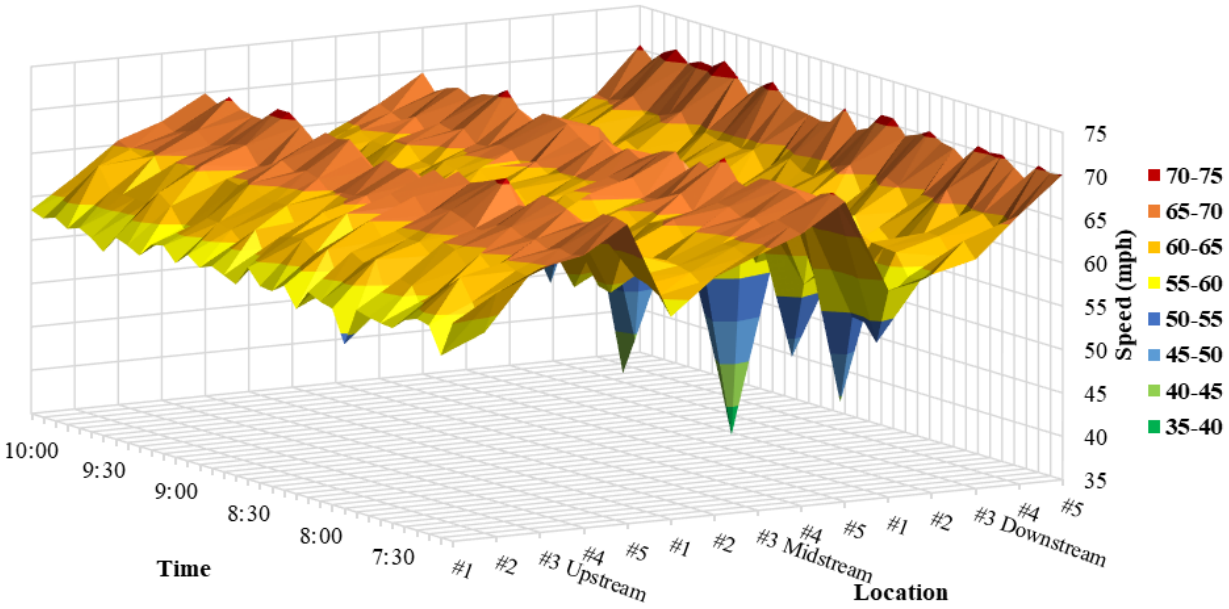


**Figure 12. Speed Variations by Locations and Time-of-day**

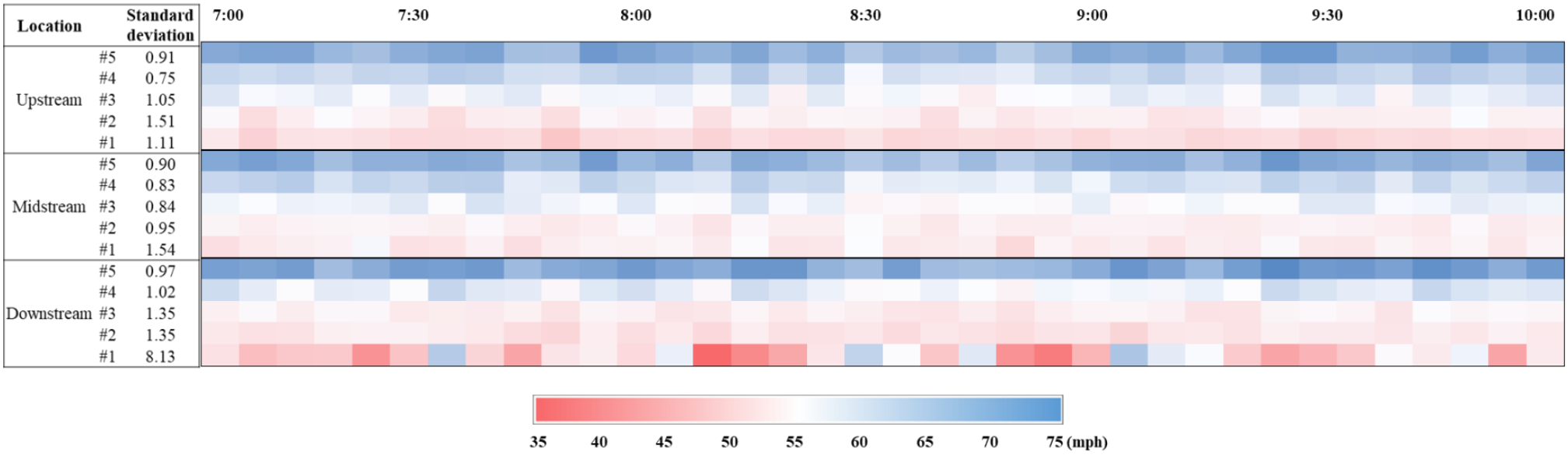


**Figure 13. Spatiotemporal Distribution of Bottleneck Speed**

### 4.4.3 Sensitivity Analysis of Speed Decrements

To delve deeper into the agent's performance across various scenarios, this study introduces simulations of adverse weather conditions by applying different speed limit decrements (5 and 10 mph). Adverse weather, such as rain or fog, tends to lead to reduced vehicle speeds due to obstructed vision, often resulting in lowered speed limits. While CAVs can utilize situational awareness to solve this problem, the effect is compromised by the presence of HDVs. Similarly, this section investigates the performance after DSH under 50% MPR, and uses the three key metrics to represent safety, mobility, and sustainability.

Figure 14 displays the MOEs alterations under varying speed decrements. The "normal" denotes driving in clear weather without any speed decrements. At a speed decrement of 5mph, the collision probability increases by 4.81%. The spatiotemporal pattern also shows that speed variations are more contained at higher speeds. When the speed descends significantly below the original speed limit of 65 mph on the mainline (a 10-mph speed decrement), the effect is further amplified. Correspondingly, $CO_2$ emissions also increase as the traffic situation worsens, but the change is not as obvious as that of safety. ATT alterations are least prominent at a 5-mph speed decrement with mere 0.32% change, but this effect becomes more noticeable with more severe speed reduction. Under adverse weather, the sensitivity of mobility changes is comparatively subdued in comparison to the impacts on safety and sustainability.

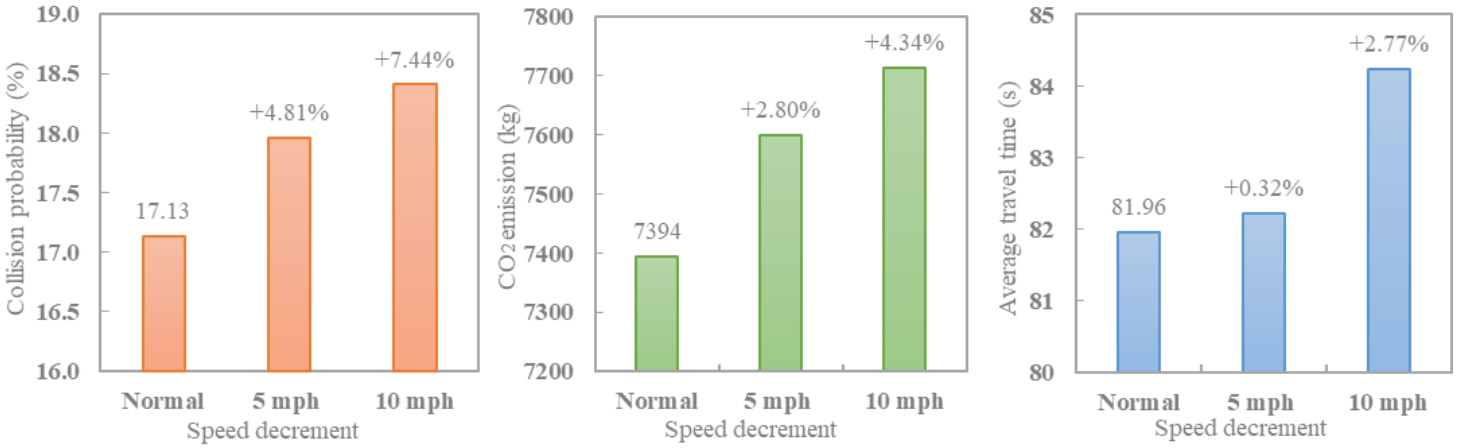


**Figure 14. MOEs Changes Under Different Speed Decrements**

### 4.4.4 Sensitivity Analysis of SSM Thresholds

The SSM parameters mentioned above rely on predefined thresholds to identify instances of traffic conflicts. It is acknowledged that this approach is both subjectively defined and statistically threshold-sensitive. Table 8 illustrates the sensitivity of SSM thresholds and how they reflect in the TTC-related collision probability. The first row (bold) is the baseline under 50% MPR after implementing the DSH strategy with the default value from the simulation.

Compared to the base, Perception-Reaction-Time (PRT) has the highest sensitivity. The collision probability increases as the PRT increases, but this effect saturates after more than 2s. The Deceleration Rate to Avoid the Crash (DRAC) becomes more obvious at a lower value, and it is not significant enough when it exceeds 3m/s$^2$. The range denotes the device's detection range in meters, and other vehicles are tracked when they are closer than a threshold to the equipped vehicle. A tree search is performed to find all vehicles close to the vehicle's current position. The sensitivity amplifies when more vehicles are identified within this range. The extra time expresses the time an encounter is tracked after not being associated with a potential conflict (ceases when a vehicle departs from a common route, crosses the conflict area, changes lanes, or exits the detection range, etc.). Again, this parameter primarily affects safety at higher threshold levels.

**Table 8. Safety Changes Under Different SSM Thresholds**

| Collision probability change (%) | | | | | | | |
|---|---|---|---|---|---|---|---|
| **DRAC=3m/s$^2$** | | **Prt=1s** | | **Range=50m** | | **Extra time=5s** | |
| 2s | 0.94% | 2s | 2.83% | 20m | -0.13% | 4s | 0.06% |
| 4s | -0.17% | 3s | 2.56% | 80m | -1.17% | 6s | -1.62% |

# Chapter 5. Extension of DSH Strategy

## 5.1 Introduction

The growing travel demand, coupled with a nearly stagnant infrastructure supply has significantly worsened traffic congestion. The negative impacts include increased collision risk, longer travel time, excessive fuel usage, Greenhouse Gas (GHG) emissions, and pollution (Tian et al., 2018). Each year more than 6 million crashes occur in the U.S., about 31,000 highway fatalities (Hughes et al.,2023). Around 9 billion hours of travel delay across the nation yearly, equating to $192 billion congestion cost (Schrank et al., 2015). Over 56 billion pounds of additional $CO_2$ are produced every year, which means 3.2 billion gallons of wasted fuel (Eisele et al., 2014).

Compared to expanding additional motorways, utilizing existing infrastructure with active traffic management is a cost-effective approach. Dynamic Speed Harmonization (DSH), also known as Variable Speed Limit (VSL), is a mature control strategy to stabilize traffic flow (Ma et al., 2016). It can be applied by adjusting speed limits shown on variable message signs (VMS) based on current conditions in the downstream bottleneck (Papageorgiou et al., 2008). In this way, speed oscillations resulting from the upstream propagation of shock waves are dampened, and a smoother transition of upcoming traffic can be achieved (Kušić et al., 2020).

However, low compliance rates of drivers restrict the effectiveness of this technique to a great extent. Emerging technologies such as Connected and Automated Vehicles (CAVs) provide potentially cutting-edge solutions to improve various Measures of Effectiveness (MOEs) (Wang et al., 2016; Hua and Fan, 2023). With the advancement of Vehicle-to-Everything (V2X) communication technology and automation, CAVs are anticipated to completely comply with the control system and have no information delay. Before the large-scale deployment of CAVs considering the public acceptance, the mixed traffic involves Human-Driven Vehicles (HDVs) and CAVs will exist in the long term (Li et al., 2022). Therefore, an efficient integration of the DSH technique and CAV technology to optimize freeway operations becomes a critical transportation issue.

Since field tests or on-road practice are costly and might lead to unforeseen and severe effects on existing traffic if implemented incorrectly, the simulation-based method has gained popularity (Lu and Shladover, 2014). Traditional control logic often exhibits a noticeable reaction delay, the traffic breakdown may already appear when DSH is activated (Malikopoulos et al., 2018). In contrast, the proactive method allows for taking measures promptly (Khondaker and Kattan, 2015). However, it highly depends on an accurate macroscopic prediction model and cannot reflect disturbances in driving behaviors. Instead, Deep Reinforcement Learning (DRL) can adapt to traffic dynamics without the explicit model, it does not rely on the fundamental diagram to tune the controller (Lu et al., 2023;

Hua and Fan, 2024). Compared to the Single-Agent Reinforcement Learning (SARL) system, the distributed multi-agent system (MARL) can be employed flexibly without concern for the breakdown of the central controller. Using agents to work simultaneously can obtain a better coordination effect (Wang et al., 2019).

To address the aforementioned limitations, this chapter investigates the effects of MARL-based DSH strategy in mixed traffic on the freeway. The main points are as follows: (a) to prevent getting stuck in local optimization, a Multi-Agent Dynamic Speed Harmonization (MADSH) system is developed; (b) a lane-based strategy is considered to verify the feasibility of setting differential speed limits for each lane; (c) the impacts on safety, mobility, and sustainability, and interactions between MOEs are quantified through a holistic evaluation; (d) to thoroughly comprehend how CAVs improve the operational performance, effects of CAVs at varying Market Penetration Rates (MPRs) are explored; and (e) sensitivity analysis under multiple traffic scenarios is conducted to test the adaptation of the model. This study provides essential insights to foster a deeper understanding of the transformative potential of the CAV-powered DSH technique in promoting intelligent transportation systems.

The structure of this chapter is as follows: Section 5.2 introduces the proposed MADSH framework. Section 5.3 describes the experimental settings and microscopic simulation platform. Section 5.4 presents the numerical results and discussions.

## 5.2 Methodology

### 5.2.1 Proposed Framework

A MADSH system can be defined as: “a group of intelligent, interacting agents using speed harmonization on a managed freeway segment.” Within this landscape, two noteworthy points stand out: robustness of agents to learn the system dynamics, and adaptation to other agents’ evolving action (Busoniu et al., 2008). Therefore, the coordination between multiple agents should be guaranteed. Figure 15 demonstrates the lane-based MADSH system in this research. When the traffic demand exceeds the freeway capacity, the merging and diverging behaviors that occur in the weaving area will result in a downstream bottleneck. The downstream detector on each lane collects the congestion information and sends it to the Transportation Management Center (TMC). The TMC then activates the predefined MADSH strategy embedded in the controllers to optimize the traffic flow of the control section. To facilitate the implementation of the system, the Roadside Units (RSU) along the freeway and the VMS in the gantry display the optimal speed limits derived from the controllers. The RSUs transmit information to CAVs through Infrastructure-to-Vehicle (I2V) communication, and CAVs execute the control commands automatically. The CAVs within the communication range exchange their vehicular

information via Vehicle-to-Vehicle (V2V) communication. The VMS on each lane provides the information for the HDVs. This research assumes that there is no communication delay or information loss during the control period.

Different from the previous DSH technique, which divided an entire freeway segment longitudinally into multiple consecutive sections, this study considers a lane-based control strategy, which divides the control section latitudinally into parallel segments based on the number of lanes. Setting a homogeneous and synchronized speed limit across all lanes is inefficient. When two traffic flows interfere on the outer lane in the weaving area, the innermost overtaking lane is actually not affected. Shi and Liu (2019) have revealed that implementing a differential variable speed limit on each lane can mitigate the speed variations triggered by slow-moving vehicles occupying any lane, and ensure safety and mobility.

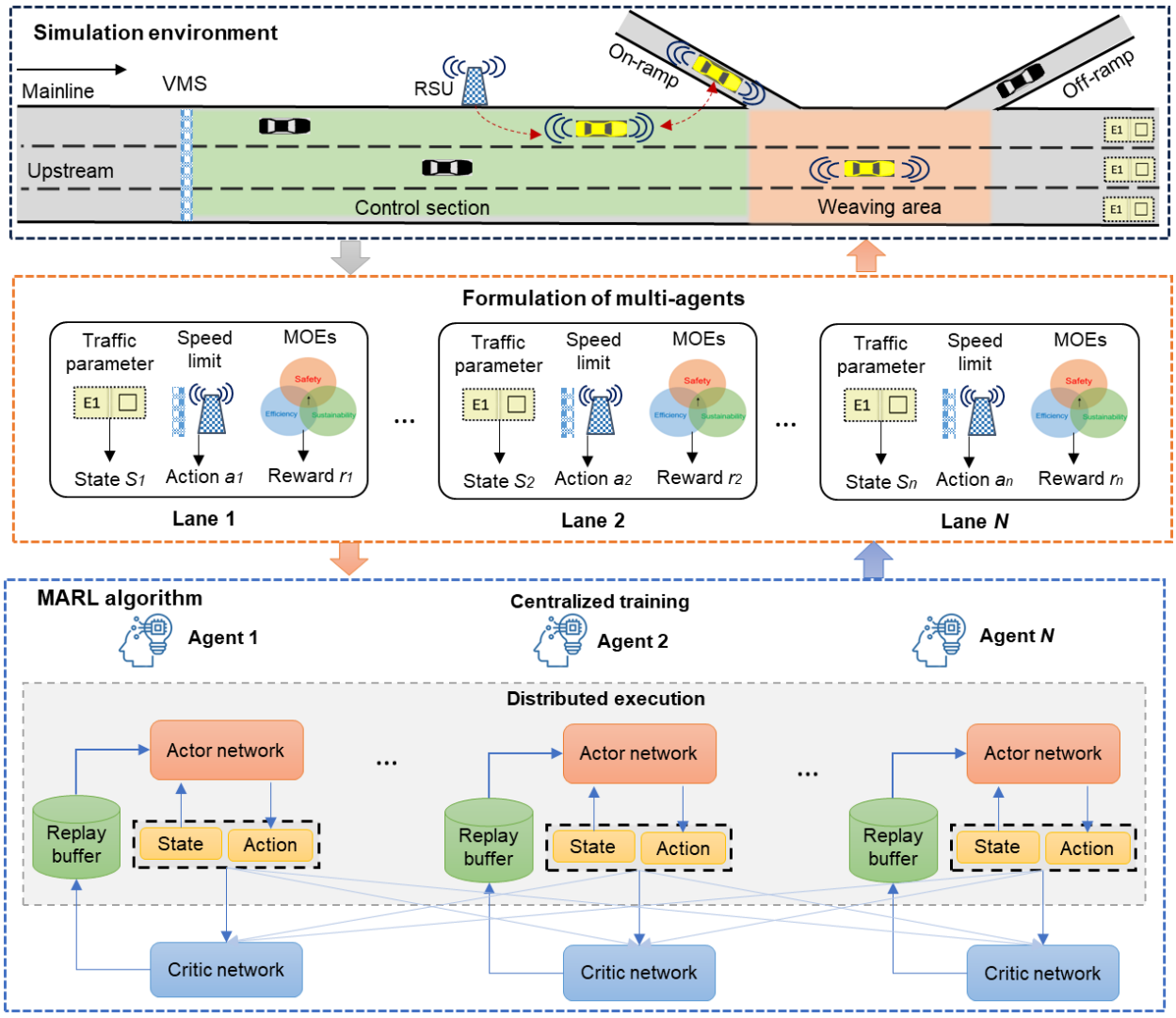


**Figure 15. Scheme of MADSH Control System**

5.2.2 MARL Formulation

The MARL-based DSH issue can be formalized as an *N*-agent interacting Markov decision process. In the lane-based strategy, the controller on each lane acts as an agent. The quantity and serial number of the agents correspond to the number and sequence of lanes. MARL extends the standard RL paradigm with multiple agents, the MARL can be defined as a tuple (*S*, *A*, *P*, *R, N*), where *S* represents a set of local states *s*, the global state composed of the local states perceived by all agents, *A* denotes a set of actions *a*, *P* is the transition probability from the last step *a* in *s* based on policy $\pi$ that leads to next state $s'$, *R* is the reward with a discount factor $\gamma$ from 0 to 1 for the agent after the transition, and *N* denotes the number of agents. The goal of each agent is to learn a $\pi$ that maximizes its own cumulative reward. It is hard to achieve a common goal with maximum individual rewards when agents cooperate with each other. The configurations of state, action, and reward function are the same as previous settings.

Two main categories are applicable to classify MARL algorithms: centralized and decentralized learning. The centralized algorithms transform the MARL into a single-agent problem that is usually inefficient considering the large state-action space. The decentralized algorithms are limited due to the non-stationarity issue (Sunehag et al., 2017). The Centralized Training and Decentralized Execution (CTDE) framework has been paid more attention. The centralized training enables each agent to evaluate the policies of other agents, and the non-stationarity problem caused by the learning dynamics of other agents can be mitigated. The decentralized execution allows each agent to separately take actions based on its local observations. The robustness and adaptation of algorithms can be ensured in this way.

This study employs the MADDPG (Multi-Agent Deep Deterministic Policy Gradient) algorithm (Lowe et al., 2017) for a strategy solution. By considering the behaviors of other agents, MADDPG improves coordination among agents. It is designed to handle environments where agents have partial observability, and is well-suited for continuous action spaces where precise control is necessary. This makes it applicable to a wide range of real-world problems. The $\pi$ is built on an Actor-Critic architecture that provides both value- and policy-based function approximation in the deep neural networks. The actor network generates action based on the local state and its own policy. When all agents have executed actions, the environment returns a reward to each agent. The critic network evaluates the action according to the global state of all agents. Then, the experience is stored in the replay buffer to support sampling. Each agent will take a mini-batch of samples to update the parameters of the online actor network $\theta^{\mu}$, critic network $\theta^{c}$, and corresponding target actor network $\theta^{\mu\prime}$, critic network $\theta^{c\prime}$. During the training process, each $\theta^{c}$ is updated by minimizing the Temporal Difference error:

$$L(\theta_i^c) = E_{s,\, s',a,r}[(Q_i^c(s,a|\theta_i^c - y)^2] \tag{14}$$

$$y = r_i + \gamma Q_i^{c'}(s', \mu_i'(s_i|\theta_i^{\mu'})|_{a_i'=\theta_i^{c'}(s_i')} \quad (15)$$

The weights of $\theta^{\mu}$ can be updated by taking the deterministic policy gradient:

$$\nabla_{\theta^{\mu}} = \frac{1}{N_{\text{sample}}} \sum_1^N \nabla_{a_i} Q_i^c(s, a|\theta_i^c)\,|_{a_i=\mu_i(s_i)} \nabla_{\theta_i^{\mu}} \mu_i\left(s\middle|\theta_i^{\mu}\right)|_{s=s_i} \quad (16)$$

The target actor and critic networks are replaced by the "soft updating" factor $\tau$=0.01. The memory capacity of experience replay is set to 50000. The discount factor $\gamma$ is 0.95. The light-weight deep neural networks with two layers are established, and each layer has 64 neurons. The batch size is set to 1024, the learning rate of the actor is 0.001 and the critic is 0.002. The Adam optimizer is used to adapt the learning rate. The steps of the MADDPG algorithm for DSH are summarized as follows:

Step 1: Randomly initialize parameters for the actor network $\theta^{\mu}$ and the critic network $\theta^{c}$ of all agents, and set target weights of $\theta^{\mu\prime}$ and $\theta^{c\prime}$. Then empty the replay buffer.

Step 2: Load the environment and run the simulation with state *s*. Repeat steps 2 to 4 until the episode reaches its predefined maximum value.

Step 3: During time lengths in the simulation, each agent *i* from 1 to *N* explores action *a* based on the current policy $\pi$ and noise decay.

Step 4: Select variable speed limits derived from action *a*, and observe reward *r* and new state *s'* of the local lane. Store the useful experience in the reply buffer.

Step 5: For each agent *i* from 1 to *N*, randomly sample a mini-batch of transitions from the reply buffer. Update the critic network by the loss function (6.1). Update the actor network by the Equation (6.3).

Step 6: For each agent, update the target actor $\theta^{\mu'}$ and critic $\theta^{c'}$networks by $\tau\theta^{\mu} + (1-\tau)\theta^{\mu'}$ and $\tau\theta^{c} + (1-\tau)\theta^{c'}$ until convergence.

## 5.3 Experimental Settings

The study site is the same as previous chapters. Each simulation lasts for 4 hours in the peak-hour morning, and the control cycle is set to 60s. To truly reflect the dynamics of real traffic flow, the demand is stochastically generated for each round. Two types of vehicles are defined to model the mixed flow. The Krauss car-following model is used to simulate HDVs, with a 70% driver compliance rate. The IDM car-following model is used for simulating CAVs, assuming perfectly comply with the system. The headway is set as 1.2 s for HDVs and 0.9 s for CAVs considering their features and performances (Treiber et al., 2000; Hua and Fan, 2022). The lang-changing model uses the default LC2013 in SUMO. Three MPRs are set to investigate the effects of CAVs at different deployment

stages, which 10% signifies early-stage adoption, 50% represents mid-term deployment, and 90% portrays future expectations.

To quantify the performance on safety, mobility, and sustainability, a series of MOE metrics are evaluated. It consists of safety represented by collision probability, mobility represented by Average Travel Time (ATT), and sustainability represented by main GHG ($CO_2$) and pollutant emission (CO is selected due to the much higher proportion), and fuel consumption. The mobility metrics can be directly derived from the output of the simulation. To measure the sustainability metrics, this study uses the Handbook Emission Factors for Road Transport version 3 (HBEFA3) model. It is suitable for a gasoline-powered Euro norm 4-passenger car. Emission factors are provided for $CO_2$, CO, NOx, PMx, HC, and fuel consumption by computing weighted value per vehicle type, emission stage, fuel type, or sub-segment (= vehicle type/size class/emission stage) and traffic situation.

## 5.4 Results and Discussions

To evaluate the effectiveness of the proposed strategy, this study compares it with two baselines. (1) no control, in which there is no DSH activated the whole period, and each vehicle runs within the original speed limit, (2) SADSH, in which only one controller is managing the whole control section. The components of RL in single-agent are consistent with MADSH. Figure 16 displays the cumulative rewards of baselines and the proposed method under 50% MPR after 500 episodes of training. The initial values of both DSHs are comparatively lower than no control, however, the subsequent learning curves experience a significant enhancement. It is discovered that the agents stay in a narrow range and seldom gain any valuable experience before 73 episodes for MADSH and 127 episodes for SADSH. As more experience is learned, the reward of SADSH hits its first peak approximately 140 episodes. When more useful information is extracted, reaching the maximum value in the $217^{th}$ episode, there is not any improvement in the following learning process. The difference for MADSH is that the turning point occurs earlier than that of SADSH, which means that the multi-agent system can capture useful information more quickly and make optimizations earlier. More importantly, MADSH can further improve the reward based on the single-agent system, which shows that SADSH actually falls into local optimization. The highest reward value of MADSH appears in the $307^{th}$ episode, then the learning process progressively stabilizes and no higher value is observed.

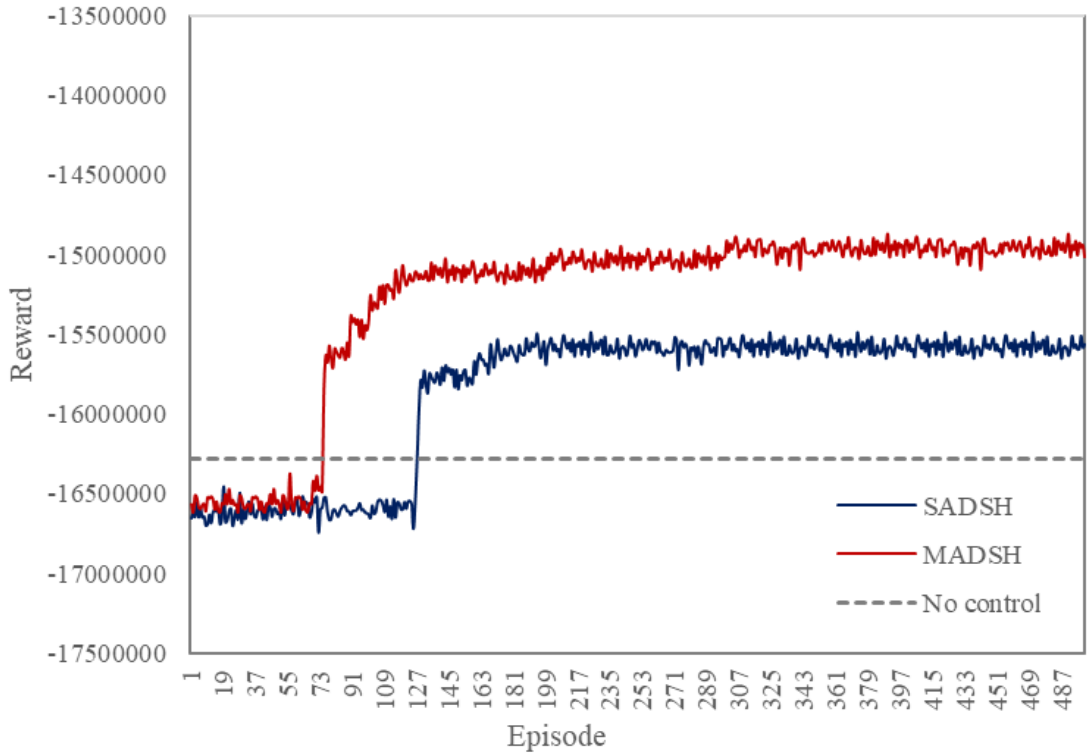


**Figure 16. Learning Process**

To quantify the impacts of the proposed strategy on safety, mobility, and sustainability, and interactions between various MOE metrics, a holistic evaluation is conducted. Moreover, the effects of CAVs at varying MPRs are explored to thoroughly understand how CAVs improve operational performance at different deployment stages. The results are displayed in Table 9.

For the safety represented by collision probability, both SADSH and MADSH are better than no control at any MPR. Compared with SADSH, MADSH can further reduce the risk of collision. In addition, this improvement shows differences under different MPRs. In the early deployment stage of CAV, MADSH can reduce the risk to up to 20.7%, which is a significant improvement compared to no control. As more CAVs are introduced into the network, the collision probability can be reduced to up to 9.2%. However, it is worth noting that compared with baselines, MADSH at higher MPR does not improve safety as much as in the initial stage. This shows that in the future stages of the development of intelligent transportation systems, traffic flow can be optimized solely by relying on the performance of CAV itself, and DSH technology will become obsolete.

For mobility, the changes in ATT under different MPRs are not obvious, which indicates that depending only on CAV cannot improve operating efficiency. In fact, DSH's primary goal is to ensure safety, sometimes will scarify mobility. Fortunately,

DSH-powered CAV can significantly reduce ATT, and MADSH can further shorten the time from approximately 82s to 74s. This also reflects the superiority of the multi-agent system.

From the perspective of sustainability, for main GHG emission ($CO_2$ constitutes around 26% of all GHGs such as $O_3$, $CH_4$, and $N_2O$), with the increase of MPR, MADSH can effectively reduce GHG emissions, which can make 9722 kg at 10% MPR dropped to 8348 kg at 90% MPR. Interestingly, fuel consumption and $CO_2$ emissions show the same changing trend, which reflects that there is a proportional relationship between the two. It is worth noting that the use of DSH in mixed traffic flows may result in a slight increase in gas emissions. This may be caused by several reasons: (a) Inconsistent Driving Patterns, in which the erratic behavior caused by the unpredictability of HDVs can force CAVs to adapt constantly, leading to less efficient driving, (b) Suboptimal Traffic Flow, the stop-and-go movement by HDVs can disrupt the smooth flow of CAVs, causing CAVs to brake and accelerate more frequently, (c) Interaction Dynamics, because CAVs are programmed to adopt more conservative driving strategies to ensure safety, while HDVs might not respect these gaps or might cut in more aggressively, causing CAVs to frequently adjust their speed and lane position, (d) System Limitations, CAVs may need to adapt to infrastructure not optimized for mixed traffic or non-communicative vehicles, which can hinder the fuel efficiency gains expected in an all-CAV environment. In addition, the same changing trend can also be found in the harmful gas CO.

**Table 9. The Performance of MOE Metrics**

| MPR | Control | Collision probability (%) | ATT (s) | $CO_2$ (kg): | CO (kg): | Fuel usage (L): |
|---|---|---|---|---|---|---|
| 10% | No | 31.4% | 82.41 | 9722 | 144 | 4179 |
| | SADSH | 22.0% | 78.30 | 10573 | 172 | 4545 |
| | MADSH | 20.7% | 74.39 | 10150 | 165 | 4363 |
| 50% | No | 23.8% | 82.52 | 8871 | 120 | 3814 |
| | SADSH | 16.6% | 78.77 | 9497 | 145 | 4082 |
| | MADSH | 15.6% | 74.83 | 9117 | 139 | 3919 |
| 90% | No | 14.5% | 82.12 | 8100 | 101 | 3482 |
| | SADSH | 9.8% | 78.70 | 8696 | 125 | 3738 |
| | MADSH | 9.2% | 74.77 | 8348 | 120 | 3589 |

# Chapter 6. Summary and Conclusions

## 6.1 Conclusions

In the vicinity of weaving areas, freeway congestion is nearly unavoidable due to their negative effects on the continuous freeway mainline flow. The adverse impacts include increased collision risks, extended travel time, and excessive emissions and fuel consumption. DSH has the potential to dampen traffic oscillation during congestion. However, the effectiveness of this strategy is typically limited by the low compliance rates of drivers and delays in information access. CAVs are introduced as part of the intelligent transportation systems to enhance a variety of MOEs. This research investigates the effects of DSH in mixed traffic flow involving HDVs and CAVs on the freeway. The main contributions are as follows: (a) a MADSH system is developed, (b) a lane-based strategy is considered to verify the feasibility of setting differential speed limits for each lane, (c) the impacts on safety, mobility, and sustainability, and interactions between MOEs are quantified through a holistic evaluation, (d) to thoroughly comprehend how CAVs improve the operational performance, effects of CAVs at varying MPRs are explored, (e) sensitivity analysis under multiple traffic scenarios is conducted to test the adaptation of the model. This study provides essential insights to foster a deeper understanding of the transformative potential of the CAV-powered DSH technique in promoting intelligent transportation systems.

The results show that the suggested approach can improve safety and freeway mobility during recurrent congestion, while also enhancing environmental sustainability at higher MPRs. The bottleneck speed's spatiotemporal characteristics demonstrate how DSH driven by CAVs might lessen speed variations in particular regions. Headway sensitivity indicates that high-level CAVs can improve performance substantially. As MPRs increase, the technique can improve safety and mobility for nonrecurrent congestion. Although special events might worsen congestion, their impact can be partially mitigated through speed controls. Spatiotemporal patterns of speed variations show how the controller can lessen oscillations and improve traffic flow. Sensitivity analyses also show how the agent responds to varying parametric thresholds and how flexible it is in inclement weather. Moreover, the application of MADSH system can prevent the proposed strategy falling into local optimization.

## 6.2 Future Work

Future directions can be devoted to the following aspects: (a) the effect of road configuration on the performance, such as the length of weaving area and control section deserves more attention, and the position and length of the control area can also be optimized; (b) in addition to the headway, there are other parameters describing the characteristics of the vehicle can be investigated, and mixed flow with trucks can also be

considered; c) the transferability of SSM considering the features of HDV and CAV, and a universal set of indicators is highly needed to evaluate the transportation safety in mixed traffic flow environment; (d) more abundant state representation of RL such as the trajectory data can be considered to better describe the traffic situation; (e) the explainability of the model should be explored by combining RL with other advanced techniques; and (f) the DSH can be integrated with other ATMs such as ramp metering to investigate the effects of merging control.

# References


1. Abdel-Aty, M., Dilmore, J., & Dhindsa, A. (2006). Evaluation of variable speed limits for real-time freeway safety improvement. Accident analysis & prevention, 38(2), 335-345.
2. Ahmed, M. S., & Cook, A. R. (1979). Analysis of freeway traffic time-series data by using Box-Jenkins techniques (No. 722).
3. Alessandri, A., Di Febbraro, A., Ferrara, A., & Punta, E. (1999). Nonlinear optimization for freeway control using variable-speed signaling. IEEE Transactions on vehicular technology, 48(6), 2042-2052.
4. Barth, M., & Boriboonsomsin, K. (2009). Energy and emissions impacts of a freeway-based dynamic eco-driving system. Transportation Research Part D: Transport and Environment, 14(6), 400-410.
5. Busoniu, L., Babuska, R., & De Schutter, B. (2008). A comprehensive survey of multiagent reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 38(2), 156-172.
6. Cooper, D. F., & Ferguson, N. (1976). Traffic studies at T-Junctions. 2. A conflict simulation Record. Traffic Engineering & Control, 17(Analytic).
7. Eisele, W. L., Fossett, T., Schrank, D. L., Farzaneh, M., Meier, P. J., & Williams, S. P. (2014). Greenhouse Gas Emissions and Urban Congestion: Incorporation of Carbon Dioxide Emissions and Associated Fuel Consumption into Texas A&M Transportation Institute Urban Mobility Report. Transportation Research Record, 2427(1), 73-82.
8. Fang, J., Luo, Y., Hadiuzzaman, M., Liu, G., & Qiu, T. Z. (2015). Safety oriented variable speed limit control method with enhanced driver response modeling (No. 15-2393).
9. Ghiasi, A., Li, X., & Ma, J. (2019). A mixed traffic speed harmonization model with connected autonomous vehicles. Transportation Research Part C: Emerging Technologies, 104, 210-233.
10. Gora, P., Katrakazas, C., Drabicki, A., Islam, F., & Ostaszewski, P. (2020). Microscopic traffic simulation models for connected and automated vehicles (CAVs)–state-of-the-art. Procedia Computer Science, 170, 474-481.
11. Gregurić, M., Kušić, K., & Ivanjko, E. (2022). Impact of Deep Reinforcement Learning on Variable Speed Limit strategies in connected vehicles environments. Engineering Applications of Artificial Intelligence, 112, 104850.
12. Gregurić, M., Kušić, K., Vrbanić, F., & Ivanjko, E. (2020, September). Variable speed limit control based on deep reinforcement learning: A possible implementation. In 2020 International Symposium ELMAR (pp. 67-72). IEEE.

13. Gregurić, M., Vrbanić, F., & Ivanjko, E. (2023). Towards the spatial analysis of motorway safety in the connected environment by using explainable deep learning. Knowledge-Based Systems, 269, 110523.
14. Grumert, E., Ma, X., & Tapani, A. (2015). Analysis of a cooperative variable speed limit system using microscopic traffic simulation. Transportation research part C: emerging technologies, 52, 173-186.
15. Ha, T.-J., Kang, J.-G. & Park, J.-J. (2003), The effects of automated speed enforcement systems on traffic-flow characteristics and accidents in Korea, ITE Journal, 73(2), 28.
16. Hadiuzzaman, M., & Qiu, T. Z. (2013). Cell transmission model based variable speed limit control for freeways. Canadian Journal of Civil Engineering, 40(1), 46-56.
17. Han, Y., Chen, D., & Ahn, S. (2017). Variable speed limit control at fixed freeway bottlenecks using connected vehicles. Transportation Research Part B: Methodological, 98, 113-134.
18. Hayward, J. C. (1972). Near-miss determination through use of a scale of danger. Highway Research Record, 24-34.
19. Hegyi, A., De Schutter, B., Hellendoorn, H., & Van Den Boom, T. (2002, May). Optimal coordination of ramp metering and variable speed control-an MPC approach. In Proceedings of the 2002 American Control Conference (IEEE Cat. No. CH37301) (Vol. 5, pp. 3600-3605). IEEE.
20. Hegyi, A., Schutter, B. D. & Hellendoorn, H. (2005), Model predictive control for optimal coordination of ramp metering and variable speed limits, Transportation Research Part C: Emerging Technologies, 13(3), 185–209.
21. https://international.fhwa.dot.gov/pubs/pl07012/images/figure_1.cfm
22. Hua, C., & Fan, W. D. (2022). Freeway Traffic Speed Prediction under the Intelligent Driving Environment: A Deep Learning Approach. Journal of Advanced Transportation, 2022.
23. Hua, C., & Fan, W. D. (2024). Safety-oriented dynamic speed harmonization of mixed traffic flow in nonrecurrent congestion. Physica A: Statistical Mechanics and its Applications, 634, 129439.
24. Hua, C., & Fan, W. (2023). Dynamic speed harmonization for mixed traffic flow on the freeway using deep reinforcement learning. IET Intelligent Transport Systems, 17(12), 2519-2530.
25. Hughes, J. E., Kaffine, D., & Kaffine, L. (2023). Decline in traffic congestion increased crash severity in the wake of COVID-19. Transportation research record, 2677(4), 892-903.
26. Islam, M. T., Hadiuzzaman, M., Fang, J., Qiu, T. Z., & El-Basyouny, K. (2013). Assessing mobility and safety impacts of a variable speed limit control strategy. Transportation research record, 2364(1), 1-11.

27. Khondaker, B. & Kattan, L. (2015), Variable speed limit: a microscopic analysis in a connected vehicle environment, Transportation Research Part C: Emerging Technologies, 58(September), 146–59.
28. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
29. Kušić, K., Dusparic, I., Guériau, M., Gregurić, M., & Ivanjko, E. (2020, September). Extended variable speed limit control using multi-agent reinforcement learning. In 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC) (pp. 1-8). IEEE.
30. Kušić, K., Ivanjko, E., Gregurić, M., & Miletić, M. (2020). An overview of reinforcement learning methods for variable speed limit control. Applied Sciences, 10(14), 4917.
31. Lee, C., Hellinga, B., & Saccomanno, F. (2006). Evaluation of variable speed limits to improve traffic safety. Transportation research part C: emerging technologies, 14(3), 213-228.
32. Li, D., & Wagner, P. (2019). Impacts of gradual automated vehicle penetration on motorway operation: a comprehensive evaluation. European transport research review, 11(1), 1-10.
33. Li, Y., Li, Z., Wang, H., Wang, W., & Xing, L. (2017). Evaluating the safety impact of adaptive cruise control in traffic oscillations on freeways. Accident Analysis & Prevention, 104, 137-145.
34. Li, Y., Pan, B., Xing, L., Yang, M., & Dai, J. (2022). Developing dynamic speed limit strategies for mixed traffic flow to reduce collision risks at freeway bottlenecks. Accident Analysis & Prevention, 175, 106781.
35. Li, Y., Xu, C., Xing, L., & Wang, W. (2017). Integrated cooperative adaptive cruise and variable speed limit controls for reducing rear-end collision risks near freeway bottlenecks based on micro-simulations. IEEE transactions on intelligent transportation systems, 18(11), 3157-3167.
36. Li, Z., Li, Y., Liu, P., Wang, W., & Xu, C. (2014). Development of a variable speed limit strategy to reduce secondary collision risks during inclement weathers. Accident Analysis & Prevention, 72, 134-145.
37. Li, Z., Xu, C., Guo, Y., Liu, P., & Pu, Z. (2020). Reinforcement learning-based variable speed limits control to reduce crash risks near traffic oscillations on freeways. IEEE Intelligent Transportation Systems Magazine, 13(4), 64-70.
38. Li, Z., Zhu, X., Liu, X., & Qu, X. (2019, October). Model-based predictive variable speed limit control on multi-lane freeways with a line of connected automated vehicles. In 2019 IEEE Intelligent Transportation Systems Conference (ITSC) (pp. 1989-1994). IEEE.
39. Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., ... & Wierstra, D. (2015). Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.

40. Lowe, R., Wu, Y. I., Tamar, A., Harb, J., Pieter Abbeel, O., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. Advances in neural information processing systems, 30.
41. Lu, W., Yi, Z., Gu, Y., Rui, Y., & Ran, B. (2023). TD3LVSL: A lane-level variable speed limit approach based on twin delayed deep deterministic policy gradient in a connected automated vehicle environment. Transportation Research Part C: Emerging Technologies, 153, 104221.
42. Lu, X. Y., & Shladover, S. E. (2014). Review of variable speed limits and advisories: Theory, algorithms, and practice. Transportation research record, 2423(1), 15-23.
43. Lu, X. Y., Qiu, T. Z., Varaiya, P., Horowitz, R., & Shladover, S. E. (2010, June). Combining variable speed limits with ramp metering for freeway traffic control. In Proceedings of the 2010 american control conference (pp. 2266-2271). IEEE.
44. Ma, J., Li, X., Shladover, S., Rakha, H. A., Lu, X. Y., Jagannathan, R., & Dailey, D. J. (2016). Freeway speed harmonization. IEEE Transactions on Intelligent Vehicles, 1(1), 78-89.
45. Malikopoulos, A. A., Hong, S., Park, B. B., Lee, J., & Ryu, S. (2018). Optimal control for speed harmonization of automated vehicles. IEEE Transactions on Intelligent Transportation Systems, 20(7), 2405-2417.
46. Milanés, V., & Shladover, S. E. (2014). Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data. Transportation Research Part C: Emerging Technologies, 48, 285-300.
47. Müller, E. R., Carlson, R. C., & Kraus, W. (2016). Cooperative mainstream traffic flow control on freeways. IFAC-PapersOnLine, 49(32), 89-94.
48. Papageorgiou, M., Kosmatopoulos, E., & Papamichail, I. (2008). Effects of variable speed limits on motorway traffic flow. Transportation Research Record, 2047(1), 37-48.
49. Payne, H. J. (1971). Model of freeway traffic and control. Mathematical Model of Public System, 51-61.
50. Peng, C., & Xu, C. (2022). Combined variable speed limit and lane change guidance for secondary crash prevention using distributed deep reinforcement learning. Journal of Transportation Safety & Security, 14(12), 2166-2191.
51. Piao, J., & McDonald, M. (2008, October). Safety impacts of variable speed limits-a simulation study. In 2008 11th International IEEE Conference on Intelligent Transportation Systems (pp. 833-837). IEEE.
52. Robinson, M. D. (2000), Examples of variable speed limit applications, in Transportation Research Board, 79th Annual Meeting, Washington DC.
53. Salles, D., Kaufmann, S., & Reuss, H. C. (2020, October). Extending the intelligent driver model in SUMO and verifying the drive off trajectories with aerial measurements. In SUMO User Conference.
54. Schrank, D., Eisele, B., Lomax, T., & Bak, J. (2015). 2015 urban mobility scorecard.

55. Seliman, S., Sadek, A. W., & He, Q. (2020). Automated vehicle control at freeway lane-drops: a deep reinforcement learning approach. Journal of Big Data Analytics in Transportation, 2(2), 147-166.
56. Shi, J., & Liu, M. (2019). Impacts of differentiated per-lane speed limit on lane changing behaviour: A driving simulator-based study. Transportation research part F: traffic psychology and behaviour, 60, 93-104.
57. Smulders, S. (1990), Control of freeway traffic flow by variable speed signs, Transportation Research Part B: Methodological, 24(2), 111–32.
58. Sunehag, P., Lever, G., Gruslys, A., Czarnecki, W. M., Zambaldi, V., Jaderberg, M., ... & Graepel, T. (2017). Value-decomposition networks for cooperative multi-agent learning. arXiv preprint arXiv:1706.05296.
59. Tajalli, M., & Hajbabaie, A. (2018). Dynamic speed harmonization in connected urban street networks. Computer‐Aided Civil and Infrastructure Engineering, 33(6), 510-523.
60. Talebpour, A., Mahmassani, H. S., & Hamdar, S. H. (2013). Speed harmonization: Evaluation of effectiveness under congested conditions. Transportation research record, 2391(1), 69-79.
61. Tian, D., Wu, G., Boriboonsomsin, K., & Barth, M. J. (2018). Performance Measurement Evaluation Framework and Co-Benefit\/Tradeoff Analysis for Connected and Automated Vehicles (CAV) Applications: A Survey. IEEE Intelligent Transportation Systems Magazine, 10(3), 110-122.
62. Treiber, M., Hennecke, A., & Helbing, D. (2000). Congested traffic states in empirical observations and microscopic simulations. Physical review E, 62(2), 1805.
63. Vahidi, A., & Sciarretta, A. (2018). Energy saving potentials of connected and automated vehicles. Transportation Research Part C: Emerging Technologies, 95, 822-843.
64. Vinitsky, E., Parvate, K., Kreidieh, A., Wu, C., & Bayen, A. (2018, November). Lagrangian control through deep-rl: Applications to bottleneck decongestion. In 2018 21st International Conference on Intelligent Transportation Systems (ITSC) (pp. 759-765). IEEE.
65. Vrbanić, F., Ivanjko, E., Kušić, K., & Čakija, D. (2021). Variable speed limit and ramp metering for mixed traffic flows: A review and open questions. Applied Sciences, 11(6), 2574.
66. Waller, S. T., Ng, M., Ferguson, E., Nezamuddin, N., & Sun, D. (2009). Speed harmonization and peak-period shoulder use to manage urban freeway congestion (No. FHWA/TX-10/0-5913-1). University of Texas at Austin. Center for Transportation Research.
67. Wang, C., Xie, Y., Huang, H., & Liu, P. (2021). A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling. Accident Analysis & Prevention, 157, 106157.

68. Wang, C., Zhang, J., Xu, L., Li, L., & Ran, B. (2019). A new solution for freeway congestion: Cooperative speed limit control using distributed reinforcement learning. IEEE Access, 7, 41947-41957.
69. Wang, M., Daamen, W., Hoogendoorn, S. P., & van Arem, B. (2016). Connected variable speed limits control and car-following control with vehicle-infrastructure communication to resolve stop-and-go waves. Journal of Intelligent Transportation Systems, 20(6), 559-572.
70. Wang, Z., Chen, X. M., Ouyang, Y. & Li, M. (2015), Emission mitigation via longitudinal control of intelligent vehicles in a congested platoon, Computer-Aided Civil and Infrastructure Engineering, 30(6), 490–506.
71. Wu, Y., Tan, H., Qin, L., & Ran, B. (2020). Differential variable speed limits control for freeway recurrent bottlenecks via deep actor-critic algorithm. Transportation research part C: emerging technologies, 117, 102649.
72. Xiao, D., Kang, S., Xu, X., & Shen, Z. (2022). Reinforcement learning based mainline dynamic speed limit adjustment of expressway off‐ramp upstream under connected and autonomous vehicles environment. IET Intelligent Transport Systems, 16(12), 1809-1819.
73. Yu, M., & Fan, W. D. (2019). Optimal variable speed limit control in connected autonomous vehicle environment for relieving freeway congestion. Journal of Transportation Engineering, Part A: Systems, 145(4), 04019007.
74. Zhang, L., Liu, H., Sun, J., & Wang, D. (2015). Variable speed limit strategy to improve the safety and environmental impact of freeway traffic. Paper presented at the 94th Annual Meeting of the Transportation Research Board, Washington, DC, USA.