https://doi.org/10.1016/j.apenergy.2019.114030

# A Car-Following Model for Electric, Connected and Automated Vehicles to Dampen Traffic Oscillations and Improve Energy Consumptions: A Reinforcement Learning Based Approach

Xiaobo Qu[a], Mofan Zhou[b], Yang Yu[a,c], Chin-Teng Lin[d], and Xiangyu Wang[e*]

[a] *Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg 41296, Sweden*
[b] *Tencent Holdings Limited, Shenzhen 518057, China*
[c] *School of Civil and Environmental Engineering, University of Technology Sydney, Sydney 2007, Australia*
[d] *School of Software, University of Technology Sydney, Sydney 2007, Australia*
[e] *School of Civil Engineering and Architecture, East China Jiaotong University, Nanchang 330013, China*

***Abstract*—It has been well recognized that human driver's limits, heterogeneity, and selfishness substantially compromise the performance of our urban transport systems. In recent years, in order to deal with these deficiencies, our urban transport systems have been transforming with the blossom of key vehicle technology innovations, most notably, connected and automated vehicles. In this paper, we develop a car-following model for electric, connected and automated vehicles based on reinforcement learning with the aim to dampen traffic oscillations (stop and go traffic waves) caused by human drivers and improve electric energy consumptions. Compared to classical modelling approaches, the proposed reinforcement learning based model significantly reduces the modelling constraints and has the capability of self-learning and self-correction. Experiment results demonstrate that the proposed model is able to improve travel efficiency by reducing the negative impact of traffic oscillations, and it can also reduce the average electric energy consumptions.**

***Keywords*—Electric Vehicles; Connected and Automated Vehicles; Reinforcement Learning; Car-Following; Machine Learning; Deep Deterministic Policy Gradient; Traffic Oscillations; Energy Consumptions.**

## 1. Introduction

Urbanization is rapidly taking place globally. According to [1], the urbanized population will contribute 66% in 2050. The rapid urbanization unavoidably causes severe transport and mobility challenges, especially in large cities like London, New York and Shanghai. There is no doubt that the transport challenges (safety, congestion, sustainability, etc.) significantly undermine a large city's liveability and the wellbeing of its residents: i) traffic accidents result in 1.25 million fatalities and 50 million injuries worldwide every year and, more importantly, they are the leading causes of death for < 45 years old [2]; ii) gridlocks occur more and more frequently in our urban cities, especially during peak hours; and iii) transport sector contributes over 1/3 of the greenhouse gas (GHG) emissions [3].

It has been well recognized that human driver's limits (e.g. slow reaction time, limited information processing capability), heterogeneity (e.g. different reactions among drivers), and selfishness (non-cooperativeness) substantially compromise the performance of our urban transport systems [4]. Most existing traffic control strategies and technologies (e.g. traffic signal) aim to regulate or control a collective and aggregated group of vehicles, with an attempt to accommodate the aforementioned human driver's deficiencies [5-9]. In order to fully utilize the potential of our urban transport infrastructure, a series of vehicle technology innovations have been proposed in recent years, most notably, connected vehicles, and automated (self-driving) vehicles. Connected vehicles basically enable real time information sharing and communications among individual vehicles and infrastructure control units [10]. Automated vehicles aim to replace a human driver with a robot that constantly receives environmental information via various sensor technologies (as compared to human eyes and ears) [11], and consequently determines vehicle control decisions with proper computer algorithms (as compared to human brains) and vehicle control mechanics [12]. The development of connected and automated vehicles (CAVs) has far outpaced the existing traffic control systems in that individual vehicle can be controlled and regulated in a real-time manner [13]. With these CAVs, individual vehicle based control to fully overcome or minimize the negative effects caused by human driver's limits, heterogeneity, and non-cooperativeness becomes feasible [14-16]. In other words, the traffic flow management can be transformed from a reactive, aggregated/collective, and non-cooperative infrastructure based paradigm to a proactive, disaggregated/individual, and cooperative vehicle based paradigm [12].

A number of studies have been developed with an attempt to modify and improve classical models for controlling CAVs [12, 15, 17-23]. These models have yielded abundant knowledge and control methods in understanding and utilizing this emerging technology in highway traffic management. However, as most of these models were primarily developed based on human-behavioural theories without any room for self-learning and self-corrections, they have limited flexibility and adaptivity, and further modifications and improvements are likely to be constrained by their specific empirical equations.

Machine learning has been widely used in transportation research, such as using artificial neural networks to mimic human driving behaviours [24-32]. However, if CAV driving

[*] Corresponding author. E-mail: xiangyu.wang.perth!@gmail.com

strategies are only developed based on human driving paradigms, it would be hard for CAVs to overcome the intrinsic limitations of human drivers (e.g., proneness to errors, long reaction time, unwillingness to collaborate, etc.), which have been widely criticized as causes to prevailing traffic issues [33-35] and are arguably the challenges that the masterminds behind the concepts of CAV plan to overcome [36, 37]. Therefore, appropriate driving strategies beyond the human driving framework needs to be designed in order to realize the full vision of the future CAV traffic.

Traffic oscillations refer to the stop-and-go driving conditions in congested traffic which typically form bottlenecks of transport infrastructure [32, 38]. With regard to controlling CAVs in terms of dampening or eliminating traffic oscillations, one approach is to create sufficient time buffer or shorten the responding time in traffic oscillations and stabilize overall traffic flow [38, 39]. This idea has also been validated by field experiments [16, 17, 23]. Another approach is to make a driving plan by accessing a future target state from the current state. Ma, et al. [13] developed a trajectory design model to eliminate traffic oscillation by optimizing the motion of CAVs backward from a target driving state in the future. However, this methodology only optimizes the current and future motions of vehicles by taking advantage of the communications between infrastructure and vehicles, and the awareness of a future target state. But in most cases where there is no target state or it is difficult to obtain one, the CAVs cannot plan in advance, and the current state matters more in terms of decision-making. Moreover, such planning methods will need much computation time, as it has to search into the future. Differently, the study in this paper focuses more on obtaining a general CAV model that considers current driving information only. A Reinforcement Learning (RL) approach is applied for this purpose.

A recent breakthrough in RL challenges human in gaming disciplines [40-42]. RL is capable of generating appropriate rules to achieve a certain goal without human supervision. Additionally, RL-based solutions may come from a great number of search attempts in a solution space while human may only be capable of accessing a subset of this space. In this regard, the RL approach can overcome human limitations. As such, a properly designed RL-based model can be an ideal alternative for the design of CAV driving strategies. Zhou and Qu [43] applied one of the RL methods named Deep Q-Network (DQN) [41, 44] and Desjardins and Chaib-Draa [45] applied Policy Gradient (PG) [46, 47] to design a CAV driving controller. The result shows a specific driving strategy can be learned by CAVs by implementing an appropriate reward-guided system. However, the learned CAV acceleration model only works in a discrete action space due to the fact that a discrete action function approximation can simplify the possible action outputs. However, it is impractical that the CAV can only drive with a series of discrete actions. Further, discretized actions in RL require outputting multiple values and choosing the action with the maximum value. When including a great deal of discretized actions, it consumes more computation time compared with outputting a single but continuous action. To the best of our knowledge, there is no research study on using RL to develop car-following models for CAVs in a continuous action space. Thus, in this paper, we extend aforementioned studies to a continuous action space (acceleration space), which resolves the demerits of the existing RL-based models and is more suitable for real-time control of CAVs. Furthermore, in addition to learning from the prior experiences, the proposed model shall be able to mitigate or eliminate current traffic problems, particularly on traffic oscillations or stop-and-go traffic waves.

On the other hand, electric vehicles (EV) have been developing quite fast in recent years in the background of reducing GHG emissions and protecting environment. However, it is always difficult to improve electric energy consumptions of electric, manually-driven vehicles (e-MV) through optimizing human driver behaviours due to their heterogeneity and non-cooperativeness, as mentioned above. Fortunately, the above goal becomes possible if the vehicle involved is an electric CAV (e-CAV) as accurate vehicle controls can be achieved.

In this paper, we develop a novel RL-based, reward-guided car following model for e-CAVs to learn and maximize their accumulative rewards. After the training process, the e-CAVs trained by the proposed model are able to dampen the effect of sudden traffic disturbances. Further simulations under traffic oscillations indicate that the travel efficiency of transport systems as a whole is improved significantly and the average electric energy consumptions are also reduced.

This rest of this paper is organized as follows. Section 2 summarizes the relevant literature review. Section 3 introduces the proposed RL-based methodology and details of the experiment design. Section 4 presents the experiment results and the relevant discussions. Section 5 concludes.

## 2. Literature review

### *2.1. Classical Car-following models*

A classical car-following model is a mathematical expression with respect to how one car follows another. Different expressions have been established from the mid of twentieth century, e.g. the GHR model [48, 49], the CA model [50] and Gipps' model [51]. In order to overcome the deficiencies of the above pioneering models, some other classical models are developed in order to better establish the relationship between vehicle motion and traffic conditions. The Optimal Velocity (OV) model [52-54] determines acceleration rates based on gap distance and velocity. The OV model extracts the gap distance information and converts it into a desired optimal velocity, then compute the acceleration rate from the difference between the desired optimal velocity and the real velocity. Models proposed in [55-58] are modifications of the original OV model for a better performance. Another widely used model is the Intelligent Driver Model (IDM) [59], which decomposes the acceleration into two aspects consisting of a free flow acceleration and a brake deceleration. Some other IDM improvements include the Human Driver Model (HDM) [60, 61], the IDM with Constant-Acceleration Heuristic (CAH) [20], and the IDM for cooperative adaptive cruise control [17].

Although these classical car-following models are initially developed to simulate human driving behaviours, they were also applied to analyse CAV behaviours. Many studies [12, 17, 21, 23, 62] built their CAV models by either improving or modifying an existing classical model. However, most classical models have prescribed model structures and parameter settings that are independent of real-time/historical surrounding traffic conditions as well as prior driving experiences. Therefore, these models may not be flexible enough to describe adaptive CAV behaviours in real-world traffic. In this regard, a mainstream of future CAV control algorithms are learning-based and adaptive to constantly changing sensor feeds [63, 64].

### *2.2. Electric Vehicle (EV)-related technologies*

Electric Vehicle (EV) related technologies [65-73] are being widely studied, tested, and implemented in recent decades due to the urgent call of the transition of energy structure from the high-pollution, non-renewable fossil fuels to the environmental-friendly, renewable energies such as electricity. The history of EV can be dated back to as early as the beginning of 20th Century [73]. One advantage of EV is that regenerative brakings can be perfectly integrated with it, which means the EV can use its motor as a generator during braking maneuvers to transform the kinetic energies of the EV during brakings into electric energy, and thus improve energy utilization rate and reduce energy consumptions [74, 75]. In addition, to evaluate the performance of EVs, many studies have also focused on either estimating EV energy consumptions [76], or evaluating capacity degradations of EV battery cells [77, 78].

### *2.3. Traffic oscillation and CAV-based solutions*

Traffic oscillations are believed to be unavoidable under heavy traffic flow conditions due to the heterogeneity of drivers' responses. Indeed, human being's reaction time causes delay in responding, and this delay can quickly form traffic oscillations in a dynamic traffic condition.

With the increased traffic demands caused by urbanization, traffic oscillations become more and more frequent, which subsequently imposes negative impacts on transport safety, efficiency and sustainability [79]. The formation and propagation mechanisms of traffic oscillations have been intensively investigated during recent years. For example, Li, et al. [80] pointed out that drivers are often forced to be engaged in repeated deceleration-acceleration cycles in a congested highway segment. The trigger of this phenomenon includes ramp-merging, lane changes and changes in roadway geometric features [81].

Inter-Vehicle-Communication (IVC) [82-85] and Cooperative Adaptive Cruise Control (CACC) [12, 17, 22, 23, 62] are the foundation of future CAVs. CAVs can possibly be utilized in a way that is able to optimize the travelling conditions and solve the issue of inefficient driving behaviours. In fact, CAVs could be designed to minimize the negative impact caused by traffic oscillations. To evaluate the performance of CAVs in this circumstance, simulations conducted in [12, 86] show that CAVs can perform beyond human drivers when having a specific design corresponding to a typical circumstance such as highway-merging.

Modelling CAVs to maximize travel efficiency in traffic oscillations is difficult due to many constraints and unknown parameters in classical car-following models. The state-of-art machine learning techniques such as the reinforcement learning approach make it possible to simplify and solve this problem.

### *2.4. Reinforcement learning*

A standard reinforcement learning framework consists of interactions between an agent and an environment. At each time $t$, the agent receives an observation $s_t$, takes an action $a_t$ based on $s_t$ and receives a reward $r_t$ from the environment. In a typical RL system, the observation obtained from the environment is called state. The behaviour of an agent is defined by a policy $\pi$. Under the policy $\pi$, the agent takes current state $s_t$ and output a probability distribution $P(a) = \pi(s_t)$ over the action set $a$. For the environment, it provides the transition dynamics $p(s_{t+1} \mid s_t, a_t)$ and reward $r(s_t, a_t)$ for the agent who takes action $a_t$ at state $s_t$ at time $t$.

In a reinforcement learning framework, an agent learns to match the future reward with its experience. A discounting factor $\gamma \in [0,1]$ is applied to compute a return which is defined as the sum of discounted future reward $R_t = \sum_{i=t}^{T} \gamma^{(i-t)} r(s_i, a_i)$. The goal in reinforcement learning is to learn a policy that maximizes the expected return from the current state. Many approaches in reinforcement learning use the Bellman Equation (Eq. (1)) to represent the recursive relationship in the future return.

$$Q^{\pi}(s_t, a_t) = \mathrm{E}_{r_t, s_{t+1} \sim E}\left[ r(s_t, a_t) + \gamma \mathrm{E}_{a_{t+1} \sim \pi}[Q^{\pi}(s_{t+1}, a_{t+1})] \right] \tag{1}$$

where $Q^{\pi}$ is the state-action value based on a stochastic policy $\pi$. For a deterministic target policy $\mu$ instead of $\pi$, the inner expectation disappears, and the above equation can be rewritten as

$$Q^{\mu}(s_t, a_t) = \mathrm{E}_{r_t, s_{t+1} \sim E}\left[ r(s_t, a_t) + \gamma Q^{\mu}\left(s_{t+1}, \mu(s_{t+1})\right) \right] \tag{2}$$

With this deterministic policy $\mu$, the agent is possible to learn $Q^{\mu}$ off-policy, which makes use of the state-action-reward pairs generated from other agents or from this agent but a different time. The Q-learning algorithm [87] is commonly used as an off-policy algorithm by considering the greedy policy $\mu(s) = \arg\max_a Q(s, a)$. Utilizing a neural network as a function approximator is a shortcut to many complex RL problems. Thus, a policy $\mu$ can be parameterized by a neural network $\theta^Q$, which can be optimized by minimizing the loss:

$$L(\theta^Q) = \mathrm{E}_{s_t, a_t, r_t}\left( \left( Q(s_t, a_t \mid \theta^Q) - y_t \right)^2 \right) \tag{3}$$

where

$$y_t = r(s_t, a_t) + \gamma Q\left(s_{t+1}, \mu(s_{t+1}) \mid \theta^Q\right). \tag{4}$$

The $y_t$ in Eq. (3) and (4) is typically recognized as a Q-target in RL, and it is also dependent on $\theta^Q$. Due to the unstable and non-convergence problem related to the use of complicated nonlinear function approximators, researchers and practitioners in the past rarely apply a large scale or nonlinear function approximator for evaluating the $Q$. In recent years, Mnih, et al. [41] proposed a variation of the Q-learning algorithm named DQN which learns to play video games from pixel inputs. After that, Lillicrap, et al. [88] adapted the concept in DQN and applied it with Deterministic Policy Gradient (DPG) [89], and renamed the DPG as Deep Deterministic Policy Gradient (DDPG). Their results show DDPG is able to achieve a better control when involving a continuous action domain.

In transportation research, a few RL approaches have been applied to model human driving behaviours [90] or CAVs [43, 45]. Although (e-)CAV control should be considered as a continuous action problem, all existing models treat it as a discrete action problem, as mentioned before. To bridge this gap, we propose a DDPG-based car following model and apply it for controlling e-CAV acceleration rate.

## 3. Model development and experimental design

### *3.1. Deep deterministic policy gradient*

Lillicrap, et al. [88] proposed Deep Deterministic Policy Gradient (DDPG). DDPG belongs to a Policy Gradient (PG) [46, 47] that is suitable to perform in continuous action spaces. However, the basic PG is limited by an episodic update rule – updating the behaviour policy must be at the end of this episode [91]. The use of the Actor-Critic method [92] and a function approximator for PG [91] dramatically improves its performance in terms of speeding up training process and increasing the ability in non-linearity.

The usual stochastic policy gradient such as Actor-Critic may not be efficient when learning in an environment that only needs a deterministic behaviour policy. Therefore, Silver, et al. [89] proposed a Deterministic Policy Gradient (DPG) method that leads to an efficiency improvement in training. As the DPG is an upgraded version of Actor-Critic, the updating procedure can be separated into two parts: the update for the actor and the update for the critic. These updates aim to maximize the average reward that an agent receives, so the objective function for this purpose can be shown in Eq. **Error! Reference source not found.**.

$$J_\beta(\mu^\theta) = \int_S \rho^\beta(s) Q^\mu(s, \mu^\theta(s)) ds \qquad (5)$$

where $\mu^\theta$ is the target policy parameterized by $\theta$, $\rho^\beta(s)$ denotes the behaviour policy at state $s$, $Q^\mu(s, \mu^\theta(s))$ represents the state-action value or Q-value evaluated from a critic and its action which comes from the target policy. We can rewrite this objective function to obtain the parameter update in Eq. **Error! Reference source not found.**.

$$\nabla_\theta J_\beta(\mu^\theta) = \mathrm{E}_{s\sim\rho^\beta}\left[\nabla_\theta \mu^\theta(s) \nabla_a Q^\mu(s,a)\big|_{a=\mu^\theta(s)}\right] \qquad (6)$$

This equation gives the off-policy deterministic policy gradient. It indicates that the actor parameters updated by moving its parameters in the direction of the critic can maximize its Q-value. While on the critic side, its update is shown in Eq. **Error! Reference source not found.** and Eq. **Error! Reference source not found.**. To sum up, the actor and critic update in the DPG can be unified as following:

$$\delta_t = r_t + \gamma Q^w(s_{t+1}, \mu^\theta(s_{t+1})) - Q^w(s_t, a_t) \qquad (7)$$

$$w_{t+1} = w_t + \alpha_w \delta_t \nabla_w Q^w(s_t, a_t) \qquad (8)$$

$$\theta_{t+1} = \theta_t + \alpha_\theta \nabla_\theta \mu^\theta(s_t) \nabla_a Q^w(s_t, a_t)\big|_{a=\mu^\theta(s)} \qquad (9)$$

where $\delta_t$ denotes the TD-error in one-step update; $r_t$ represents the reward received at time $t$ when taking action $a_t$ and the state changes from $s_t$ to $s_{t+1}$; $Q^w$ is the estimated state-action value or Q-value by a function approximator (typically a nonlinear neural network) parameterized by $w$; $\alpha_w$ and $\alpha_\theta$ are the learning rates of actor and critic, respectively; $\mu^\theta(s_t)$ denotes the target policy parameterized by $\theta$.

Although this DPG algorithm successfully improves the learning efficiency, the convergence and stability problems still exist due to the on-policy updating and combining a nonlinear function approximator. In other words, a correlation between two successive state updates introduces unstable issues for challenging problems. These problems can be solved by bringing the advantages in DQN [88].

The DDPG is a combination of DQN and DPG in terms of creating a memory buffer and target networks for DPG in order to de-correlate successive updates. Both of the actor and critic in DDPG have an evaluating network ($\mu^\theta$ and $Q^w$) and a target network ($\mu^{\bar{\theta}}$ and $Q^{\bar{w}}$). The parameters update in a target network is delayed for the purpose of de-correlating successive updates. The formal updating rule is shown as the following equations.

$$\delta_t = r_t + \gamma Q^{\bar{w}}(s_{t+1}, \mu^{\bar{\theta}}(s_{t+1})) - Q^w(s_t, a_t) \qquad (10)$$

$$w_{t+1} = w_t + \alpha_w \delta_t \nabla_w Q^w(s_t, a_t) \qquad (11)$$

$$\theta_{t+1} = \theta_t + \alpha_\theta \nabla_\theta \mu^\theta(s_t) \nabla_a Q^w(s_t, a_t)\big|_{a=\mu^\theta(s)} \qquad (12)$$

The $\bar{w}$ and $\bar{\theta}$ in $Q^{\bar{w}}$ and $\mu^{\bar{\theta}}$ are then assigned to $w$ and $\theta$ after a particular amount of time steps. In addition, the state-action-reward transitions are stored in a memory buffer and randomly selected during the update process.

### *3.2. Training environment and parameter settings*

The training procedure for our proposed DDPG-based car following model is demonstrated in **Fig. 1** as e-CAVs can interact with the driving environment in real-time and simultaneously collect local and global traffic information, including their own speed, gap distance and the relative speed with their preceding vehicle. The collected information is stored in a memory buffer in the RL (DDPG) system. A batch of experiences randomly sampled from this memory buffer is used to update the actor and the critic at each time step. The actor is responsible for choosing an appropriate acceleration for all e-

CAVs and the inputs and output of the actor can be simplified as $a \leftarrow Actor(\Delta v, \Delta x, v)$, where $\Delta v$ and $\Delta x$ denote relative speed and distance gap with its preceding vehicle respectively, and $v$ denotes its own speed.

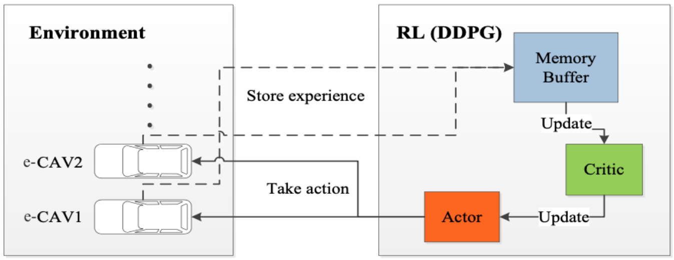

**Fig. 1.** Learning diagram of the DDPG-based car following model.

Note that there is only one actor in this system, thus, all e-CAVs share the same DDPG model, and each of them contributes equally to the system update. A virtual training environment shown in **Fig. 2** is built. We choose a vehicle platoon of 10 e-CAVs that forms a circular driving loop in order to simulate the oscillation effect in a consecutive traffic flow.

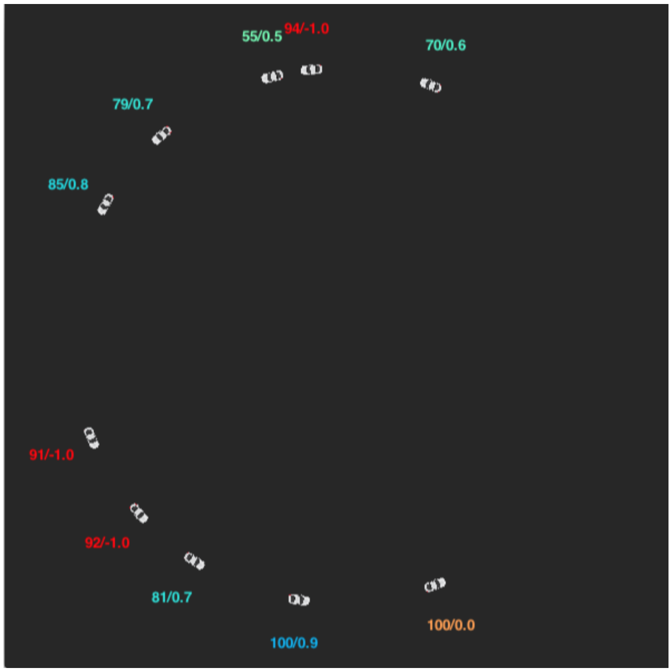

**Fig. 2.** 10 e-CAVs follow one another in a circular loop. The numbers next to each e-CAV refer to its speed and reward, and these numbers are coloured based on the reward. A higher reward turns to blue and a lower one turns to red.

The e-CAVs are trained on 2,000 episodes, each of which consists of 300 time steps. Each time step is equivalent to a 0.1 second updating interval. We initialize each episode with randomness in order to reduce the sensitivity of the final model and all initialization setups are described as follows.

- Updating interval: 0.1 second;
- Vehicle acceleration range: $[-5\ m/s^2, 3\ m/s^2]$;
- Vehicle length: 5 meters.
- Fix the initial speed of leading vehicle in each episode as 100 $km/h$;
- Initial distance gaps for subsequent vehicles are randomly selected in a range of $[15\ m, 50\ m]$;
- Initial speeds for subsequent vehicles are randomly selected in a range of $[40\ km/h, 130\ km/h]$.

The leading vehicle is not allowed to accelerate or decelerate during the whole episode, while other vehicles in this 10-vehicle platoon adjust their acceleration rate by the actor in the DDPG model. The final goal for each episode is to stabilize following condition from a disordered initialization. This is an indirect but faster training procedure for e-CAVs to learn how to handle traffic oscillations compared with randomly disturbing the behaviours of the leader of the platoon.

The hyper-parameters for the DDPG model are carefully selected: we select $\alpha_w = 1e-5$ and $\alpha_\theta = 2e-5$ in Eq. **Error! Reference source not found.** and **Error! Reference source not found.** as the learning rates for critic and actor respectively. Further, the discount factor $\gamma = 0.9$ in Eq. **Error! Reference source not found.**. In order to cover the experience in serval episodes, we choose the memory capacity as 100,000 transitions. The update frequency for target networks $Q^{\bar{w}}$ and $\mu^{\bar{\theta}}$ are selected as 1,000 time steps. The evaluation networks $Q^w$ and $\mu^\theta$ are updated each step using RMSprop [93] with a batch size of 64.

### *3.3. Reward function design*

In practice, a carefully designed reward function could result in a particular solution. In a car-following problem, without loss of generality, we apply a time-headway-based reward function such as the one mentioned in [43, 45]. By doing so, one can easily learn a reasonable car-following rule. With the learned rule, an e-CAV can drive with collision-free behaviour, which, however, may not also lead to a travel efficiency improvement. This makes us to reconsider and design a reward function from the perspective of optimizing traffic flow dynamics.

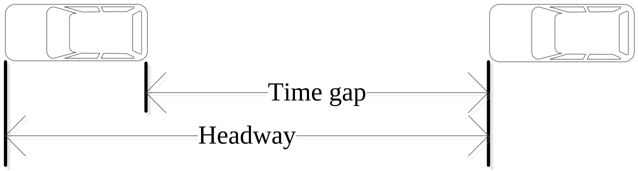

**Fig. 3.** Illustration of time gap and time headway between successive vehicles.

The time headway is essential for a safety concern, but the time gap shown in **Fig. 3** is more crucial since it excludes the impact of preceding vehicle's length. Therefore, in the following section, we adopt time gap instead of time headway as a part of the reward function illustrated in **Fig. 4**.

The reward function consists of two aspects: speed and time gap, as shown in **Fig. 4**. First, we define a maximum speed of 110 $km/h$. In a homogeneous traffic follow, there is no doubt that if all e-CAVs are travelling with higher speeds, the entire travel efficiency will increase, and the entire/individual electric energy consumptions will be reduced as well due to less travel times and less congestions expected (less repeated deceleration-acceleration maneuvers). Further, in a traffic oscillation scenario, an e-CAV stabilizing its following condition while maintaining a higher speed should be rewarded due to the same reason. Thus, within the range of 0 $km/h$ to 110 $km/h$, the reward is set monotonically increasing from 0 to 1. In this research, we compare three types of monotonic reward curve including “Linear”, “Concave” and “Convex”, and the results can be found in the next section. Whenever the speed exceeds the maximum speed, a reward of -1 is assigned as a punishment to avoid over-speeding. Further, whenever the time gap is less than a minimum safe time-gap for e-CAVs (0.6 second is adopted, as was found in [17]), a reward of -1 is given to reduce the risk of collision.

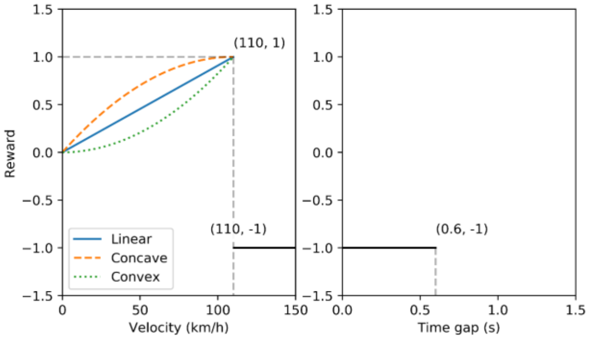

**Fig. 4.** The design of reward function.

## 4. Results and discussion

### 4.1. Training result

In order to validate the robustness of the DDPG model, we conduct six epochs with different random seeds and plot the moving averaged episode reward in Fig. 5. In a typical RL training process, the reward for each episode can be noisy. Therefore, we apply a moving averaging method to show the tendency of the total reward change, which is computed as

$$R_t \leftarrow 0.99R_{t-1} + 0.01R_t$$

where $R_t$ denotes the reward at time step $t$.

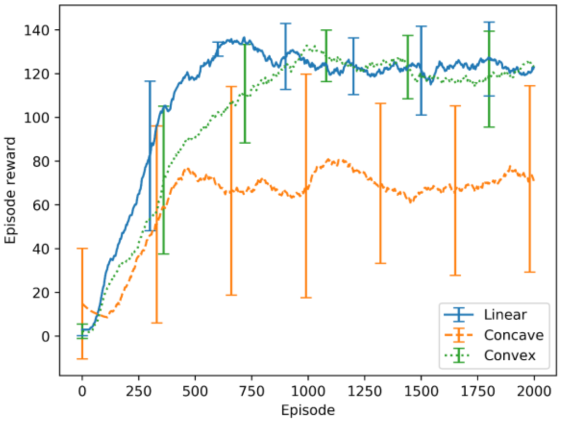

**Fig. 5.** Moving averaged episode reward comparison.

The accumulated reward in an episode grows up quickly from the beginning of training. The linear reward function achieves 120 episode rewards faster than other reward curves. The concave reward curve introduces higher variance in training and has only about half of the episode rewards at the end of training compared with other reward curves. In conclusion, both linear and convex reward curves are acceptable for training e-CAVs and the linear curve has the fastest convergence. Therefore, in the rest of the research, we select the DDPG model trained by the linear reward function.

### 4.2. Comparing with electric, manually-driven vehicles (e-MVs) as per travel efficiencies

Many car-following models have been developed for modelling MVs. The IDM [59] is one of the classical car-following models that has been intensively studied [16, 20, 60, 61, 94]. Therefore, in the following sections, we compare e-CAVs trained by the DDPG model with e-MVs controlled by the IDM in terms of both travel efficiencies and electric energy consumptions in various scenarios involving traffic disturbances/oscillations. The default IDM parameters in [59] are used for the rest of evaluations as they perform well under not only free-flow but also congested flow traffic [12, 32].

#### 4.2.1. High speed scenario

This test is to compare the performance between DDPG-based e-CAVs (e-CAV platoon) and IDM-based e-MVs (e-MV platoon) in handling a disturbance encountered when they are travelling in high speeds. We fix the leading vehicle’s behaviour by a sequence of acceleration patterns: 1) constant speed (100 $km/h$) for 100 seconds; 2) decelerate ($-2\ m/s^2$) for 15 seconds (if the speed is decreases to zero, vehicle stops and the deceleration rate is set to zero); 3) accelerate ($1\ m/s^2$) to the original speed (100 $km/h$); 4) constant speed (100 $km/h$) for 200 seconds. 50 following e-CAVs and 50 following e-MVs are generated respectively with a uniformly 2 seconds initial headway and 100 $km/h$ initial speed.

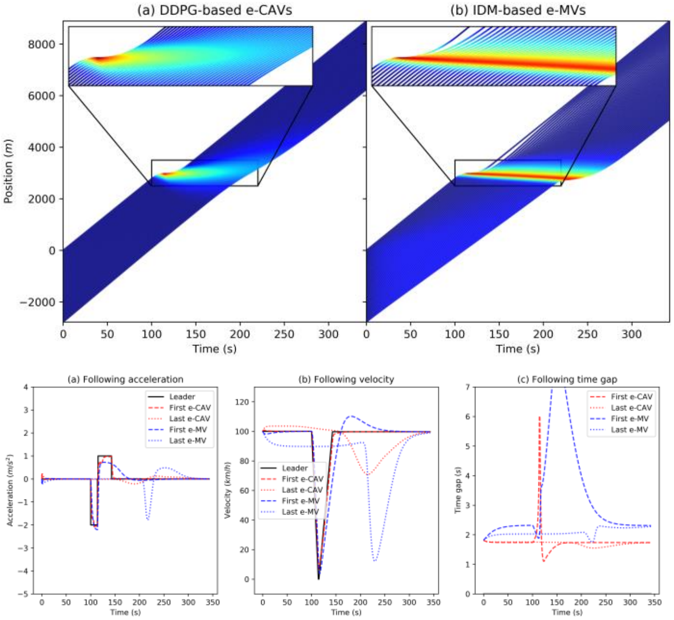

**Fig. 6.** Comparison of traffic oscillations under a high speed scenario.

We plot the simulated trajectories and travelling details of the leading vehicle, the first and last follower from e-CAV platoon and e-MV platoon respectively in **Fig. 6**. Based on the simulation results, it is clear that the disturbance caused by the leading vehicle creates a series of chain reactions in both e-CAV and e-MV platoons. Specifically, an obvious propagative oscillation is observed throughout the whole e-MV platoon. In contrast, in the e-CAV platoon, the disturbance quickly dissipates and the oscillation gradually disappears. Additionally, the acceleration, speed and time gap details also draw the same conclusion. The acceleration and speed details indicate that the first e-CAV follower is more responsive to the changes in its following condition, which results in a faster stabilization. The time gap results indicate that an DDPG-based e-CAV tends to maintain a smaller time gap compared with an IDM-based e-MV. The comparison of travel efficiencies of e-CAVs and e-MVs in this test is quantified in Table 1.

### 4.2.2. *Low speed scenario*

Since the training environment for the DDPG model is set in a high-speed condition, it is also necessary to test the model performance under a low speed condition to evaluate the robustness of the trained model. We run another test by adopting the same testing configuration as the last test but with an 40 $km/h$ initial speed for all vehicles.

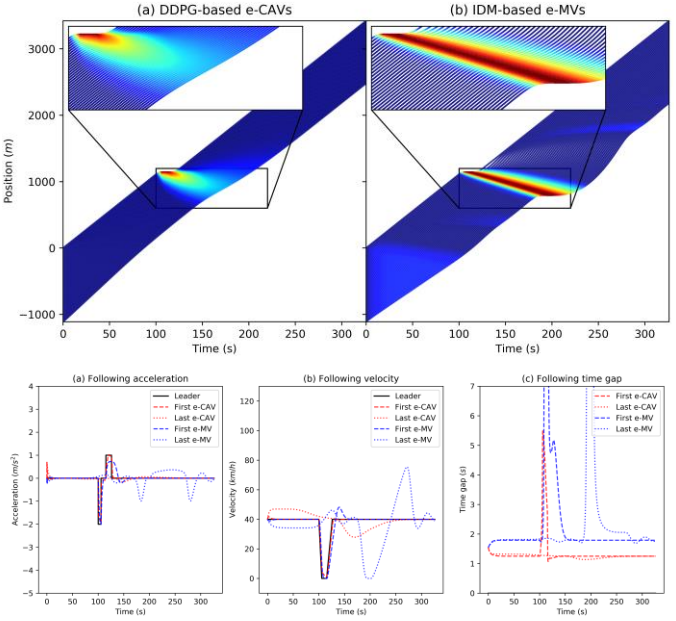

**Fig. 7.** Comparison of traffic oscillations under a low speed scenario.

Although the overall speed condition differs from the last test, the results shown in Fig. 7 draw exactly the same conclusion as the last test. We also listed the travel efficiencies of both models (DDPG model and IDM) in Table 1.

**Table 1**
Travel efficiency comparison of DDPG-based e-CAVs and IDM-based e-MVs

| | Vehicle type | Average travel time (*min/km*) | Time mean speed (*km/h*) |
|---|---|---|---|
| High speed | e-CAV | 0.64 | 94.08 |
| | e-MV | 0.69 | 87.60 |
| Low speed | e-CAV | 1.55 | 38.73 |
| | e-MV | 1.62 | 37.05 |
| Leader Stopping | e-CAV | 0.69 | 86.90 |
| | e-MV | 0.79 | 75.93 |

### 4.2.3. *Leader stopping scenario*

Traffic oscillations often consist of stop-and-go phases. In this section, we evaluate the trained model under a long stopping phase followed by an acceleration phase. The behaviour of the first vehicle is fixed by: 1) constant speed (100 $km/h$) for 3 seconds; 2) decelerate ($-4\ m/s^2$) for 30 seconds (if the speed decreases to zero, vehicle stops and the deceleration rate is set to zero); 3) accelerate ($2\ m/s^2$) to the original speed (100 $km/h$); 4) constant speed (100 $km/h$) for 200 seconds. 50 following e-CAVs/e-MVs are initialized by the same configuration as the first test.

From the result shown in **Fig. 8**, the e-CAV platoon controlled by the DDPG model successfully eliminates the leader stopping effect. In contrast, traffic oscillations in the e-MV platoon propagate to the last vehicle in the platoon.

Considering smoothing traffic oscillations by ramp metering or traffic light control on this oscillated flow, the lengths of time intervals needed (**Fig. 9**) for the last vehicle are 34.4 seconds and 72.8 seconds for e-CAVs and e-MVs respectively. If it is measured by distance, the buffer distances are 955.3 meters and 2023.0 meters respectively, which indicates an over 100% efficiency improvement for an oscillation-free e-CAV controlled by the DDPG model than an IDM-based e-MV.

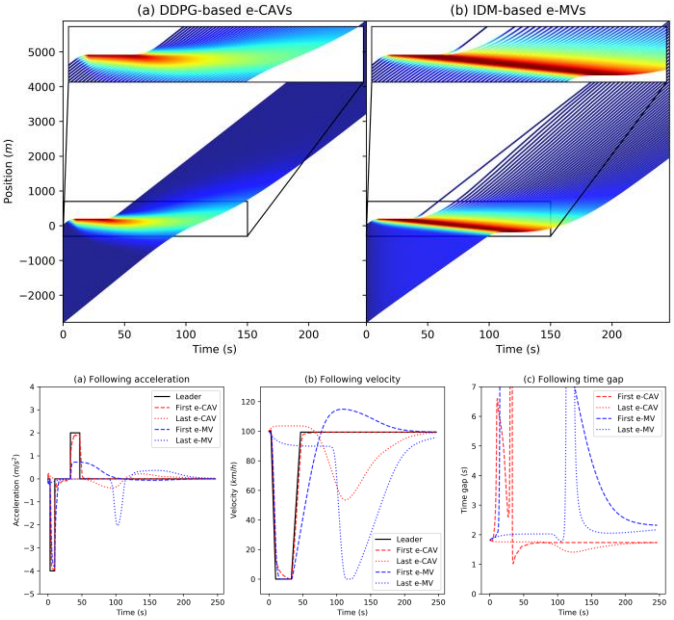

**Fig. 8.** Comparison of traffic oscillations under a leader stopping phase.

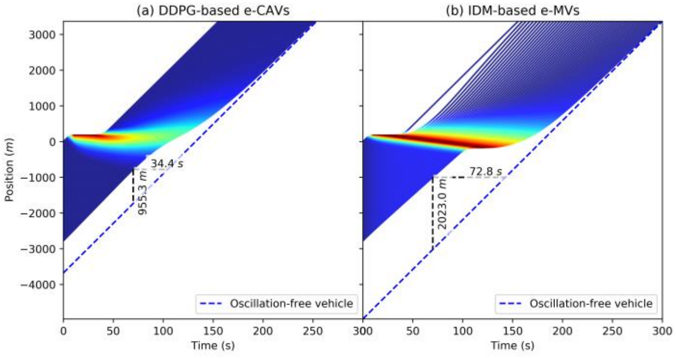

**Fig. 9.** Buffer time and distance for the last e-CAV/e-MV.

A quantified comparisons of travel efficiencies between vehicles controlled by both models under the leader stopping effect is available in Table 1. Further, we also cross compare both models under all above three scenarios by scaling down all the values in Table 1 based on the performance of e-MVs and use the performance of e-MVs as a baseline. The result is displayed in Fig. 10 and it indicates that on average, e-CAVs controlled by the proposed DDPG model outperform e-MVs controlled by IDM in all cases involving traffic oscillations/disturbance (up to 14.4% more efficient on average). And the efficiency improvement will further amplify towards the end of vehicle platoon (up to 26.5% more efficient when only comparing the last e-CAV/e-MV in the vehicle platoon), which further validates the better ability of the proposed DDPG car following model in dampening/dissipating traffic oscillations/disturbance. Note that the above result only includes a one-off traffic disturbance, a greater travel efficiency improvement will be expected in a real, congested road.

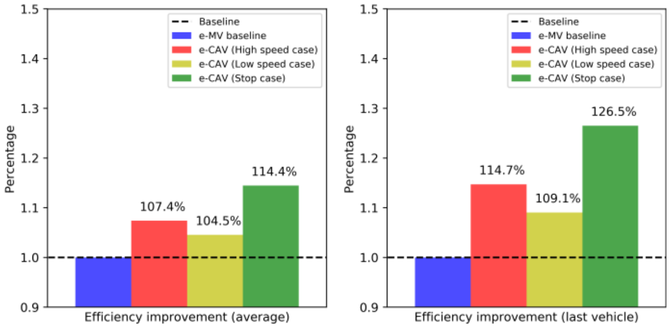

**Fig. 10.** Efficiency comparison against the baseline of e-MV.

### 4.2.4. *Mixed traffic flow scenario*

There is no doubt that (e-)CAVs will soon share roads with (e-)MVs. As such, we also test the performance of the proposed model in a mixed traffic flow consisting of e-CAVs and e-MVs. As can be seen in Figs 11 and 12, with an increase in e-CAV penetration rate (an increase in the proportion of e-CAVs in the mixed flow), the mixed flow can better accommodate traffic disturbances/oscillations.

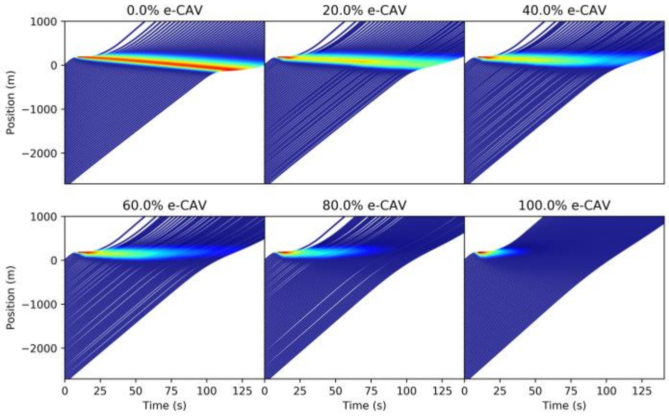

**Fig. 11.** Comparison of traffic oscillations in various e-CAV penetration rates.

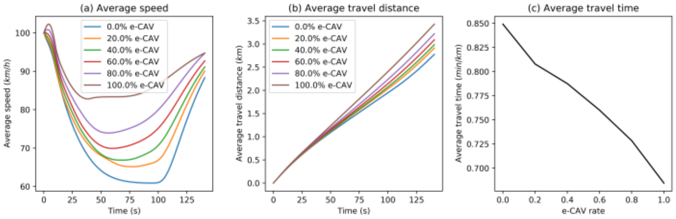

**Fig. 12.** Comparison of travel efficiencies in various e-CAV penetration rates.

Table 2 quantifies the model performances under different e-CAV penetration rates. Note that the column 'Travel efficiency improvement' represents the percentage of an average speed increment based on the average speed under 0% e-CAVs.

**Table 2**
Comparison of average travel efficiencies in different e-CAV penetration rates

| e-CAV rate | Average speed (*km/h*) | Average travel distance (*km*) | Average travel time (*min/km*) | Travel efficiency improvement |
|---|---|---|---|---|
| 0% | 71.12 | 2.77 | 0.85 | 0.00% |
| 20% | 74.59 | 2.91 | 0.81 | 4.89% |
| 40% | 76.42 | 2.98 | 0.78 | 7.45% |
| 60% | 79.11 | 3.09 | 0.76 | 11.23% |
| 80% | 82.49 | 3.22 | 0.73 | 15.99% |
| 100% | 87.78 | 3.43 | 0.68 | 23.43% |

### 4.3. *Comparing with electric, manually-driven vehicles (e-MVs) as per electric energy consumptions*

In this section, we attempt to compare the electric energy consumptions of e-CAVs and e-MVs in the aforementioned traffic scenarios. To enable a fair comparison, a reasonable assumption is made that all e-CAVs and e-MVs involved share exactly the same physical and aerodynamic properties. In other words, we assume that all e-CAVs and e-MVs in this research are equivalent in terms of calculating energy consumptions. Without loss of generality, we predefine a set of physical and aerodynamic property values of a typical electric vehicle based on the data from [74, 76] to calculate the energy consumptions, and the details are listed as follows:

- Mass of vehicle $m$: 2575 $kg$;
- The frontal area of vehicle $A_f$: 2.5 $m^2$;
- Rolling resistance of vehicle tyre with road surface $C_r$: 0.01;
- Aerodynamic drag coefficient of vehicle $C_D$: 0.3;
- Vehicle battery type: Li-Ion battery cells;
- Rated battery voltage of vehicle $U$: 316.8 $V$;
- Rated battery capacity of vehicle $Q$: 252.525 $Ah$;
- Electric motor (battery cell) efficiency of vehicle $\eta_m$: 0.9;
- Generator efficiency of vehicle $\eta_g$: 0.85, assuming that all brakings of the e-CAV/e-MV are energy regenerative brakings and the generatory efficiency is a constant value;
- Air mass density $\rho_{air}$: 1.2256 $kg/m^3$;
- Gravitational acceleration $g$: 9.8066 $m/s^2$.
- Road slope $\alpha$: 0;

Based on the above assumptions and property values, the instant traction force or braking force $F_w(t)$ of an e-CAV or e-MV at a specific time step $t$ can be calculated by the following equations [74, 95]:

$$m \cdot a(t) = F_w(t) - F_{air}(t) - F_r(t) - F_G(t) \quad (13)$$

$$F_{air}(t) = \frac{1}{2}\rho_{air} A_f C_D v(t)^2 \quad (14)$$

$$F_r(t) = mgC_r cos\alpha \quad (15)$$

$$F_G(t) = mgsin\alpha \quad (16)$$

where Eq. (13) is the fundamental dynamic model of vehicle in the moving direction; $a(t)$ and $v(t)$ are the acceleration and speed of the vehicle at time step $t$; $F_{air}(t)$, $F_r(t)$, and $F_G(t)$ refer to the air drag force, rolling resistance force, and gravity force the vehicle suffers at time step $t$, respectively. Then, given that all the brakings of the e-CAVs/e-MVs are energy regenerative, the instant power output or recovered energy input $p_b(t)$ of the vehicle at time step $t$ can be calculated by the following equation [74, 76, 95]:

$$p_b(t) = \begin{cases} F_w(t)v(t)/\eta_m\,, & if\ F_w(t) \geq 0 \\ F_w(t)v(t)\eta_g, & if\ F_w(t) < 0 \end{cases} \quad (17)$$

where a positive $p_b(t)$ represents an instant power output at time step $t$ while a negative $p_b(t)$ represent an instant recovered energy input at time step $t$, respectively. Finally, the total electric energy consumption $E$ of the vehicle over a specific travel distance $s$ $(= s_d - s_o)$ can be calculated by:

$$E = \int_{s_o}^{s_d} \frac{1}{v(t)} p_b(t) ds \quad (18)$$

And the average energy consumption per kilometer (*Ah/km*) of this vehicle can be easily acquired accordingly.

The average energy consumption per kilometer of all the DDPG-based e-CAVs and IDM-based e-MVs in the above high speed, low speed, and leader stopping secnarios are summarized in Table 3, along with the corresponding standard deviations. And the average energy consumption per kilometer of each of these e-CAVs and e-MVs are displayed in Fig. 13.

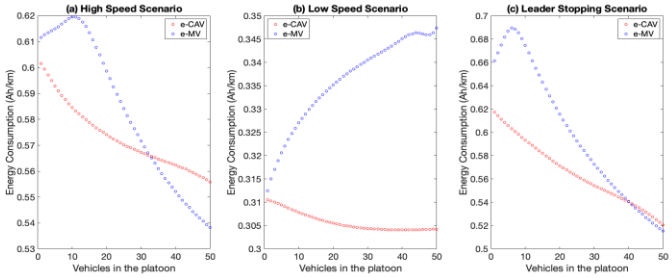

**Fig. 13.** Comparison of electric energy consumptions of vehicles in DDPG-based e-CAV platoon and IDM-based e-MV platoon in different scenarios

**Table 3**
Comparison of electric energy consumptions between DDPG-based e-CAVs and IDM-based e-MVs in different scenarios

| | Vehicle type | Avg energy consumption (*Ah/km*) | Standard deviation (*Ah/km*) | Energy consumption improvement |
|---|---|---|---|---|
| High speed | e-CAV | 0.5729 | 0.0122 | 1.70% |
| | e-MV | 0.5828 | 0.0287 | |
| Low speed | e-CAV | 0.3057 | 0.0020 | 9.00% |
| | e-MV | 0.3360 | 0.0094 | |
| Stopping effect | e-CAV | 0.5649 | 0.0269 | 5.58% |
| | e-MV | 0.5983 | 0.0574 | |

It is easy to observe from Fig. 13 that, in most times, the average energy consumption of each DDPG-based e-CAV in the e-CAV platoon is lower than that of the corresponding IDM-based e-MV in the e-MV platoon, which is particularly obvious in both low speed and leader stopping scenarios. The above finding is further validated in Table 3, where the highest average energy consumption improvement of DDPG-based e-CAV is identified as 9.00%, which is achieved in the low speed scenario.

In addition, we can also find from Table 3 that the standard deviations of average energy consumptions of all the 50 e-CAVs under all the three scenarios are also substantially smaller than the corresponding deviations of the 50 e-MVs, which further proves that the proposed DDPG model can better dampen traffic oscillations and faster stabilize vehicle behaviours in a platoon level. This can also be concluded from Fig. 3 where the energy consumptions of a single e-CAV in all three scenarios smoothly decrease towards the end of platoon, but this is not the case for e-MV platoon.

Finally, the average energy consumptions per vehicle (*Ah/km*) under different e-CAV penetration rates in the mixed traffic flow scenario are also calculated and compared in Table 4. The column 'Energy consumption improvement' represents the percentage of energy consumption decrement based on the average energy consumption under 0% e-CAVs. It is easy to conclude that the average energy consumption per vehicle in the traffic flow indeed reduces with the introduction of e-CAVs into the flow. Though the average energy consumption improvements are not obvious (up to around 3.5% improvement), these improvements are still very meaningful since they are simultaneously achieved along with the substantial improvements in vehicle travel efficiencies (up to 23.43% improvement when e-CAV penetration rate reaches 100%, as was summarized in Table 2) under traffic disturbances/oscillations. Moreover, as can be seen from Fig. 11 and 12, all vehicles in the mixed traffic flow went through smaller degree, less frequent deceleration-acceleration maneuvers (less short-term charge-discharge cycles to the battery cells) with the increase of e-CAV penetration rate, which is beneficial in slowing the battery degradation process from a long-term perspective [77, 78]. Besides, it is worth noting that the standard deviation of average energy consumption per vehicle in the mixed flow scenario substantially reduces when e-CAV penetration rates increases from 80% to 100%, which further validated the negative impact of human driver heterogeneity, as mentioned in the beginning of this paper. By contrast, the proposed DDPG car following model can enable individual vehicle (e-CAV) to behave smoother and more stable in a platoon, which is the reason why both vehicle travel efficiency and energy consumption can be improved even under traffic disturbances/oscillations.

**Table 4**
Comparison of electric energy consumption per vehicle in a mixed traffic flow under different e-CAV penetration rates

| e-CAV rate | Avg energy consumption (*Ah/km*) | Standard deviation (*Ah/km*) | Energy consumption improvement |
|---|---|---|---|

| | | | |
|---|---|---|---|
| 0% | 0.4983 | 0.1546 | 0% |
| 20% | 0.4829 | 0.1538 | 3.09% |
| 40% | 0.4811 | 0.1505 | 3.45% |
| 60% | 0.4845 | 0.1431 | 2.77% |
| 80% | 0.4898 | 0.1275 | 1.71% |
| 100% | 0.4904 | 0.0752 | 1.59% |

At last, it is also worth mentioning that since all the above energy consumption tests are conducted based on a very ideal, fully regenerative braking paradigm with satisfactory generator efficiency (0.85), the energy consumption improvements of the DDPG-based e-CAVs in all the above scenarios are expected to be more significant in reality.

## 5. Conclusions

e-CAVs can not only free human from driving, but also be considered as an optimization tool for improving traffic operation through, for instance, dampening or even eliminating traffic oscillations, and an effective approach to reduce GHG emissions. The design of e-CAVs should focus on not only levels of automation from the perspective of vehicle manufacturers, but also efficient and energy-saving traffic operations from the perspective of transport managers/users. This paper provides an RL (DDPG)-based car following model for e-CAVs other than following the template of classical car-following models. In doing so, the e-CAV's driving rules are no longer constrained by the physical formworks in classical car-following models. By generating an appropriate e-CAV driving strategy using the RL (DDPG) approach, the e-CAVs can learn to drive with the capabilities of both improving the travel efficiency of the transport system as a whole and reducing average electric energy consumptions under various traffic disturbance/oscillation scenarios.

## References

[1] (2014). *World's population increasingly urban with more than half living in urban areas*. [Online] Available: http://www.un.org/en/development/desa/news/population/world-urbanization-prospects-2014.html

[2] (2017). *Road traffic injuries*. [Online] Available: http://www.who.int/mediacentre/factsheets/fs358/en/

[3] (2017). *Sources of Greenhouse Gas Emissions*. [Online] Available: https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

[4] X. Qu, J. Zhang, and S. Wang, "On the stochastic fundamental diagram for freeway traffic: Model development, analytical properties, validation, and extensive applications," *Transportation Research Part B: Methodological*, vol. 104, no. Supplement C, pp. 256-271, 2017, doi: https://doi.org/10.1016/j.trb.2017.07.003.

[5] M. Papageorgiou and A. Kotsialos, "Freeway ramp metering: an overview," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 4, pp. 271-281, 2002, doi: 10.1109/TITS.2002.806803.

[6] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and W. Yibing, "Review of road traffic control strategies," *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2043-2067, 2003, doi: 10.1109/JPROC.2003.819610.

[7] X. Qu, S. Wang, and J. Zhang, "On the fundamental diagram for freeway traffic: A novel calibration approach for single-regime models," *Transportation Research Part B: Methodological*, vol. 73, no. Supplement C, pp. 91-102, 2015, doi: https://doi.org/10.1016/j.trb.2015.01.001.

[8] X. Qu and S. Wang, "Long-Distance-Commuter (LDC) Lane: A New Concept for Freeway Traffic Management," *Computer-Aided Civil and Infrastructure Engineering*, vol. 30, no. 10, pp. 815-823, 2015.

[9] D. Zhao, Y. Dai, and Z. Zhang, "Computational Intelligence in Urban Traffic Signal Control: A Survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 485-494, 2012, doi: 10.1109/TSMCC.2011.2161577.

[10] F. Zhou, X. Li, and J. Ma, "Parsimonious shooting heuristic for trajectory design of connected automated traffic part I: Theoretical analysis with generalized time geography," *Transportation Research Part B: Methodological*, vol. 95, no. Supplement C, pp. 394-420, 2017, doi: https://doi.org/10.1016/j.trb.2016.05.007.

[11] Y. Xu, D. Xu, S. Lin, T. X. Han, X. Cao, and X. Li, "Detection of Sudden Pedestrian Crossings for Driving Assistance Systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 3, pp. 729-739, 2012, doi: 10.1109/TSMCB.2011.2175726.

[12] M. Zhou, X. Qu, and S. Jin, "On the Impact of Cooperative Autonomous Vehicles in Improving Freeway Merging: A Modified Intelligent Driver Model-Based Approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1422-1428, 2017, doi: 10.1109/TITS.2016.2606492.

[13] J. Ma, X. Li, F. Zhou, J. Hu, and B. B. Park, "Parsimonious shooting heuristic for trajectory design of connected automated traffic part II: Computational issues and optimization," *Transportation Research Part B: Methodological*, vol. 95, pp. 421-441, 2017, doi: http://doi.org/10.1016/j.trb.2016.06.010.

[14] W. J. Schakel, B. van Arem, and B. D. Netten, "Effects of Cooperative Adaptive Cruise Control on traffic flow stability," in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference*, 2010, pp. 759-764, doi: 10.1109/ITSC.2010.5625133.

[15] I. H. Zohdy and H. A. Rakha, "Intersection Management via Vehicle Connectivity: The Intersection Cooperative Adaptive Cruise Control System Concept," *Journal of Intelligent Transportation Systems*, vol. 20, no. 1, pp. 17-32, 2016, doi: 10.1080/15472450.2014.889918.

[16] V. Milanés and S. E. Shladover, "Handling Cut-In Vehicles in Strings of Cooperative Adaptive Cruise Control Vehicles," *Journal of Intelligent Transportation Systems*, vol. 20, no. 2, pp. 178-191, 2016, doi: 10.1080/15472450.2015.1016023.

[17] V. Milanés and S. E. Shladover, "Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data," *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 285-300, 2014, doi: 10.1016/j.trc.2014.09.001.

[18] M. Wang, M. Treiber, W. Daamen, S. P. Hoogendoorn, and B. van Arem, "Modelling supported driving as an optimal control cycle: Framework and model characteristics," *Transportation Research Part C: Emerging Technologies*, vol. 36, no. 0, pp. 547-563, 2013, doi: http://dx.doi.org/10.1016/j.trc.2013.06.012.

[19] M. Treiber and A. Kesting, *Traffic flow dynamics: Data, Models and Simulation*, 1 ed. Springer-Verlag Berlin Heidelberg, 2013, pp. XIV, 506.

[20] A. Kesting, M. Treiber, and D. Helbing, "Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1928, pp. 4585-4605, 2010.

[21] S. Yu and Z. Shi, "The effects of vehicular gap changes with memory on traffic flow in cooperative adaptive cruise control strategy," *Physica A: Statistical Mechanics and its Applications*, vol. 428, pp. 206-223, 2015, doi: 10.1016/j.physa.2015.01.064.

[22] M. Wang, W. Daamen, S. P. Hoogendoorn, and B. van Arem, "Rolling horizon control framework for driver assistance systems. Part II: Cooperative sensing and cooperative control," *Transportation Research Part C: Emerging Technologies*, vol. 40, pp. 290-311, 2014, doi: 10.1016/j.trc.2013.11.024.

[23] V. Milanes, S. E. Shladover, J. Spring, C. Nowakowski, H. Kawazoe, and M. Nakamura, "Cooperative Adaptive Cruise Control in Real Traffic Situations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 296-305, 2014, doi: 10.1109/TITS.2013.2278494.

[24] K. Aghabayk, M. Sarvi, N. Forouzideh, and W. Young, "Modelling heavy vehicle car-following behaviour in congested traffic

conditions," *Journal of Advanced Transportation,* Article vol. 48, no. 8, pp. 1017-1029, 2014, doi: 10.1002/atr.1242.

[25] Y. Zhang and H. Ge, "Freeway travel time prediction using takagi-sugeno-kang fuzzy neural network," *Computer-Aided Civil and Infrastructure Engineering,* vol. 28, no. 8, pp. 594-603, 2013, doi: 10.1111/mice.12014.

[26] T. V. Mathew and K. V. R. Ravishankar, "Neural network based vehicle-following model for mixed traffic conditions," *European Transport - Trasporti Europei*, no. 52, pp. 1-4, 2012. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84856350860&partnerID=40&md5=a5711cb815b9932295109e45df2d4698.

[27] A. Khodayari, A. Ghaffari, R. Kazemi, and R. Braunstingl, "A modified car-following model based on a neural network model of the human driver effects," *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans,* Article vol. 42, no. 6, pp. 1440-1449, 2012, Art no. 6193221, doi: 10.1109/TSMCA.2012.2192262.

[28] L. Chong, M. M. Abbas, and A. Medina, "Simulation of driver behavior with agent-based back-propagation neural network," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2249, pp. 44-51, 2011, doi: 10.3141/2249-07.

[29] S. Panwai and H. Dia, "Neural Agent Car-Following Models," *IEEE Transactions on Intelligent Transportation Systems,* vol. 8, no. 1, pp. 60-70, 2007, doi: 10.1109/TITS.2006.884616.

[30] H. Jia, Z. Juan, and A. Ni, "Develop a car-following model using data collected by "five-wheel system"," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC,* 2003, vol. 1, pp. 346-351, doi: 10.1109/ITSC.2003.1251975. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-33847760622&doi=10.1109%2fITSC.2003.1251975&partnerID=40&md5=7db2faecdc75cd1ec6b4ec80b79434ad

[31] J. Morton, T. A. Wheeler, and M. J. Kochenderfer, "Analysis of Recurrent Neural Networks for Probabilistic Modeling of Driver Behavior," *IEEE Transactions on Intelligent Transportation Systems,* vol. 18, no. 5, pp. 1289-1298, 2017, doi: 10.1109/TITS.2016.2603007.

[32] M. Zhou, X. Qu, and X. Li, "A recurrent neural network based microscopic car following model to predict traffic oscillation," *Transportation Research Part C: Emerging Technologies,* vol. 84, no. Supplement C, pp. 245-264, 2017, doi: https://doi.org/10.1016/j.trc.2017.08.027.

[33] M. Guériau, R. Billot, N.-E. El Faouzi, J. Monteil, F. Armetta, and S. Hassas, "How to assess the benefits of connected vehicles? A simulation framework for the design of cooperative traffic management strategies," *Transportation Research Part C: Emerging Technologies,* vol. 67, pp. 266-279, 2016.

[34] D. Fajardo, T.-C. Au, S. Waller, P. Stone, and D. Yang, "Automated Intersection Control," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2259, pp. 223-232, 2011, doi: 10.3141/2259-21.

[35] S. A. Bagloee, M. Tavana, M. Asadi, and T. Oliver, "Autonomous vehicles: challenges, opportunities, and future implications for transportation policies," *Journal of Modern Transportation,* journal article vol. 24, no. 4, pp. 284-303, 2016, doi: 10.1007/s40534-016-0117-3.

[36] Qualcomm Technologies, "Accelerating C-V2X commercialization," 2017. [Online]. Available: https://www.qualcomm.com/documents/path-5g-cellular-vehicle-everything-c-v2x

[37] (2017). *How Connected Vehicles Work*.

[38] R. E. Stern *et al.*, "Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments," *ArXiv e-prints,* vol. 1705. [Online]. Available: https://arxiv.org/abs/1705.01693v1

[39] S. Cui, B. Seibold, R. Stern, and D. B. Work, "Stabilizing traffic flow via a single autonomous vehicle: Possibilities and limitations," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 11-14 June 2017 2017, pp. 1336-1341, doi: 10.1109/IVS.2017.7995897.

[40] D. Silver *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature,* Article vol. 529, no. 7587, pp. 484-489, 2016, doi: 10.1038/nature16961 http://www.nature.com/nature/journal/v529/n7587/abs/nature16961.html#supplementary-information.

[41] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature,* vol. 518, no. 7540, pp. 529-533, 2015.

[42] D. Silver *et al.*, "Mastering the game of Go without human knowledge," *Nature,* Article vol. 550, no. 7676, pp. 354-359, 2017, doi: 10.1038/nature24270 http://www.nature.com/nature/journal/v550/n7676/abs/nature24270.html#supplementary-information.

[43] M. Zhou and X. Qu, "Microscopic Car-Following Model for Autonomous Vehicles Using Reinforcement Learning," presented at the Symposium on Innovations in Traffic Flow Theory and Characteristics and TFT Midyear Meeting, 2016.

[44] V. Mnih *et al.*, "Playing atari with deep reinforcement learning," *ArXiv e-prints*, vol. 1312. [Online]. Available: https://arxiv.org/abs/1312.5602

[45] C. Desjardins and B. Chaib-Draa, "Cooperative adaptive cruise control: A reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems,* Article vol. 12, no. 4, pp. 1248-1260, 2011, Art no. 5876320, doi: 10.1109/TITS.2011.2157145.

[46] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine learning,* vol. 8, no. 3-4, pp. 229-256, 1992.

[47] R. J. Williams, *Toward a theory of reinforcement-learning connectionist systems*. Northeastern University, 1988.

[48] R. Herman, D. C. Gazis, and R. B. Potts, "Car-Following Theory of Steady-State Traffic Flow," *Operations Research,* vol. 7, no. 4, pp. 499-505, 1959, doi: 10.2307/166948.

[49] R. E. Chandler, R. Herman, and E. W. Montroll, "Traffic dynamics: studies in car following," *Operations research,* vol. 6, no. 2, pp. 165-184, 1958.

[50] E. Kometani and T. Sasaki, *Dynamic behaviour of traffic with a non-linear spacing-speed relationship*. Amsterdam: Elsevier publishing co., 1961.

[51] P. G. Gipps, "A behavioural car-following model for computer simulation," *Transportation Research Part B: Methodological,* vol. 15, no. 2, pp. 105-111, 1981.

[52] M. Bando, K. Hasebe, K. Nakanishi, and A. Nakayama, "Analysis of optimal velocity model with explicit delay," *Physical Review E,* vol. 58, no. 5, p. 5429, 1998.

[53] A. D. Mason and A. W. Woods, "Car-following model of multispecies systems of road traffic," *Physical Review E,* vol. 55, no. 3, pp. 2203-2214, 1997. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.55.2203.

[54] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, "Dynamical model of traffic congestion and numerical simulation," *Physical Review E,* vol. 51, no. 2, pp. 1035-1042, 1995, doi: 10.1103/PhysRevE.51.1035.

[55] K. Hasebe, A. Nakayama, and Y. Sugiyama, "Equivalence of linear response among extended optimal velocity models," *Physical Review E,* vol. 69, no. 1, p. 017103, 2004. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.69.017103.

[56] K. Hasebe, A. Nakayama, and Y. Sugiyama, "Dynamical model of a cooperative driving system for freeway traffic," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics,* vol. 68, no. 2 2, pp. 026102/1-026102/6, 2003, Art no. 026102. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-37649028928&partnerID=40&md5=8946484dc3d9457d59d3a9bfca74b399.

[57] R. Jiang, Q. Wu, and Z. Zhu, "Full velocity difference model for a car-following theory," *Physical Review E,* vol. 64, no. 1, p. 017101, 2001.

[58] Y. Yu, R. Jiang, and X. Qu, "A Modified Full Velocity Difference Model with Acceleration and Deceleration Confinement: Calibrations, Validations, and Scenario Analyses," *IEEE Intelligent Transportation Systems Magazine*, 2019.

[59] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics,* vol. 62, no. 2 B, pp. 1805-1824, 2000. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-0034239164&partnerID=40&md5=628e3d1d53dba5eb70824488e5b2b39d.

[60] M. Treiber, A. Kesting, and D. Helbing, "Influence of Reaction Times and Anticipation on Stability of Vehicular Traffic Flow," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 1999, no. 1, pp. 23-29, 2007.

[61] M. Treiber, A. Kesting, and D. Helbing, "Delays, inaccuracies and anticipation in microscopic traffic models," *Physica A: Statistical Mechanics and its Applications*, vol. 360, no. 1, pp. 71-88, 2006, doi: 10.1016/j.physa.2005.05.001.

[62] P. Kachroo, N. Shlayan, S. Roy, and M. Zhang, "High-Performance Vehicle Streams: Communication and Control Architecture," *IEEE Transactions on Vehicular Technology,* vol. 63, no. 8, pp. 3560-3568, 2014, doi: 10.1109/TVT.2014.2307551.

[63] S. Lefevre, A. Carvalho, and F. Borrelli, "Autonomous car following: A learning-based approach," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, June 28 2015-July 1 2015 2015, pp. 920-926, doi: 10.1109/IVS.2015.7225802.

[64] J. Wei, J. M. Snider, T. Gu, J. M. Dolan, and B. Litkouhi, "A behavioral planning framework for autonomous driving," in *2014 IEEE Intelligent Vehicles Symposium Proceedings*, 2014, pp. 458-464, doi: 10.1109/IVS.2014.6856582.

[65] Z. Wei, J. Xu, and D. Halim, "Braking force control strategy for electric vehicles with load variation and wheel slip considerations," *IET Electrical Systems in Transportation,* Article vol. 7, no. 1, pp. 41-47, 2017, doi: 10.1049/iet-est.2016.0023.

[66] Z. Chen, W. Liu, and Y. Yin, "Deployment of stationary and dynamic charging infrastructure for electric vehicles along traffic corridors," *Transportation Research Part C: Emerging Technologies,* Article vol. 77, pp. 185-206, 2017, doi: 10.1016/j.trc.2017.01.021.

[67] J. Schmidt, M. Eisel, and L. M. Kolbe, "Assessing the potential of different charging strategies for electric vehicle fleets in closed transport systems," *Energy Policy,* vol. 74, pp. 179-189, 11// 2014, doi: https://doi.org/10.1016/j.enpol.2014.08.008.

[68] M. Hanazawa, N. Sakai, and T. Ohira, "SUPRA: Supply underground power to running automobiles: Electric vehicle on electrified roadway exploiting RF displacement current through a pair of spinning tires," in *2012 IEEE International Electric Vehicle Conference, IEVC 2012*, 2012, doi: 10.1109/IEVC.2012.6183164. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84860817732&doi=10.1109%2fIEVC.2012.6183164&partnerID=40&md5=23438ad319355d39e4454366821b3f38

[69] K. Clement-Nyns, E. Haesen, and J. Driesen, "The Impact of Charging Plug-In Hybrid Electric Vehicles on a Residential Distribution Grid," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 371-380, 2010, doi: 10.1109/TPWRS.2009.2036481.

[70] A. Emadi, Y. J. Lee, and K. Rajashekara, "Power Electronics and Motor Drives in Electric, Hybrid Electric, and Plug-In Hybrid Electric Vehicles," *IEEE Transactions on Industrial Electronics,* vol. 55, no. 6, pp. 2237-2245, 2008, doi: 10.1109/TIE.2008.922768.

[71] C.-S. Wang, O. H. Stielau, and G. Covic, "Design considerations for a contactless electric vehicle battery charger," 2005.

[72] B. Powell, K. Bailey, and S. Cikanek, "Dynamic modeling and control of hybrid electric vehicle powertrain systems," *IEEE Control Systems,* vol. 18, no. 5, pp. 17-33, 1998.

[73] C. Chan, "An overview of electric vehicle technology," *Proceedings of the IEEE,* vol. 81, no. 9, pp. 1202-1213, 1993.

[74] B. Varocky, H. Nijmeijer, S. Jansen, I. Besselink, and R. Mansvelder, "Benchmarking of regenerative braking for a fully electric car," *TNO Automotive, Helmond & Technische Universiteit Eindhoven (TU/e),* 2011.

[75] G. Xu, K. Xu, C. Zheng, X. Zhang, and T. Zahid, "Fully Electrified Regenerative Braking Control for Deep Energy Recovery and Maintaining Safety of Electric Vehicles," *IEEE Transactions on Vehicular Technology,* Article vol. 65, no. 3, pp. 1186-1198, 2016, Art no. 7055278, doi: 10.1109/TVT.2015.2410694.

[76] C. Fiori, K. Ahn, and H. A. Rakha, "Power-based electric vehicle energy consumption model: Model development and validation," *Applied Energy,* vol. 168, pp. 257-268, 2016.

[77] L. Lam and P. Bauer, "Practical capacity fading model for Li-ion battery cells in electric vehicles," *IEEE transactions on power electronics,* vol. 28, no. 12, pp. 5910-5918, 2012.

[78] S. S. Choi and H. S. Lim, "Factors that affect cycle-life and possible degradation mechanisms of a Li-ion cell based on LiCoO2," *Journal of Power Sources,* vol. 111, no. 1, pp. 130-136, 2002.

[79] D. Chen, J. Laval, Z. Zheng, and S. Ahn, "A behavioral car-following model that captures traffic oscillations," *Transportation Research Part B: Methodological,* vol. 46, no. 6, pp. 744-761, 2012, doi: 10.1016/j.trb.2012.01.009.

[80] X. Li, J. Cui, S. An, and M. Parsafard, "Stop-and-go traffic analysis: Theoretical properties, environmental impacts and oscillation mitigation," *Transportation Research Part B: Methodological,* vol. 70, pp. 319-339, 2014, doi: http://dx.doi.org/10.1016/j.trb.2014.09.014.

[81] X. Li, X. Wang, and Y. Ouyang, "Prediction and field validation of traffic oscillation propagation under nonlinear car-following laws," *Transportation Research Part B: Methodological,* vol. 46, no. 3, pp. 409-423, 2012, doi: http://dx.doi.org/10.1016/j.trb.2011.11.003.

[82] X. Wang, "Modeling the process of information relay through inter-vehicle communication," *Transportation Research Part B: Methodological,* vol. 41, no. 6, pp. 684-700, 2007, doi: 10.1016/j.trb.2006.11.002.

[83] M. Schönhof, M. Treiber, A. Kesting, and D. Helbing, "Autonomous detection and anticipation of jam fronts from messages propagated by intervehicle communication," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 1999, pp. 3-12, 2007, doi: 10.3141/1999-01.

[84] M. Schönhof, A. Kesting, M. Treiber, and D. Helbing, "Coupled vehicle and information flows: Message transport on a dynamic vehicle network," *Physica A: Statistical Mechanics and its Applications,* vol. 363, no. 1, pp. 73-81, 2006, doi: 10.1016/j.physa.2006.01.057.

[85] X. Yang and W. Recker, "Simulation studies of information propagation in a self-organizing distributed traffic information system," *Transportation Research Part C: Emerging Technologies,* vol. 13, no. 5, pp. 370-390, 2005, doi: 10.1016/j.trc.2005.11.001.

[86] B. van Arem, C. J. G. van Driel, and R. Visser, "The Impact of Cooperative Adaptive Cruise Control on Traffic-Flow Characteristics," *IEEE Transactions on Intelligent Transportation Systems,* vol. 7, no. 4, pp. 429-436, 2006, doi: 10.1109/TITS.2006.884615.

[87] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning,* vol. 8, no. 3-4, pp. 279-292, 1992.

[88] T. P. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," *ArXiv e-prints*, vol. 1509. [Online]. Available: https://arxiv.org/abs/1509.02971

[89] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," presented at the ICML, 2014.

[90] L. Chong, M. M. Abbas, A. Medina Flintsch, and B. Higgs, "A rule-based neural network approach to model driver naturalistic behavior in traffic," *Transportation Research Part C: Emerging Technologies,* vol. 32, pp. 207-223, 2013, doi: 10.1016/j.trc.2012.09.011.

[91] R. S. Sutton, D. A. McAllester, S. P. Singh, and Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation," in *NIPS*, 1999, vol. 99, pp. 1057-1063.

[92] V. R. Konda and J. N. Tsitsiklis, "Actor-Critic Algorithms," in *NIPS*, 1999, vol. 13, pp. 1008-1014.

[93] T. Tieleman, G. Hinton, and K. Swersky. Lecture 6 Neural networks for machine learning.

[94] M. Treiber, A. Kesting, M. Schönhof, and D. Helbing, "Extending adaptive cruise control to adaptive driving strategies," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2000, pp. 16-24, 2007, doi: 10.3141/2000-03.

[95] E. Wilhelm, R. Bornatico, R. Widmer, L. Rodgers, and G. Soh, *Electric Vehicle Parameter Identification*. 2012, pp. 1090-1099.