Energy, Volume 279, September 2023, Article number 128139
DOI: 10.1016/j.energy.2023.128139

# Deep Reinforcement Learning-Based Eco-Driving Control for Connected Electric Vehicles at Signalized Intersections Considering Traffic Uncertainties

Jie Li[1], Abbas Fotouhi[2], Wenjun Pan[1], Yonggang Liu[1, *], Yuanjian Zhang[3] and Zheng Chen[4, **]

[1]State Key Laboratory of Mechanical Transmissions & College of Mechanical and Vehicle Engineering Chongqing University 400044, Chongqing, China
[2]Advanced Vehicle Engineering Centre, School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, Bedfordshire, MK43 0AL, UK
[3]Department of Aeronautical and Automotive Engineering, Loughborough University, Leicestershire, LE11 3TU, UK
[4]Faculty of Transportation Engineering, Kunming University of Science and Technology, Kunming, 650500, China
* Corresponding Authors: Yonggang Liu (andyliuyg@cqu.edu.cn) and Zheng Chen (chen@kust.edu.cn)

**Abstract:** Eco-driving control poses great energy-saving potential at multiple signalized intersection scenarios. However, traffic uncertainties can often lead to errors in ecological velocity planning and result in increased energy consumption. This study proposes an eco-driving approach with a hierarchical framework to be leveraged at signalized intersections that considers the impact of traffic uncertainty. The proposed approach leverages a queue-based traffic model in the upper level to estimate the impact of traffic uncertainty and generate dynamic modified traffic light information. In the lower level, a deep reinforcement learning-based controller is constructed to optimize velocity subject to the constraints from the traffic lights and traffic uncertainty, thereby reducing energy consumption while ensuring driving safety. The effectiveness of the proposed control strategy is demonstrated through numerous simulation case studies. The simulation results show that the proposed method significantly improves energy economy and prevents unnecessary idling in uncertain traffic scenarios, as compared to other approaches that ignore traffic uncertainty. Furthermore, the proposed method is adaptable to different traffic scenarios and showcases energy efficiency.

**Key Words:** Eco-driving, deep reinforcement learning, velocity optimization, signalized intersection, connected electric vehicle.

## I. INTRODUCTION

Due to the persistent increment in energy consumption in the transportation sector, energy-saving solutions

have attracted huge attention from researchers all over the world [1]. Among those methods, eco-driving control has been gradually recognized as a prospective manner to promote energy-saving transportation [2]. With the rapid development of connected and automated vehicles (CAVs), vehicles are expected to have access to various driving environment information and automatedly plan the velocity profiles to reduce energy consumption, referred to as eco-driving technique, which has been widely demonstrated to remarkably reduce energy consumption in different scenarios [3] and is therefore the research focus of this study.

In this study, eco-driving control simply means optimization of the vehicle's velocity profile to avoid unnecessary acceleration or deceleration and maintain the vehicle to operate in an efficient status. According to different driving scenarios, eco-driving can be divided into single-vehicle and car-following scenarios. For the latter field, the basic idea of eco-driving control is to maintain a safe inter-vehicle distance with preceding vehicles and to reduce energy consumption during the car-following scenarios. In [4], car-following scenarios are divided into four modes, and the model predictive control (MPC) is leveraged to optimize the motor torque for electric CAVs in different driving modes. Xie *et al*. [5] proposed a predictive ecological adaptive cruise control approach for plug-in hybrid electric vehicles (PHEVs), in which the motion of the preceding vehicle is predicted and the driving speed is optimized accordingly based on MPC. In [6], reinforcement learning (RL) is exploited to improve the fuel economy of CAVs in the car-following scenarios by considering the nonlinear efficiency of the powertrain system. The eco-driving control in car-following scenarios can achieve energy-saving, safety and driving comfort simultaneously. However, these approaches are designed for specific driving scenarios and are more feasible for congested traffic conditions. For eco-driving control in single-vehicle scenarios, existing methods mainly work based on planning the driving speed by considering complex constraints such as the speed limit, powertrain efficiency and traffic light [7, 8]. For instance, in [9], a novel step-size discrete dynamic programming (DP) algorithm is exploited to optimize driving speed and consequently reduce energy consumption at different speed limits. Li *et al*. [10] constructed a data-driven powertrain energy consumption model to achieve co-optimization of velocity and energy management for PHEVs with high calculation efficiency. In particular, signal phase and timing (SPaT) information of the traffic lights, which can be acquired via vehicle-to-infrastructure (V2I) communication, is quite valuable for eco-driving control in urban roadways. Since velocity profiles can be interrupted by signalized intersections, they result in extra velocity fluctuations and more energy consumption.

Taking SPaT information into consideration, the eco-driving control technique has been widely investigated to promote energy economy and traffic throughput [11]. Early researches on eco-driving control at signalized intersections typically calculate a constant reference speed according to the SPaT information when driving through the intersections [12]. However, the employment of constant driving speed is not practical and cannot achieve optimal performance. In [13], the optimal velocity profile at signalized intersections is simplified into a two- or three-stage driving rule, and the Dijkstra algorithm is exploited to solve the optimal eco-driving control among multiple intersections. In [14], a constant-speed energy consumption map and a bi-level DP method are developed to optimize velocity with high computational efficiency. With the rapid development of machine learning algorithms, RL has received increasing attention in solving optimal control problem (OCP) for energy-saving [15]. To mitigate the high computational burden of traditional optimization-based approaches, a deep reinforcement learning (DRL) approach is proposed to plan an ecological velocity for connected PHEVs in a model-free method [16]. In [17], variable light spacing and trigonometric signal models are employed to train a DRL controller, thereby enhancing the adaptability to multi-light scenarios. However, the aforementioned approaches do not account for the influence of traffic uncertainties. Consequently, these methods cannot guarantee passing through traffic lights without unnecessary idling on real-world road.

To address the influence of traffic uncertainty at signalized intersections, the existing research on eco-driving control can be categorized into two groups to account for traffic uncertainty. The first group focuses on eco-driving control with constraints of traffic lights and preceding vehicles. It involves switching between three driving modes based on real-time driving conditions: traffic lights priority, car-following priority mode and emergency braking mode. For instance, in [18], a predictive cruise control system is designed based on a bi-level MPC algorithm, which employs a car-following-oriented and a SPaT-oriented MPC controller to save energy consumption on urban roads. Similarly, Bai et al. [19] proposed a hybrid DRL-based eco-driving control strategy that combines rule-based and RL-based control policies, wherein the RL-based controller addresses traffic light constraints, and the rule-based policy focuses on velocity planning for car-following scenarios and driving safety. In [20], a DRL-based controller is constructed to execute eco-driving control in urban scenarios based on a twin-delayed deep deterministic policy gradient agent that takes into account the state of the host vehicle, preceding vehicle and next traffic light. However, these approaches may result in unnecessary idling and deceleration before signalized intersections, as they do not

predict the possible waiting queues due to traffic jams before the intersections. Consequently, the controllers are typically forced to switch to the car-following mode or emergency braking to ensure driving safety when the host vehicle is approaching preceding cars, which can lead to a deterioration in the energy-saving performance of the approaches that neglect traffic uncertainty.

In the second group, the waiting queues before signalized intersections are incorporated into eco-driving control, and velocity is planned to avoid unnecessary idling. Sun *et al.* [21] formulated a data-driven chance-constrained robust optimization problem based on empirical sample data in which an effective red-light duration is proposed to illustrate the feasible passing time of a signalized intersection. In [22], an improved queue discharge prediction method is implemented to predict the discharge time of the waiting queue before a signalized intersection, and a hierarchical control framework is conduced to solve the ecological velocity. Dong *et al.* [23] designed a vehicle queue predictor based on intelligent driver model (IDM), followed by a spatial-domain optimal control strategy to realize energy consumption minimization and speed tracker. In addition, traffic flow models based on shockwave theory are gradually employed in prediction of the waiting queues [24]. In [25], a kinematic wave model is established to predict traffic dynamics and vehicle queue length on a signalized road. The optimal velocity is then solved by the direct multiple shooting algorithm with the consideration of the constraints of modified traffic lights. In [26], a long short-term memory neural network is implemented to predict dynamics traffic flow, thereby promoting to waiting queue prediction. However, the above-mentioned methods are typically executed by assuming that a waiting queue is already generated or will emerge before signalized intersections at red phase. In real-world traffic conditions, a waiting queue is typically formed gradually when the host vehicle is approaching the next traffic light, and it is difficult to anticipate the emergence of the waiting queue. Moreover, slow traffic flow might also affect the motion of the host vehicle even at the green phase.

The eco-driving control with incorporating the queue effect have gained substantial attention recently. Motivated by the above discussions, several problems are still unsolved in that area. Firstly, most existing studies are executed via an optimization algorithm. However, the balance between the energy-saving performance and the computational efficiency is intractable to be addressed especially in complicated urban scenarios. Secondly, in most studies, the influence of traffic uncertainty on the eco-driving control at signalized intersections is insufficiently considered. Moreover, most of existing studies are concerned about the traffic uncertainty executed in specific

scenarios, in which the queue progress is formed at the red phase. However, in urban scenarios, the formation of queues is difficult to be anticipated and a slow traffic flow can also affect the motion of host vehicle at the green phase. Hence, only taking the waiting queue at the red phase into consideration cannot sufficiently avoid unnecessary deceleration or idling before the signalized intersections.

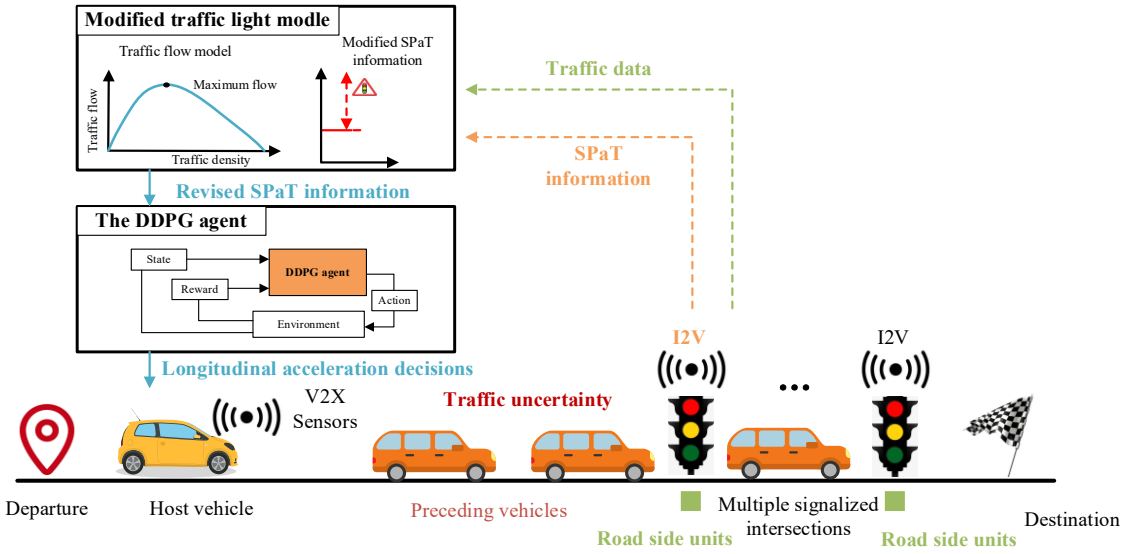



Fig. 1. Illustration of the proposed eco-driving control system.

Motivated by the above analysis, this study proposes a DRL-based eco-driving control approach at multiple signalized intersections, in which a hierarchy framework is implemented as shown in Fig 1. The upper-level controller estimates traffic uncertainty by using a queue-based traffic flow model and modifies the states of the traffic lights accordingly, while the lower-level controller uses DRL to plan an ecological velocity for the vehicle while considering driving safety and comfort constraints. The main contribution of this study is threefold: 1) an eco-driving approach is proposed for signalized intersection scenarios and accounts for the impact of traffic uncertainty. This is an important consideration as traffic flow in urban areas can be highly variable and unpredictable. 2) A dynamic modified traffic light model is created, which considers the queue-based traffic flow model to correct the states of the traffic lights. This model allows for more accurate traffic light predictions and can improve traffic flow. 3) The DRL-based eco-driving controller is designed to plan an ecological velocity at signalized intersections, while considering traffic uncertainties and the constraints of the dynamic modified traffic light. Overall, this study presents a promising approach for eco-driving control in urban areas, which can contribute to a more sustainable and efficient transportation system.

The remainder of this paper is structured as follows: Section II presents the mathematical modeling of vehicle

and traffic uncertainty and formulates the OCP for eco-driving. Section III elaborates the eco-driving control by consideration of the traffic uncertainty. Section IV presents the simulation results to verify the proposed method. Finally, the main conclusions and future work are discussed in Section V.

## II. Modeling and Problem Statement

In this section, the vehicle kinematic model, the traffic flow model, and a modified traffic light model are presented. The modified traffic light model is constructed to estimate the traffic uncertainty at signalized intersections. In addition, this section formulates the OCP of eco-driving.

### *A. Vehicle Model*

In eco-driving control, a longitudinal kinematics model is typically utilized to describe the motion of vehicles [27]. The longitudinal driving force is presented, as:

$$F_r = Mgf\cos(\alpha) + \frac{\rho_{air} C_D A v^2}{2} + Mg\sin(\alpha) \tag{1}$$

where $F_r$ is the resistance driving force, $M$ is the vehicle mass, $f$ means the rolling resistance coefficient, $g$ is the acceleration of gravity, $\alpha$ expresses the road grade, $C_D$ denotes the air drag coefficient, $A$ indicates the frontal area of the vehicle, and $\rho_{air}$ is the air density. The driving resistance force consists of rolling resistance, air resistance, acceleration resistance and road grade resistance. In this research, the lane change is neglected, and the road grade is set as zero. By assuming the state variables of the vehicle model to be velocity ($v$) and position ($d$), the longitudinal kinematics model is formulated, as:

$$\dot{\Delta} v = a \tag{2}$$

$$\dot{\Delta} d = v \tag{3}$$

$$a = \frac{\frac{T_m i_t \eta_t}{r} - F_r}{M\delta} \tag{4}$$

where $a$ is the longitudinal acceleration, $\delta$ is the correction coefficient of rotating mass, $\eta_t$ is the transmission system efficiency, $i_t$ indicates the transmission ratio, and $T_m$ stands for the torque of motor. In this study, the basic parameters of an electric vehicle is applied based on a previous study [28] to calculate the energy consumption

$E(v,a,\alpha)$, as:

$$E(v,a,\alpha) = P_b = \begin{cases} P_m / \eta_m, & P_m > 0 \\ P_m \cdot \eta_m, & P_m < 0 \end{cases} \tag{5}$$

$$P_m = \frac{T_m v}{9550 \cdot r_{tire} i_t} = \frac{T_m \omega_m}{9550} \tag{6}$$

$$\eta_m = \Psi(T_p, \omega_p) \tag{7}$$

where $P_b$ is the battery power, $P_m$ is the motor power, $\alpha$ is the road grade, $\eta_m$ is the motor efficiency, $r_{tire}$ is the radius of vehicle tire, $\omega_m$ is the rotation speed of motor, $\Psi$ represents the motor efficiency map.

*B. Modeling and Prediction of Traffic Uncertainties*

To consider traffic uncertainties at signalized intersections, a queueing-based traffic flow model [29], referred to as Lighthill-Whitham-Richards (LWR) model, is built to predict the queue's discharge time and traffic dynamics. As discussed before, not only the waiting queue has an impact on the motion of the host vehicle, but also a slow traffic flow can confine the speed of the host vehicle. Hence, the influence of traffic uncertainties at signalized intersections are classified into two categories, i.e., waiting queue and slow traffic flow. The LWR model is leveraged to describe traffic dynamics of the waiting queue, as:

$$\frac{\partial \rho(d,t)}{\partial t} + \frac{\partial q(d,t)}{\partial d} = 0 \tag{8}$$

$$v_w = \frac{\Delta q}{\Delta \rho} \tag{9}$$

where $\rho(d,t)$ and $q(d,t)$ represent the traffic density and flow at position $d$ and time $t$, respectively. The parameter $v_w$ is the effective velocity of a shockwave whereas $\Delta\rho$ and $\Delta q$ show the changes of traffic density and flow between different regions. The flow-density diagram of the LWR model in a specific region is shown in Fig. 2. With the increasement of traffic density, a growth of traffic flow is observed until it reaches the maximum traffic flow. By contrast, after reaching the peak point, the traffic flow gradually decreases to zero with the increase of the traffic density. At the maximum traffic density point, we assume a waiting queue emerges in this region. Based on (9), when the traffic light turns to green phase, the dissipating speed of waiting queue can be presented as the effective velocity of shockwave, as:

$$v_{dis} = \frac{q_d - q_u}{\rho_d - \rho_u} \tag{10}$$

where $q_d$ and $q_u$ stand for the traffic flow on the downstream and upstream of the signalized intersection, $\rho_d$ and $\rho_u$ represent the traffic density on the downstream and upstream of the intersection respectively. In most situations, the traffic flow on the downstream of a red-phase intersection can be estimated as the maximum traffic flow, and the upstream of the red-phase intersection reaches the maximum traffic density when a waiting queue is formed. Therefore, the queue's dissipation rate after the green light starts is reformulated, as:

$$v_{dis} = \frac{q_{\max}}{\rho_s - \rho_{\max}} \tag{11}$$

where $\rho_s$ is the traffic density corresponding to the maximum traffic flow. Obviously, in a specific region, the queue's dissipation rate is recognized as a constant, which can be used to predict the dissipation time of the queue.

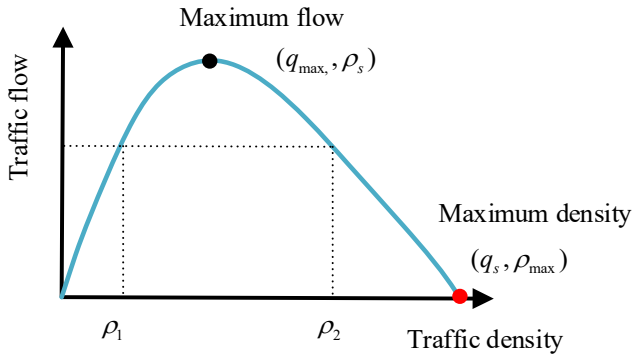



Fig. 2. The flow-density diagram for the LWR model.

To consider the influence of a slow traffic flow, the arrive time to a signalized intersection is predicted for the preceding vehicles. For this reason, the effective speed of the road is calculated based on traffic flow theory [30], as:

$$v_{traff}^{j}(t) = \frac{q^{j}(t)}{\rho^{j}(t)} \tag{12}$$

where $j$ means the $j_{\text{th}}$ region before the $j_{\text{th}}$ traffic light, $v_{traff}^{j}(t)$ indicates current effective speed at the corresponding region, whereas $q^{j}(t)$ and $\rho^{j}(t)$ are the current traffic flow and density at that corresponding region, respectively. If traffic count data is available, the traffic flow and density in a specific region can be estimated, and the effective speed in that region can be determined. In fact, the availability of traffic count data is practical and

can be achieved in the near future [31]. In this study, we assume that loop detectors are installed at each signalized intersection and in the middle of each two intersections. In addition, loop detectors will provide a traffic information vector of each passing vehicle, i.e., $[Index_{veh}, t_{index}, v_{index}]$, including the index, arrive time and velocity of the vehicles. With the support of traffic count data, the detectors at intersections can easily estimate the number of vehicles in that region, and the data from the detectors between two traffic lights can be utilized to estimate the traffic flow and density, as:

$$q^j(t) = \frac{num^j_{veh}}{t_i - t_1} \tag{13}$$

$$\rho^j(t) = \frac{num^j_{veh}(t)}{(t_i - t_1)v^j_{aver}} \tag{14}$$

$$v^j_{aver} = \frac{1}{num^j_{veh}(t)} \sum_1^i v_i \tag{15}$$

where $i = 1, 2, 3, \ldots$ means the $i_{\text{th}}$ vehicle in the region, $num^j_{veh}$ is the number of vehicles in the $j_{\text{th}}$ region, $t_i$ is the arrive time to the loop detectors for the $i_{\text{th}}$ vehicle, and $v^j_{aver}$ is the average speed of the passing vehicles. Finally, based on (12) to (15), the arrival time to the next traffic light of each vehicle can be predicted,

$$t^{pre}_i(t) = \frac{d^j_{triff}}{v^j_{triff}(t)} \tag{16}$$

where $d^j_{triff}$ denotes the length of the $j_{\text{th}}$ region. Thus, the arrival time to next traffic light of each preceding vehicle is determined and continuously updated according to the current traffic flow and traffic density. By this manner, the prediction of the traffic uncertainty includes the anticipation of queue discharge and preceding vehicles.

### *C. Optimal Eco-Driving Control Formulation*

In this study, the aim of eco-driving control at signalized intersections is to optimize the host vehicle's velocity profile for energy-saving by considering the constraints of driving safety and traffic lights. The state variables vector $x$ include the velocity and position of the host vehicle, and the control variable $u$ is the longitudinal acceleration. Therefore, the objective function, the constraints and the system dynamics of the OCP are formulated, as:

$$J = \int_0^T (\lambda_1 p_{light} + \lambda_2 p_{safety} + \lambda_3 E(x,u))dt \tag{17}$$

$$\begin{cases} a_{\min} \le a \le a_{\max} \\ v_{\min} \le v \le v_{\max} \\ v_0 = v_{initial} \\ d_0 = d_{initial} \\ T < T_{\max} \end{cases} \tag{18}$$

$$\begin{bmatrix} \dot{\Delta} d(t) \\ \dot{\Delta} v(t) \end{bmatrix} = \begin{bmatrix} v(t) \\ a(t) \end{bmatrix} \tag{19}$$

where $E(x,u)$ is the instantaneous electricity energy consumption, $\lambda$ are adjustable weight factors to balance different indexes, $p_{light}$ and $p_{safety}$ are the penalties from traffic lights and preceding vehicles, and these penalties are triggered when drive through a red light or a collision happens. $a_{\min}$ and $a_{\max}$ are the limits of acceleration, $v_{\min}$ and $v_{\max}$ are the limits of velocity, $v_0$ and $d_0$ are the initial velocity and position of the host vehicle, respectively, and $T$ is the travel time for the whole scenario. It can be found that the OCP contains nonlinear constraints from traffic lights and preceding vehicles, and consequently, it remains challenging to directly solve it using traditional optimization algorithms. Previous studies attempted to solve this problem by proposing modified optimization frameworks or algorithms, such as iterative DP algorithm and hierarchical optimization framework [23, 26]. However, any improvement of the calculation efficiency is reported to be achieved when the optimality of performance is sacrificed. In this study, a DRL-based method is leveraged to find optimal control decisions of the OCP, thereby promoting performance with a slight calculation burden.

## III. Eco-Driving Control Considering Traffic Uncertainty at Signalized Intersections

To solve the multi-objective OCP of eco-driving control at signalized intersections, RL is applied to plan velocity profiles for the host vehicle. The proposed scheme is described in Fig. 3, wherein a hierarchical framework is implemented, in which the aforementioned traffic model is exploited to predict the traffic uncertainty, and a modified traffic light model is developed to revise the SPaT information in the upper level. In addition, a DDPG agent is designed to generate acceleration control decisions at multiple intersections under the influence of traffic uncertainty.

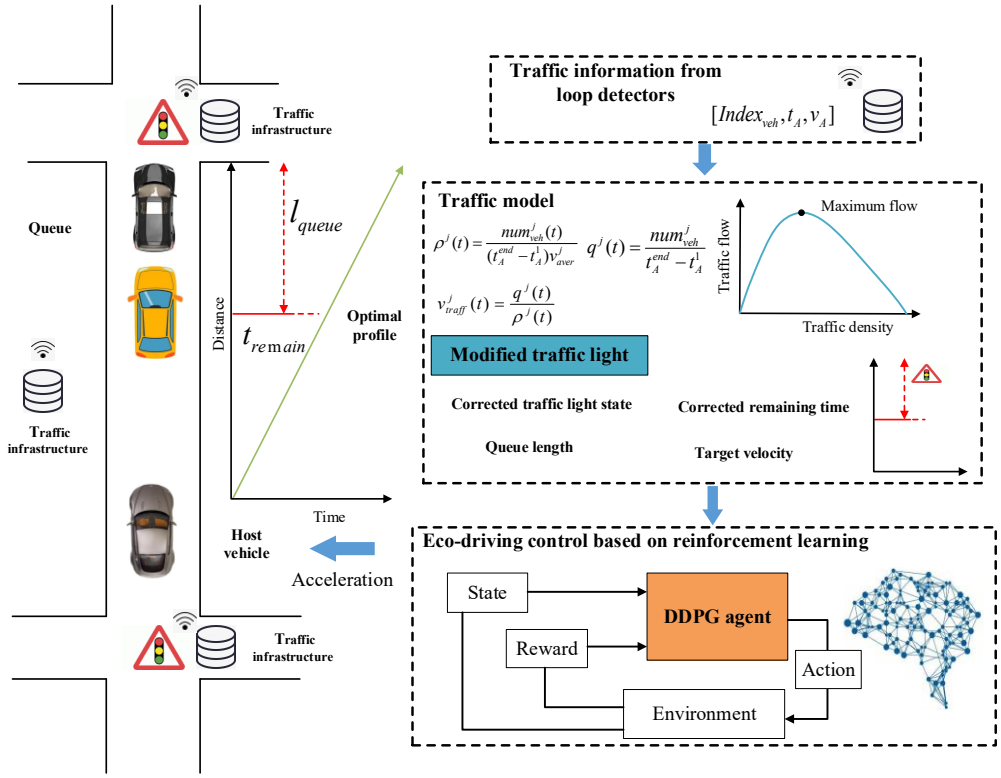



Fig. 3. Schematic of eco-driving control system at signalized intersections considering traffic uncertainties

### *A. Implementation of DRL for Eco-Driving Control*

RL methods consider control systems as a Markov decision process, in which the next state $s_{t+1}$ only depends on the current state $s_t$ and the control action $a_t$. The action is determined via a control policy $\pi(s_t, a_t)$, and a reward $r_t$ is observed from the environment. In those methods, the aim of the agent is to learn a policy for maximizing the cumulative reward, also called Q-value, which can be formulated as follows:

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma \max \sum_{i=t+1}^{T} r(s_i, a_i) \tag{20}$$

where $T$ is the duration of the whole scenario, and $\gamma$ is the discount factor. Eq. (20), referred to as the Bellman function, can be recursively formulated, as:

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma \max Q(s_{t+1}, a_{t+1}) \tag{21}$$

A classic actor-critic architecture is typically employed in RL methods. In that structure, there are two crucial components, i.e., the policy function and the Q-value function, which account for generating actions and estimating the maximum cumulative reward. To concern the requirement of non-linear approximators under complicated state spaces and continuous actions, the DDPG algorithm is utilized to solve the OCP of eco-driving control. The DDPG

agent applies the policy gradient method to train the action and critic module, which can approximate the policy and the Q-value function through deep neural networks, and generate continuous actions [32]. To balance the conflict of energy-saving and pursuing speed, a green light optimized speed advisory (GLOSA) algorithm [33] is leveraged to calculate the reference target speed for the DRL controller. The target velocity is formulated, as:

$$v_{tar} = \max(f_{tar} \cdot v_{tl\_\max}, 1.1 \cdot v_{tl\_\min}) \tag{22}$$

$$v_{tl\_\max} = \begin{cases} v_{\max}, & \text{green phase} \\ \dfrac{d_{tl}}{t_{remain}}, & \text{red phase} \end{cases} \tag{23}$$

$$v_{tl\_\min} = \begin{cases} \dfrac{d_{tl}}{t_{remain}}, & \text{green phase} \\ \dfrac{d_{tl}}{t_{remain} + t_{green}}, & \text{red phase} \end{cases} \tag{24}$$

where $v_{tl\_\max}$ and $v_{tl\_\min}$ represent the maximum and minimum velocities to pass the next traffic light at the green phase, respectively. Note that $v_{tl\_\max}$ and $v_{tl\_\min}$ should satisfy the constraints of speed limits of the road. The parameter $d_{tl}$ is the distance to the next traffic light, $t_{remain}$ is the remaining time of current phase, $t_{green}$ and $t_{red}$ are the time duration of green and red phases, and $f_{tar}$ is a factor to adjust the target velocity.

1) State and action variables

The state vector is designed to illustrate the current state of the eco-driving system. In some previous studies, all possible variables are selected to represent the state of the host vehicle, traffic light and preceding vehicles [34]. However, a quite large state-space model is entailed, leading to difficulty in learning convergence. The state variables are simplified with the help of target velocity, and the variables include host vehicle velocity $v$, host vehicle acceleration $a$, velocity deviation $\Delta v$, velocity deviation integral $\Delta v_{inte}$ and velocity deviation derivative $\Delta v'$. Note that velocity deviation represents the deviation between the host vehicle's velocity and the target velocity. The host vehicle's velocity $v$ and acceleration $a$ can describe the motion of the host vehicle whereas the other variables can provide necessary information of speed pursuing state to pass the next traffic light in a green phase. Furthermore, the acceleration is selected as the action variable to plan the velocity profiles for the host vehicle to achieve energy-saving at signalized intersections.

2）Reward function

The reward function should be designed to encourage the agent to pass traffic lights in a green phase and reduce energy consumption at the same time. Therefore, the reward function at signalized intersection scenarios mainly consists of three aspects. The first one is energy consumption reward. The second one says the travel time should be maintained within an acceptable range. The third aspect is related to driving safety, i.e., obeying traffic rules, including traffic lights and speed limits. It should be noted that an emergency braking policy is applied to avoid collisions, and this item is not included in the reward function. Based on the above discussion, the reward function corresponding to energy consumption is designed as follows:

$$r_e = -f_e \cdot E(v, a, \alpha) \tag{25}$$

where $f_e$ is a weight factor of the energy consumption reward. Furthermore, a penalty is added to the reward function by considering acceleration to avoid frequent speed variations, minimize energy consumption and ensure driving comfort. A quadratic penalty function is considered for acceleration, presented below.

$$r_a = -f_a \cdot a^2 \tag{26}$$

where $f_a$ is a weight factor. The above item generates a penalty to critic module when acceleration or deceleration emerges, especially in short maneuvers with a higher amplitude. In addition, the agent is encouraged to track the target velocity for passing traffic lights at green phase. A penalty velocity item of the reward function is formulated as follows:

$$r_v = -f_v \cdot (v_{tar} - v)^2 \tag{27}$$

where $f_v$ is a weight factor. This penalty item enables the agent to eliminate exceed velocity deviations to the target velocity. However, it does not ensure a preferable tracking of the target velocity. For instance, velocity cannot stably follow the target velocity in a small velocity deviation situation as other reward terms dominant the final reward. Moreover, a positive reward item is necessary to encourage the host vehicle to move forward to the destination. Hence, another two terms to velocity are designed, as:

$$r_{vpr} = f_{vpr} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\Delta v - \mu)^2}{2\sigma^2}\right) \tag{28}$$

$$r_{tl} = \begin{cases} \beta_{tl}, & v_{tl\_\min} \le v \le v_{tl\_\max} \\ -\beta_{tl}, & v_{tl\_\min} \ge v \,||\, v \ge v_{tl\_\max} \end{cases} \tag{29}$$

where $r_{vpr}$ is the positive reward to velocity deviation which satisfies a normal distribution, $\mu$ is the mean value of normal distribution, and $\sigma$ is the standard deviation of the normal distribution. $r_{tl}$ is the positive reward to velocity and encourages the host vehicle to move forward, as long as the velocity is maintained within the boundaries when passing the next traffic light at a green phase. $\beta_{tl}$ indicates the constant value for positive reward, and $f_{vpr}$ is a weight factor. Finally, a reward for safety is designed to penalize behaviors like violating traffic rules when the host vehicle passes traffic lights at a red phase or violate the speed limit. By contrast, the safety reward is zero if there is no unsafe behavior occurs. Hence, a piecewise function for safety reward is presented as follows:

$$r_{TL} = \begin{cases} -\beta_{safe}, & \text{traffic rules violated} \\ 0, & \text{otherwise} \end{cases} \tag{30}$$

where $\beta_{safe}$ is the penalty for safety reward. In this manner, the total reward function can be formulated by adding all items as presented below.

$$r = r_{TL} + r_{tl} + r_a + r_v + r_{vpr} + r_e \tag{31}$$

*B. Eco-Driving Approach Considering Traffic Uncertainty*

As elaborated earlier, in real-world conditions, the SPaT information of the next traffic light may not be exactly the same as practical conditions, due to the waiting queues and slow traffic flow. To address this limitation, a modified traffic light model is developed to revise the affection of traffic uncertainty on SPaT information, combining with the established traffic model. As such, the modified SPaT information is transferred to the DRL controller to realize eco-driving control considering the influence of traffic uncertainty as shown in Fig. 4.

The crucial information of traffic light consists of its state $st_{tl}$, position $p_{tl}$ and remaining time $t_{re}$. Based on this information, the target velocity can be obtained via (22) to (24). In this study, the lane change action is ignored, and the host vehicle and preceding vehicles are assumed to drive in the same lane. Thus, the existence of any preceding vehicle prevents the host vehicle from passing the signalized intersection, and the modified traffic light state is corrected to red phase to accommodate it.

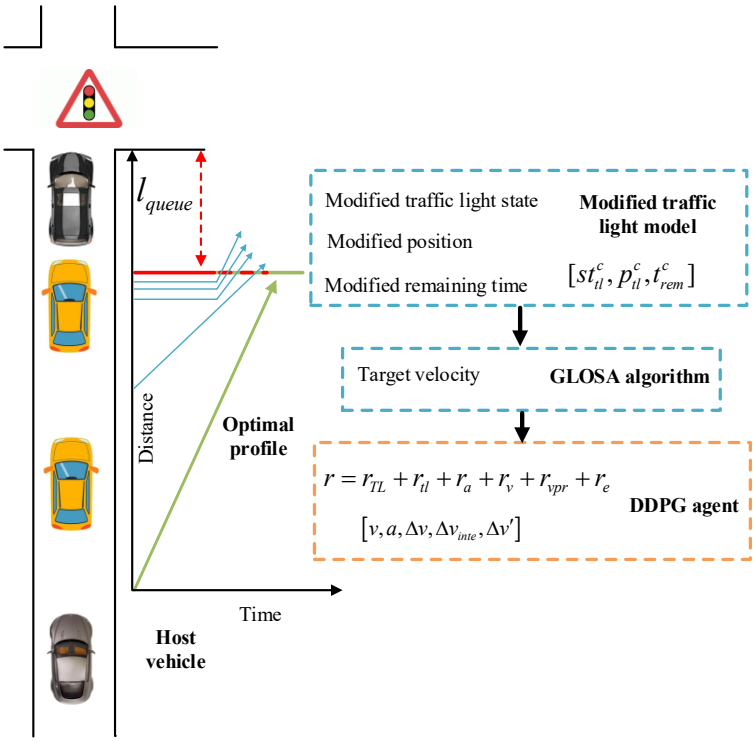



Fig. 4. Schematic of eco-driving control using a modified traffic light model and DRL

The modified state of the next traffic light is corrected as $st_{tl}^c$ according to the following rule:

$$st_{tl}^c = \begin{cases} st_{tl}, & num_{veh}^j = 0 \\ red, & num_{veh}^j \neq 0 \end{cases} \tag{32}$$

The modified state of traffic light switches to red when there are preceding vehicles in the same region, otherwise it keeps the actual state of traffic light. The motion state of the last preceding vehicle is leveraged to modify the position and remaining time of the next traffic light when $st_{tl}^c$ is transferred to red phase. The modified rules can be divided into three cases to differentiate the influence of formed waiting queues, forming waiting queues and slow traffic flow.

1) Case I: the queue process is completed, and the last preceding vehicle stops before the next traffic light. The switch conditions are shown below:

$$\begin{cases} t_{end}^{pre}(t) \in \left[ t_r^b, t_r^e \right] \\ t_{end}^{pre}(t) < t_{cur} \ \& \ v_{end} == 0 \end{cases} \tag{33}$$

where $t_{end}^{pre}(t)$ is the predicted arrival time to next traffic light for the last preceding vehicle, $t_r^b$ and $t_r^e$ are the starting and ending time of the red phase, $t_{cur}$ means the current time, and $v_{end}$ is the velocity of the last preceding

vehicle. The modified position $p_{tl}^{c}$ and the modified remaining time $t_{remain}^{c}$ of traffic light are formulated as follows:

$$p_{tl}^{c} = p_{tl} - L_{queue} \tag{34}$$

$$L_{queue} = (L_0 + L_{veh}) num_{queue} \tag{35}$$

$$t_{remain}^{c}(t) = t_{remain} + dt_{remain}(t) + \delta_t \tag{36}$$

$$dt_{remain}(t) = \frac{L_{queue}}{C_{dr}} - t_{remain} \tag{37}$$

where $L_{queue}$ is the queue length, $num_{queue}$ is the number of vehicles in the queue, $L_0$ is the standstill spacing, $L_{veh}$ is the length of each vehicle, $dt_{remain}$ is the correction value of remaining time, and $\delta_t$ is the timing buffer to ensure the safety. In this situation, a waiting queue is already formed, and a constant postponement to the actual remaining time is generated according to the predicted queue's dissipation time. This case typically occurs when the host vehicle is approaching next traffic light.

2) Case II: the waiting queue is forming, and the last preceding vehicle is approaching the next traffic light. The switch conditions are shown below:

$$\begin{cases} t_{end}^{pre}(t) \in \left[ t_r^b, t_r^e \right] \\ t_{end}^{pre}(t) > t_{cur} \ \& \ v_{end} > 0 \end{cases} \tag{38}$$

Similar to Case I, the modified position $p_{tl}^{c}$ is predicted base on (34) and (35), whereas the correction value of remaining time $dt_{remain}$ is formulated, as:

$$dt_{remain}(t) = \frac{L_{queue}(t)}{C_{dr}} + t_{remain} \tag{39}$$

In this situation, the waiting queue undergoes forming or is not formed yet. However, the traffic model predicts there will be a waiting queue ahead, and the remaining time of the current phase is dynamically updated. Therefore, the modified remaining time is predicted combining with the motion of the preceding vehicles, actual remaining time and predicted queue length. In this case, there is normally a long distance between the host vehicle and the next traffic light.

3) Case III: the last preceding vehicle can pass the next traffic light at green phase and the host vehicle will not be affected by any waiting queue. The switch conditions are shown below:

$$t_{end}^{pre}(t) \in \left[ t_g^b, t_g^e \right] \tag{40}$$

where $t_g^b$ and $t_g^e$ are the starting and ending time of the green phase. The last preceding vehicle can pass the next traffic light without idling, which indicates that the motion of the host vehicle will not be affected by any waiting queue. Hence, the modified position of the traffic light is consistent with its actual position, and the correction value of the remaining time $dt_{remain}$ is formulated, as:

$$dt_{remain}(t) = t_{pre}^{end}(t) - t_{curr} - t_{remain} \tag{41}$$

In this situation, the velocity profile of the host vehicle is affected by the preceding traffic flow instead of a waiting queue. Therefore, a time-varying postponement of the remaining time is designed to avoid unnecessary idling before a traffic light. On this basis, a modified traffic light model is developed, and the modified position $p_{tl}^c$ and modified SPaT information of the traffic light $\left[ st_{tl}^c, t_{remain}^c \right]$ are leveraged to calculate the target velocity using the GLOSA algorithm. As a result, the host vehicle can anticipate potential waiting queues or traffic flow before the next traffic light and can reasonably optimize its velocity profile to reduce energy consumption and eliminate unnecessary idling as much as possible.

## IV. Simulation Validations

To verify performance of the proposed method, several simulation case studies are conducted under a MATLAB/Simulink platform. A virtual traffic environment model [35], including traffic lights, preceding vehicles and road grade, etc., is employed to test eco-driving control strategies. In this study, the motion of preceding vehicles is controlled by IDM according to real-time driving conditions in which all preceding vehicles are assumed to have a same desired driving speed of 18 m/s. Three benchmark approaches are constructed to validate the proposed method (called MG-DRL). The first benchmark method (i.e., IDM) is an IDM-based strategy, which is selected to simulate the behavior of human drivers in the virtual traffic environment. The mathematical description of IDM is formulated, as:

$$a(t) = a_{\max}[1 - (\frac{v(t)}{v_{desired}})^{\sigma} - (\frac{s^*(v(t), \Delta v(t))}{s(t)})^2] \tag{42}$$

$$s^* = s_0 + T \cdot v(t) + \frac{v(t) \cdot \Delta v(t)}{2\sqrt{a_{\max} b}} \tag{43}$$

where $v_{desired}$ is the desired speed, $\sigma$ is the acceleration exponent, $s^*$ is the desired headway distance, $T$ is the constant time headway, $b$ is the desired deceleration, and $s_0$ is the minimum safety distance. The basic parameters of the IDM model in this study are presented in Table 1.

Table 1. The basic parameters of the IDM model

| Characteristic | Value |
|---|---|
| Desired speed (m/s) | 18 |
| Acceleration exponent | 4 |
| Constant time headway (s) | 3 |
| Desired deceleration (m) | 1.6 |
| Minimum safety distance (m) | 3 |

The second benchmark method (i.e. G-IDM) is a GLOSA-based strategy, which plans the velocity profile according to real SPaT information without considering traffic uncertainties [20]. $f_{tar}$ is set to be 0.6, which is the same as what was set in the proposed method, whereas an IDM-based model is conducted to follow the velocity planning result obtained from the GLOSA algorithm. Similarly, the third method (i.e., G-DRL) plans the reference velocity based on the GLOSA algorithm, and it uses the DRL-based controller to optimize driving speed by considering energy efficiency. The basic parameters of the host vehicle are presented in Table 2.

Table 2. The basic parameters of the host vehicle

| Characteristic | Value |
|---|---|
| Mass (kg) | 1640 |
| Frontal area ($m^2$) | 2.27 |
| Air drag coefficient | 0.3146 |
| Tire rolling radius (m) | 0.316 |
| Rolling resistance coefficient | 0.008 |
| Battery capacity (Ah) | 32.5 |
| Final gear ratio | 7.94 |

### *A. DRL Parameter Setting and Training*

The main parameters of the DRL algorithm are presented in Table 3. Two deep networks are designed to construct the actor network and the critic network, which consist of five and six hidden layers, respectively. These two networks are trained based on an actor-critic structure, and the reward is calculated according to (25) to (31) at each episode. A route with six traffic lights is selected to train the DDPG agent, and the maximum training episode is set to 500, whereas the sample time of the simulation is set to 0.1 s.

Table 3. Parameters setting of the DRL algorithm

| Parameters | Value |
|---|---|
| Number of neurons in critic network | 60 |
| Number of neurons in action network | 60 |
| Learning rate of critic network | 1.00e-3 |
| Learning rate of action network | 1.00e-4 |
| Number of layers in critic networks | 6 |
| Number of layers in action networks | 5 |
| Number of neurons in hidden layers | 120 |
| Target smooth factor | 1.00e-3 |
| Minibatch size | 268 |
| Discount factor | 0.99 |
| The weight of energy consumption $f_e$ | 25 |
| The weight of acceleration penalty $f_a$ | 0.2 |
| The weight of velocity deviation penalty $f_v$ | 0.025 |
| The positive reward for velocity $\beta_{tl}$ | 0.8 |
| The penalty for safety reward $\beta_{safe}$ | 1000 |

To train the agent under different driving scenarios, several system parameters are randomized at each episode, including the initial state of the host vehicle and preceding vehicles as well as the SPaT information. The training result is shown in Fig. 5, the average accumulative reward increases sharply at the beginning of training, and it gradually stabilizes after around 200 training episodes. Due to the randomness of initial conditions, the accumulative reward of each episode has a slight difference after convergence. Thus, a trained DRL-based controller is obtained for online execution, in which the controller can directly generate acceleration decisions according to the state of the system.

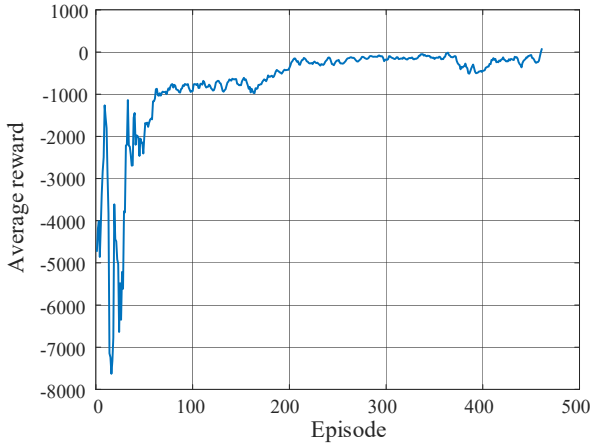



Fig. 5. Convergence of training.

*B. Analysis of Energy Efficiency*

To verify the energy-saving performance of the proposed eco-driving approach, a driving scenario with six signalized intersections and six preceding vehicles is constructed to simulate different eco-driving approaches for

comparison. Besides, the positions of traffic lights are randomized to distinguish them from the training scenarios. The simulation results of distance profiles are shown in Fig. 6 (a). Obviously, the velocity profile of the MG-DRL method can be preferably optimized in advance without unnecessary idling or deceleration, and the host vehicle can exactly pass each signalized intersection at available green phase. By contrast, the IDM-based method can only follow the desired speed, which has to stop three times before traffic lights in this scenario. As shown in Fig. 6 (b), compared with IDM-based method, the G-DRL and G-IDM methods can plan similar low speed to pass the fifth intersection while the host vehicles of these two approaches are affected by the waiting queue before the traffic light. By contrast, the proposed MG-DRL method successfully predicts the emergence of the waiting queue, and a lower driving speed is planned to avoid idling. Fig. 6 (c) illustrates the modified SPaT information to consider the slow traffic flow. In summary, the proposed modified traffic light model can accurately modify the SPaT information. A modified red phase is observed and dynamically updated when there are preceding vehicles ahead, thus impacting the velocity planning in a complicated urban road.

The generated velocity profiles are shown in Fig. 7. The host vehicle under the proposed MG-DRL method maintains a reasonable driving speed throughout the whole trip. A relative low speed is planned when a potential waiting queue is predicted ahead to pass the next traffic light smoothly (such as the one happens around t=50 s and t=300 s in the figure). In other situations, the host vehicle drives at an ecological speed to pass the traffic lights and reduce energy consumption. According to the results, velocity profiles under the G-DRL and G-IDM methods have similar trends in most of the cases. However, the G-DRL method shows fluctuations in velocity when the host vehicle is approaching the preceding vehicles. The reason is that the G-DRL method does not integrate the influence of preceding vehicles into the DDPG agent, and therefore, the simple switch logic in this method leads to frequent switches between the DRL-oriented and emergency braking modes. Fig. 8 presents the results of acceleration distribution and idling time proportions. As shown in Fig. 8 (a), compared with the benchmark methods, the host vehicle under the MG-DRL method notably reduces acceleration time by applying an ecological velocity planning. Besides, all acceleration decisions are maintained within a small range to reduce energy consumption and guarantee the driving comfort. Moreover, the idling proportion results, shown in Fig. 8 (b), reveal that unnecessary idling is eliminated under the MG-DRL method, which significantly contributes to reducing acceleration duration.

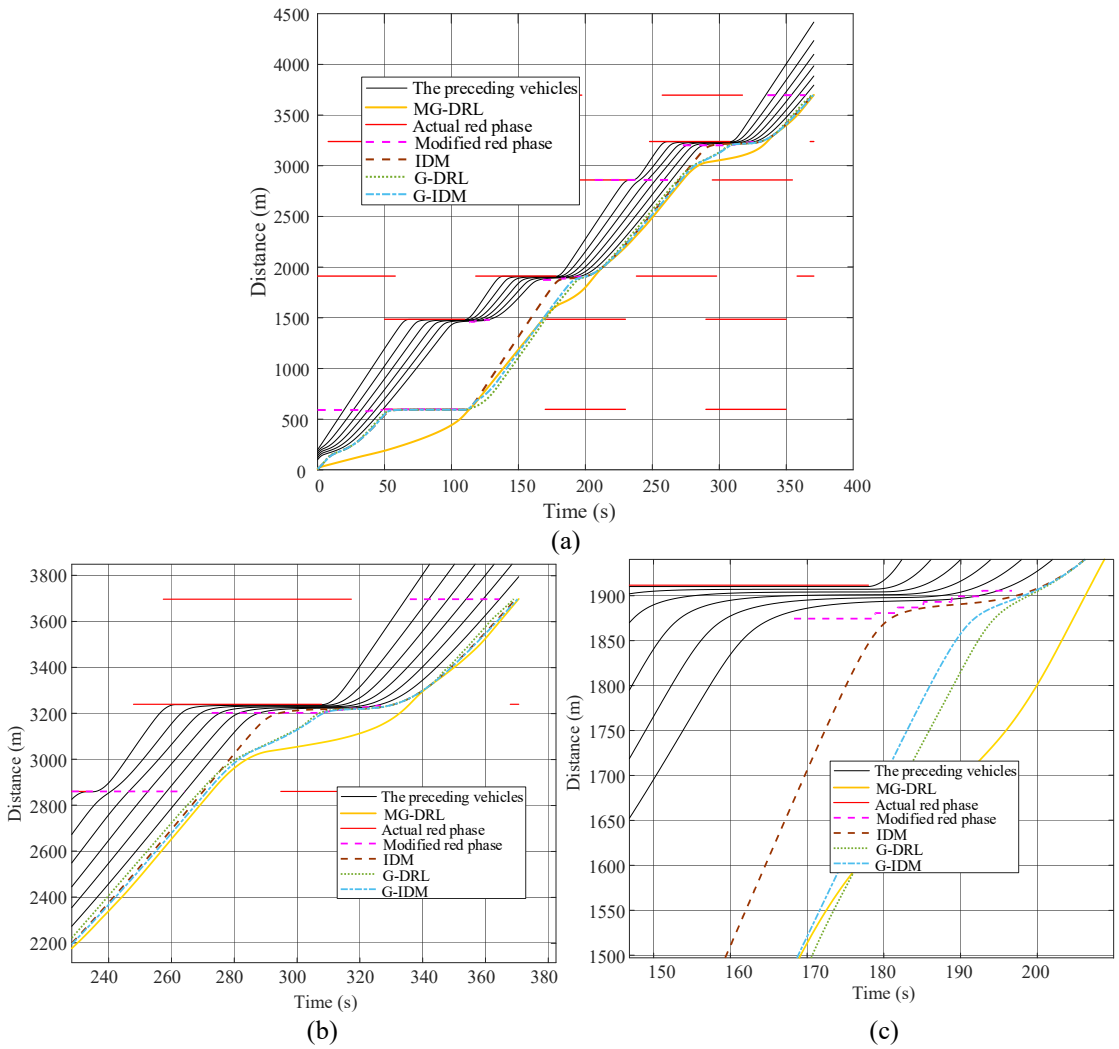



Fig. 6. Position profiles of different methods when there are six preceding vehicles. (a) the position profiles of the whole scenario, (b) the position profiles of the fifth intersection, (c) the position profiles of the third intersection.

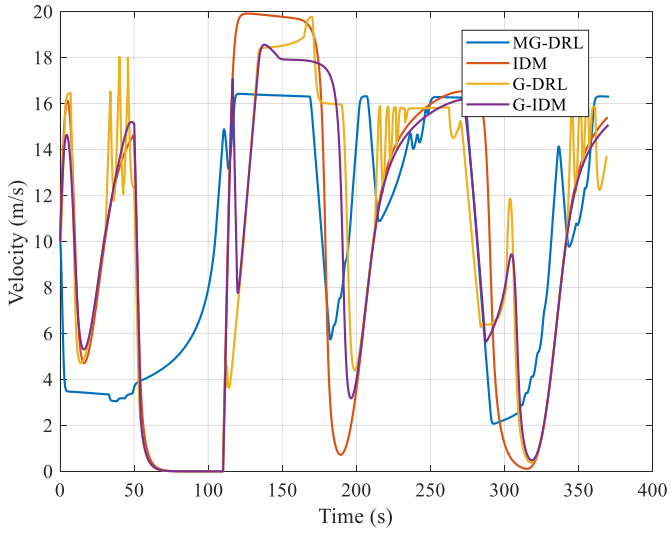
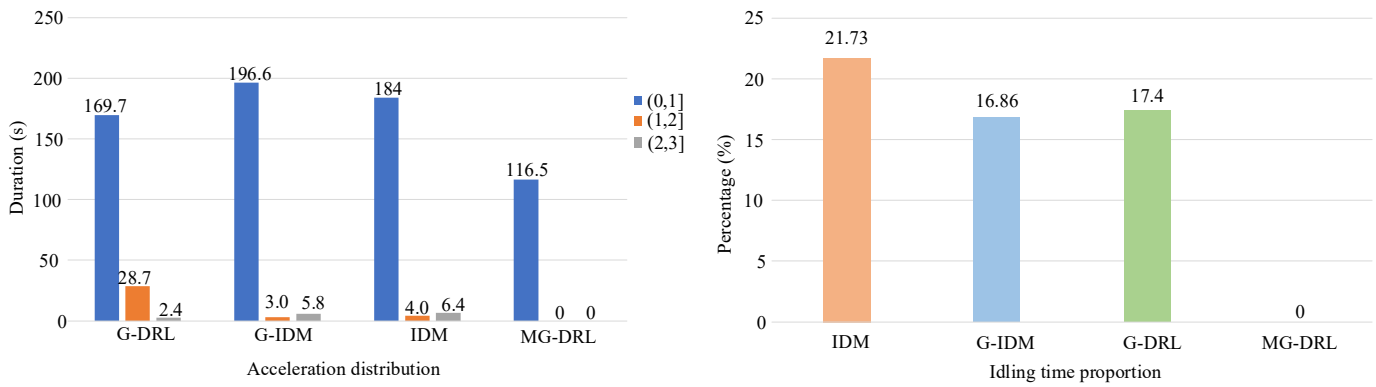



Fig. 7. Velocity profiles of different approaches.




(a) (b)

Fig. 8. Results of acceleration distribution and idling proportion. (a) acceleration distribution, (b) idling proportion

The simulation results of different approaches are summarized in Table 4. The proposed method generates a notable energy-saving improvement of 24.29% compared with the IDM method, which is regarded as human drivers in this study. Furthermore, superior performance in energy consumption is raised by the proposed method, compared with the other existing eco-driving approaches which do not integrate traffic uncertainties into their controllers. In addition, all schemes have slight difference in the whole travel time, indicating that the proposed method improves the energy efficiency without remarkably sacrificing traffic efficiency. It should be noted that, compared with the G-IDM method, the G-DRL method consumes more energy due to its frequent mode switching. To sum up, the MG-DRL method is demonstrated to effectively handle the traffic uncertainties in multiple signalized intersections, and the energy-saving performance is significantly improved compared with the benchmark approaches.

Table 4. Simulation results under different approaches

| Method | Battery energy consumption (kWh) | Improvement (%) | Travel time (s) | Idling time proportion (%) |
|---|---|---|---|---|
| IDM | 0.7 | - | 369.6 | 21.73 |
| G-IDM | 0.6 | 14.29 | 370.1 | 16.86 |
| G-DRL | 0.63 | 10 | 368.9 | 17.40 |
| MG-DRL | 0.53 | 24.29 | 370.5 | 0 |

### *C. Analysis of Adaptability*

To verify the adaptability performance of the proposed method in different traffic scenarios, a Monte Carlo simulation is conducted to fully explore the performance of the proposed method. The simulation involves 500 various trials for each method in the same road. While, the traffic parameters, including initial SPaT information of traffic light, the number of preceding vehicles and initial state of preceding vehicles, are all randomized in each interval. To fully explore the adaptability of the proposed method, the energy consumption, travel time, idling time proportion and low-speed time proportion are summarized for comparison. The speed boundary (1.39 m/s) is

selected to judge the vehicle in a low-speed state [26].

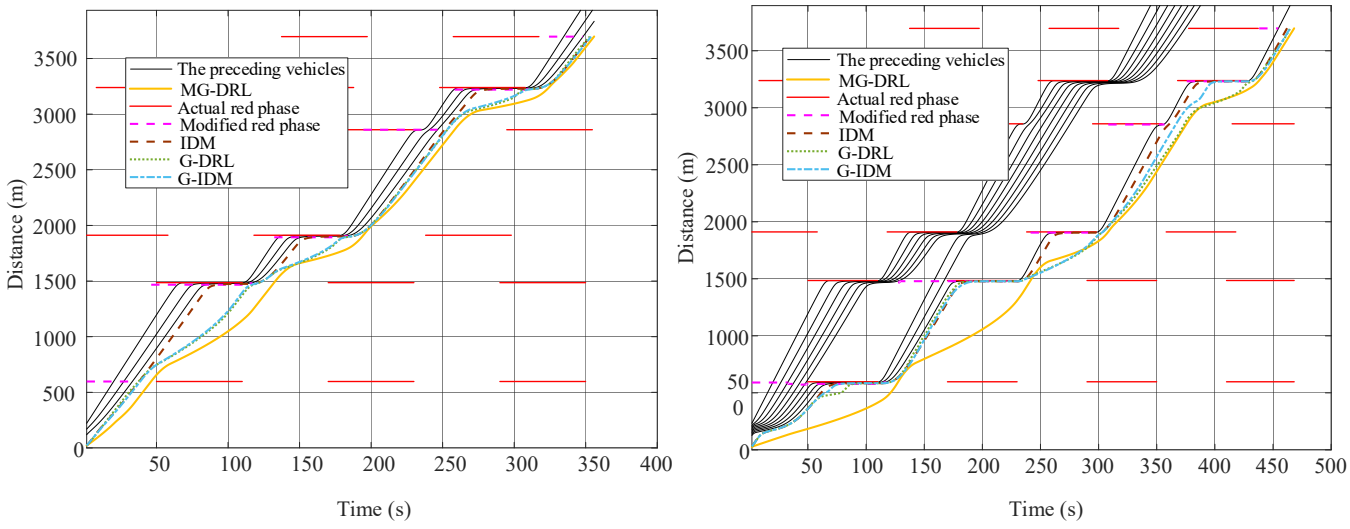
Table 5 summarizes the average energy consumption and travel time. Compared with the IDM method, an average improvement of 17.65% is observed, indicating that the proposed method can improve energy efficiency in different traffic conditions. In addition, an average energy consumption reduction of 5.77% is achieved, compared with other two eco-driving approaches. Furthermore, the travel time of the proposed method doses not have significant increase, and the idling time and low-speed time proportion are remarkably decreased. Note that the extreme congestion scenarios where idling is inevitable for the host vehicle, are not discussed in this study. The reason is that eco-driving approaches considering the SPaT information, have limited contribution to energy-saving in these extreme scenarios [25].

Table 5. Average simulation results under different traffic scenarios

| Method | Average battery energy consumption (kWh) | average energy consumption improvement (%) | Average travel time (s) | Time variation (%) | Average idling time proportion (%) | Average low-speed time proportion (%) |
|---|---|---|---|---|---|---|
| IDM | 0.68 | - | 381 | - | 13.35 | 13.02 |
| G-IDM | 0.6 | 11.76 | 370 | -2.4 | 8.72 | 7.84 |
| G-DRL | 0.59 | 13.24 | 369 | -2.8 | 5.25 | 5.69 |
| MG-DRL | 0.56 | 17.65 | 382 | 0.82 | 1.7 | 2.63 |

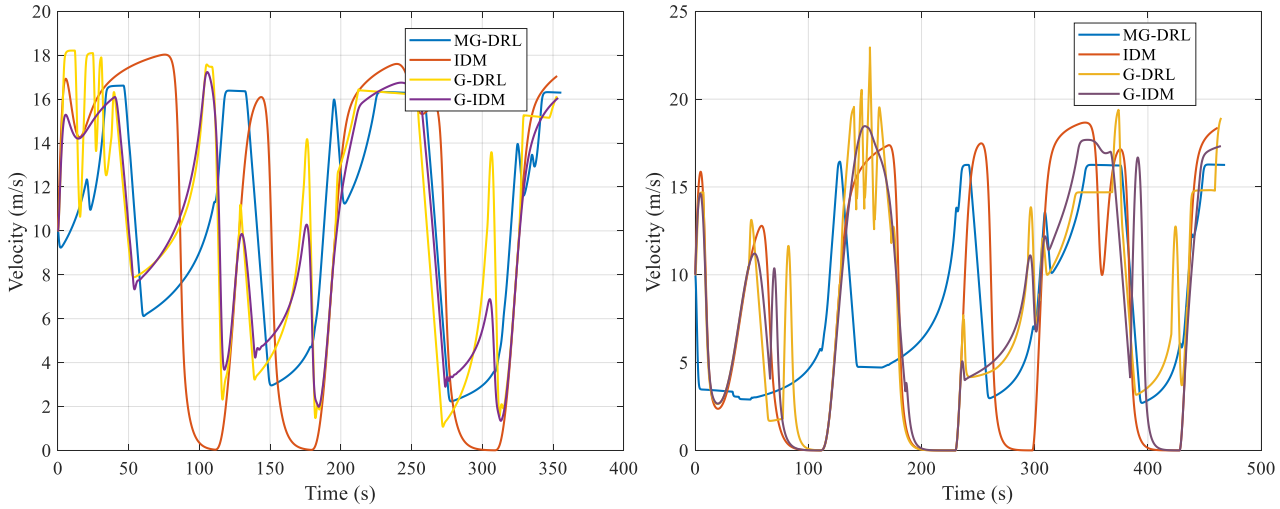
Specially, two tests with different number of preceding vehicles are selected to further analyze the influence of different traffic conditions, which are defined as Scenario A and B. The number of preceding vehicles in these two scenarios is set as 3 and 9, respectively. These two scenarios are regarded as lighter and heavier traffic congestion levels, and other traffic parameters are set the same as the simulation in Section IV-B. Fig. 9 illustrates the position profiles of these two scenarios. The results demonstrate that the proposed MG-DRL method still has superior adaptability in different traffic scenarios, and the host vehicle under the MG-DRL method can accurately pass each intersection without idling. As shown in Fig. 9 (a), all three eco-driving approaches have similar trends in position profiles in a light traffic scenario. The reason is that the waiting queue does not have significant impact on the motion of the host vehicle in this traffic scenario. Besides, the slow traffic flow is concerned in the first and second intersections, wherein the host vehicle plans an ecological speed to eliminate deceleration and extra energy consumption when approaching the preceding vehicles. By contrast, as shown in Fig. 9 (b), the MG-DRL method plans a preferable velocity profile to avoid unnecessary idling when a potential traffic intervene is predicted, especially in the first two intersections. The velocity profiles are shown in Fig. 10 (a) and (b), in different scenarios,

and it can be observed that the MG-DRL method can effectively plan driving speed within a reasonable range. Additionally, the proposed method significantly restrains speed waves, unnecessary decelerations and idling, compared with other approaches. The results verify that the host vehicle can satisfactorily tackle the traffic uncertainties in different traffic scenarios.




(a) (b)

Fig. 9. Position profiles in different traffic scenarios. (a) Scenario A, (b) Scenario B.




(a) (b)

Fig. 10. Velocity profiles in different scenarios. (a) Scenario A, (b) Scenario B.

The results of these two scenarios are summarized in Table 6 for performance comparison. It can be observed that the MG-DRL method can effectively promote the energy efficiency in different scenarios with a maximum energy economy improvement of 30.64% compared with the IDM method. Moreover, an upward tendency in the

energy efficiency improvement is observed with the increase of idling time proportion. The reason is that the waiting queues and slow traffic flow tend to emerge at signalized intersections, and the traffic uncertainty imposes more impact on the velocity planning in a high-level congestion scenario. By contrast, all three eco-driving approaches have similar contributions to the energy consumption in Scenario A which represents a light traffic condition. Compared with other eco-driving schemes, the proposed MG-DRL method provides a modest average energy efficiency improvement of 3.74% in free-flowing traffic conditions. To conclude, the proposed method is validated to show strong adaptability in different traffic scenarios, and the energy efficiency is effectively promoted in all scenarios, especially in high-level congestion scenarios.

Table 6. Simulation results under different traffic scenarios

| | Scenario A | | | | Scenario B | | | |
|---|---|---|---|---|---|---|---|---|
| Method | Energy consumption (kWh) | Improvement (%) | Travel time (s) | Idling time proportion (%) | Energy consumption (kWh) | Improvement (%) | Travel time (s) | Idling time proportion (%) |
| IDM | 0.67 | - | 352.9 | 24.6 | 0.62 | - | 464.6 | 19.84 |
| G-IDM | 0.54 | 19.4 | 353.6 | 0.93 | 0.56 | 9.68 | 464.5 | 24.2 |
| G-DRL | 0.52 | 22.39 | 352.7 | 1.70 | 0.54 | 12.9 | 464.7 | 16.4 |
| MG-DRL | 0.51 | 23.88 | 355.8 | 0 | 0.43 | 30.64 | 468.1 | 0 |

## V. CONCLUSIONS

This article proposes a DRL-based eco-driving approach to tackle the traffic uncertainty at multiple signalized intersections with limited access to traffic data. The proposed method can ecologically plan velocity profiles at signalized intersections, regarding the potential impacts of slow traffic flow and waiting queues, to avoid unnecessary idling and decelerations. A hierarchy framework is developed, in which the upper layer predicts the influence of traffic uncertainty based on a dynamics traffic model, and a modified traffic light model is constructed to revise the actual SPaT information. In the lower level, a DRL-based controller is designed to plan ecological driving speed according to the system's states, including the motion of the host vehicle, revised SPaT information, etc. Substantial simulations are conducted to validate the effectiveness of the proposed method, and the results demonstrate that it significantly promotes energy efficiency when compared with both the human driver model and other existing eco-driving approaches, with an average improvement of 17.65% and 5.77%, respectively. In addition, the proposed method is demonstrated to have preferable energy economy in different traffic scenarios, and a growth tendency of energy efficiency improvement is observed with the increase of traffic congestion level.

Our future work will be focused on investigating eco-driving approach in complicated environments by considering cooperation between multiple connected vehicles. The influence of driving style will be integrated into the eco-driving approach.

## ACKNOWLEDGMENTS

The work is funded by the National Natural Science Foundation of China (No. 52172400 and 52272395) in part, and Science and Technology Research Program of Chongqing Municipal Education Commission (No. KJQN201901539) in part. Any opinions expressed in this paper are solely those of the authors and do not represent those of the sponsors.

## REFERENCES

[1] M. Yan, G. Li, M. Li, H. He, H. Xu, and H. Liu, "Hierarchical predictive energy management of fuel cell buses with launch control integrating traffic information," *Energy Conversion and Management,* vol. 256, p. 115397, 2022.

[2] A. Vahidi and A. Sciarretta, "Energy saving potentials of connected and automated vehicles," *Transportation Research Part C: Emerging Technologies,* vol. 95, pp. 822-843, 2018.

[3] F. Zhang, X. Hu, R. Langari, and D. Cao, "Energy management strategies of connected HEVs and PHEVs: Recent progress and outlook," *Progress in Energy and Combustion Science,* vol. 73, pp. 235-256, 2019.

[4] L. Xie, Y. Luo, D. Zhang, R. Chen, and K. Li, "Intelligent energy-saving control strategy for electric vehicle based on preceding vehicle movement," *Mechanical Systems and Signal Processing,* vol. 130, pp. 484-501, 2019.

[5] S. Xie, X. Hu, T. Liu, S. Qi, K. Lang, and H. Li, "Predictive vehicle-following power management for plug-in hybrid electric vehicles," *Energy,* vol. 166, pp. 701-714, 2019.

[6] G. Li and D. Görges, "Ecological Adaptive Cruise Control for Vehicles With Step-Gear Transmission Based on Reinforcement Learning," *IEEE Transactions on Intelligent Transportation Systems,* 2019.

[7] S. Xu, S. E. Li, H. Peng, B. Cheng, X. Zhang, and Z. Pan, "Fuel-saving cruising strategies for parallel HEVs," *IEEE Transactions on Vehicular Technology,* Article vol. 65, no. 6, pp. 4676-4686, 2016, Art no. 7296694.

[8] Y. Liu, Z. Huang, J. Li, M. Ye, Y. Zhang, and Z. Chen, "Cooperative optimization of velocity planning and energy management for connected plug-in hybrid electric vehicles," *Applied Mathematical Modelling,* vol. 95, pp. 715-733, 2021.

[9] Z. Ye, K. Li, M. Stapelbroek, R. Savelsberg, M. Gunther, and S. Pischinger, "Variable Step-Size Discrete Dynamic Programming for Vehicle Speed Trajectory Optimization," *IEEE Transactions on Intelligent Transportation Systems,* Article vol. 20, no. 2, pp. 476-484, 2019, Art no. 8320319.

[10] J. Li, Y. Liu, Y. Zhang, Z. Lei, Z. Chen, and G. Li, "Data-driven based eco-driving control for plug-in hybrid electric vehicles," *Journal of Power Sources,* vol. 498, p. 229916, 2021.

[11] Z. Nie, Y. Jia, W. Wang, and R. Outbib, "Eco-Co-Optimization strategy for connected and automated fuel cell hybrid vehicles in dynamic urban traffic settings," *Energy Conversion and Management,* vol. 263, p. 115690, 2022.

[12] G. Thomas and P. G. Voulgaris, "Fuel minimization of a moving vehicle in suburban traffic," in *2013 American Control Conference*, 2013: IEEE, pp. 4009-4014.

[13] Q. Lin, S. E. Li, S. Xu, X. Du, D. Yang, and K. Li, "Eco-driving operation of connected vehicle with V2I communication among multiple signalized intersections," *IEEE Intelligent Transportation Systems Magazine,* vol. 13, no. 1, pp. 107-119, 2020.

[14] X. Wei, J. Leng, C. Sun, W. Huo, Q. Ren, and F. Sun, "Co-optimization method of speed planning and energy management for fuel cell vehicles through signalized intersections," *Journal of Power Sources,* vol. 518, p. 230598, 2022.

[15] Y. Zhang, C. Zhang, R. Fan, C. Deng, S. Wan, and H. Chaoui, "Energy management strategy for fuel cell vehicles via soft actor-critic-based deep reinforcement learning considering powertrain thermal and durability characteristics," *Energy Conversion and Management,* vol. 283, p. 116921, 2023.

[16] A. Pozzi, S. Bae, Y. Choi, F. Borrelli, D. M. Raimondo, and S. Moura, "Ecological Velocity Planning through Signalized Intersections: A Deep Reinforcement Learning Approach," in *2020 59th IEEE Conference on Decision and Control (CDC)*, 2020: IEEE, pp. 245-252.

[17] B. Liu, C. Sun, B. Wang, and F. Sun, "Adaptive Speed Planning of Connected and Automated Vehicles Using Multi-light Trained Deep Reinforcement Learning," *IEEE Transactions on Vehicular Technology,* 2021.

[18] Z. Nie and H. Farzaneh, "Real-time dynamic predictive cruise control for enhancing eco-driving of electric vehicles, considering traffic constraints and signal phase and timing (SPaT) information, using artificial-neural-network-based energy consumption model," *Energy,* vol. 241, p. 122888, 2022.

[19] Z. Bai, P. Hao, W. Shangguan, B. Cai, and M. J. Barth, "Hybrid Reinforcement Learning-Based Eco-Driving Strategy for Connected and Automated Vehicles at Signalized Intersections," *IEEE Transactions on Intelligent Transportation Systems,* 2022.

[20] M. Wegener, L. Koch, M. Eisenbarth, and J. Andert, "Automated eco-driving in urban scenarios using deep reinforcement learning,"

*Transportation research part C: emerging technologies,* vol. 126, p. 102967, 2021.
[21] C. Sun, J. Guanetti, F. Borrelli, and S. J. Moura, "Optimal Eco-Driving Control of Connected and Autonomous Vehicles Through Signalized Intersections," *IEEE Internet of Things Journal,* Article vol. 7, no. 5, pp. 3759-3773, 2020, Art no. 8964352.
[22] H. Dong, W. Zhuang, B. Chen, G. Yin, and Y. Wang, "Enhanced eco-approach control of connected electric vehicles at signalized intersection with queue discharge prediction," *IEEE Transactions on Vehicular Technology,* vol. 70, no. 6, pp. 5457-5469, 2021.
[23] H. Dong *et al.*, "A comparative study of energy-efficient driving strategy for connected internal combustion engine and electric vehicles at signalized intersections," *Applied Energy,* vol. 310, p. 118524, 2022.
[24] H. Yang, H. Rakha, and M. V. Ala, "Eco-cooperative adaptive cruise control at signalized intersections considering queue effects," *IEEE Transactions on Intelligent Transportation Systems,* vol. 18, no. 6, pp. 1575-1585, 2016.
[25] S. Dong, H. Chen, B. Gao, L. Guo, and Q. Liu, "Hierarchical energy-efficient control for CAVs at multiple signalized intersections considering queue effects," *IEEE Transactions on Intelligent Transportation Systems,* 2021.
[26] C. Sun, C. Zhang, H. Yu, W. Liang, Q. Ren, and J. Li, "An Eco-driving Approach with Flow Uncertainty Tolerance for Connected Vehicles against Waiting Queue Dynamics on Arterial Roads," *IEEE Transactions on Industrial Informatics,* 2021.
[27] S. Halbach, P. Sharer, S. Pagerit, A. Rousseau, and C. Folkerts, "Model architecture, methods, and interfaces for efficient math-based design and simulation of automotive control systems," *SAE Technical Paper,* vol. 20100241, 2010.
[28] A. Fotouhi, N. Shateri, D. Shona Laila, and D. J. Auger, "Electric vehicle energy consumption estimation for a fleet management system," *International Journal of Sustainable Transportation,* vol. 15, no. 1, pp. 40-54, 2021.
[29] H. X. Liu, X. Wu, W. Ma, and H. Hu, "Real-time queue length estimation for congested signalized intersections," *Transportation research part C: emerging technologies,* vol. 17, no. 4, pp. 412-427, 2009.
[30] N. Vandaele, T. Van Woensel, and A. Verbruggen, "A queueing based traffic flow model," *Transportation Research Part D: Transport and Environment,* vol. 5, no. 2, pp. 121-135, 2000.
[31] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies,* vol. 79, pp. 1-17, 2017.
[32] Y. Huang, H. Hu, J. Tan, C. Lu, and D. Xuan, "Deep reinforcement learning based energy management strategy for range extend fuel cell hybrid electric vehicle," *Energy Conversion and Management,* vol. 277, p. 116678, 2023.
[33] K. Katsaros, R. Kernchen, M. Dianati, and D. Rieck, "Performance study of a Green Light Optimized Speed Advisory (GLOSA) application using an integrated cooperative ITS simulation platform," in *2011 7th International Wireless Communications and Mobile Computing Conference*, 2011: IEEE, pp. 918-923.
[34] J. Li, X. Wu, M. Xu, and Y. Liu, "Deep reinforcement learning and reward shaping based eco-driving control for automated HEVs among signalized intersections," *Energy,* vol. 251, p. 123924, 2022.
[35] F. Xu *et al.*, "Real-time energy optimization of HEVs under connected environment: ECOSM 2021 benchmark problem and a case study," *IEEE/ASME Transactions on Mechatronics,* vol. 24, pp. 1-36, 2020.

# Deep reinforcement learning-based eco-driving control for connected electric vehicles at signalized intersections considering traffic uncertainties

Li, Jie

2023-06-19



Li J, Fotouhi A, Pan W, et al., (2023) Deep reinforcement learning-based eco-driving control for connected electric vehicles at signalized intersections considering traffic uncertainties, Energy, Volume 279, September 2023, Article Number 128139

https://doi.org/10.1016/j.energy.2023.128139